Technická univerzita v Košiciach
Fakulta elektrotechniky a informatiky

KATEDRA MATEMATIKY A
TEORETICKEJ INFORMATIKY

Aplikovaná štatistika

Anna Grinčová – Jana Petrillová

Košice 2019

Recenzovali: RNDr. Miriam Andrejiová, PhD.
RNDr. Štefan Berežný, PhD.

1. vydanie

Za odbornú stránku učebného textu zodpovedajú autori.
Rukopis neprešiel redakčnou ani jazykovou úpravou.

ISBN 978-80-553-3213-0

Obsah

Úvod

Táto učebnica je určená pre študentov prvého ročníka druhého stupňa štúdia Fakulty elektrotechniky a informatiky Technickej univerzity v Košiciach (FEI TU), ale môže poslúžiť aj študentom iných fakúlt.

Učebnica je rozdelená do siedmich kapitol, ktoré obsahujú základné teoretické poznatky potrebné k riešeniu príkladov a vzorové riešené úlohy k učivu, ktoré je preberané v predmete Aplikovaná štatistika.

Cieľom tejto učebnice nebolo podať ucelený teoretický prehľad riešenej problematiky, preto je vhodné kombinovať používanie tejto učebnice s vysokoškolskou učebnicou Štatistické metódy v praxi autorky Miriam Andrejiová aj s voľne dostupnými e-learningovými materiálmi Katedry matematiky a teoretickej informatiky FEI TU v Košiciach.

Na záver ďakujeme RNDr. Miriam Andrejiovej, PhD. a RNDr. Štefanovi Berežnému, PhD. za starostlivé prečítanie rukopisu a za cenné pripomienky, ktorými prispeli k zlepšeniu textu tejto učebnice. Zároveň sa chceme vopred ospravedlniť za možné jazykovo–štylistické chyby, pretože daný text neprešiel redakčnou ani jazykovou úpravou.

1 Úvod do štatistiky

Štatistiku ako vednú disciplínu je možné rozdeliť na dve časti:

- *popisnú (deskriptívnu) štatistiku*, ktorá sa zaoberá metódami zberu, spracovania, prezentácie a analýzy údajov súčasne s ich opisom za pomoci špeciálnych štatistických prostriedkov a metód,
- *výberovú (induktívnu) štatistiku*, ktorá nachádza uplatnenie vtedy, ak zozbieranie všetkých údajov nie je z akýchkoľvek dôvodov možné (napr. súbor je veľmi rozsiahly). Vtedy robíme závery o skúmanom jave na základe čiastkových (vybraných) údajov. Zber (výber) týchto údajov tvorí predmet induktívnej štatistiky, ktorej základom je teória pravdepodobnosti. Podstatnou zložkou induktívnej štatistiky je štatistická analýza, ktorej úlohou je poznanie pravidelností, súvislostí a vývojových tendencií hromadných javov. Patrí sem teória odhadu a testovanie štatistických hypotéz.

Popisná a výberová štatistika sú vzájomne prepojené. Predmetom štatistiky ako vednej disciplíny sú *hromadné náhodné javy* (prírodné alebo spoločenské), ktoré sa za presne definovaných vecných, časových a priestorových podmienok viackrát vyskytujú, resp. opakujú (napr. narodenia, úmrtia, spotreba alebo produkcia konkrétnych výrobkov). Skúmaním hromadného náhodného javu zistíme jeho podstatu, vývoj, pravidelnosť, či vzájomné vzťahy.

Prvok, na ktorom pozorujeme konkrétny prejav hromadného javu (napr. osoba, objekt, výrobok), nazývame *štatistická jednotka.*

Množinu všetkých štatistických jednotiek, ktoré majú požadované spoločné vlastnosti, nazývame *štatistický súbor. Rozsah štatistického súboru* charakterizujeme ako počet štatistických jednotiek v danom štatistickom súbore. *Štatistický znak* predstavuje určitú vlastnosť hromadného javu, ktorá je predmetom štatistického skúmania (napr. vek respondentov, veľkosť výrobku, povolanie respondentov).

Štatistické znaky delíme na:

- *kvalitatívne znaky*, ktoré používame, ak sú vlastnosti štatistickej jednotky vyjadrené slovne (napr. spokojnosť zákazníka, povolanie respondentov),
- *kvantitatívne znaky*, ktoré používame, ak sú vlastnosti štatistickej jednotky vyjadrené číselne, merateľné jednotky (napr. hmotnosť výrobku, vek zákazníka).

Kvantitatívne štatistické znaky delíme na:

- *spojité*, ktoré môžu nadobúdať akékoľvek reálne hodnoty z nejakého ohraničeného aj neohraničeného intervalu (napr. šírka výrobku, výška respondenta),

- *diskrétne*, ktoré môžu nadobúdať len niektoré konkrétne hodnoty v rámci nejakého intervalu (napr. počet výrobkov).

Kvalitatívne štatistické znaky delíme na:

- *alternatívne (dichotomické)*, ktoré môžu nadobúdať len dve možné hodnoty (napr. kvalita - nekvalita),
- *množné (viackategoriálne)*, ktoré môžu nadobúdať viac ako dve možné hodnoty (napr. poškodené oblečenie - málo poškodené oblečenie - nepoškodené oblečenie).

Štatistické skúmanie je možné spravidla rozdeliť do troch etáp:

1. *Štatistické zisťovanie*, ktorého úlohou je zozbierať štatistické údaje o skúmaných hromadných javoch a procesoch.
2. *Štatistické spracovanie zistených údajov*, ktorého úlohou je zozbierané štatistické údaje prehľadne usporiadať, roztriediť a spracovať.
3. *Štatistické vyhodnocovanie*, ktorého úlohou je získané výsledky analyzovať, vyhodnotiť a na základe týchto výsledkov sformulovať závery.

2 Popisná štatistika

Na začiatku štatistického zisťovania získame spravidla veľké množstvo údajov, ktoré zapisujeme v takom poradí, v akom sme ich zaznamenali. Prvým krokom pri spracovaní týchto údajov je ich triedenie, čím docielime prehľadnosť štatistického súboru. Zároveň môžu vyniknúť určité charakteristické rysy a zákonitosti analyzovaného súboru.

2.1 Triedenie štatistického súboru

Triedenie štatistického súboru je usporiadanie štatistických jednotiek do skupín, resp. tried homogénnych z hľadiska určitého štatistického znaku. Podľa počtu triediacich znakov rozlišujeme triedenie na *jednostupňové* (triedenie podľa jedného triediaceho znaku, napr. vek respondenta, výška výrobku) a *viacstupňové* (triedenie podľa viacerých triediacich znakov, napr. vek a pohlavie respondenta, výška a šírka výrobku). My budeme uvažovať len jednostupňové triedenie.

Vzhľadom na množstvo rôznych hodnôt, rozdeľujeme triedenie štatistického súboru na *jednoduché* a *intervalové*. V tejto učebnici budeme pracovať najmä s kvantitatívnymi štatistickými znakmi.

Pri triedení štatistického súboru je potrebné dodržať dve pravidlá:

- *pravidlo úplnosti*, t. j. každú štatistickú jednotku musíme zatriediť,
- *pravidlo jednoznačnosti*, t. j. každú štatistickú jednotku musíme zatriediť práve do jednej triedy.

2.1.1 Jednoduché triedenie

Jednoduché triedenie použijeme, ak je štatistický znak diskrétny alebo spojitý a pritom nadobúda málo rôznych, často sa opakujúcich hodnôt. Nech $x_1, x_2, \ldots, x_n$ sú získané hodnoty sledovaného štatistického znaku, pričom rozsah štatistického súboru je n. Nech celkový počet rôznych hodnôt štatistického znaku je k. Z toho vyplýva, že $k \leqq n$. Každá rôzna hodnota x_j štatistického znaku tvorí jednu triedu. Hodnoty $x_1, x_2, \ldots, x_n$ usporiadame do neklesajúcej postupnosti, čím dostávame *variačný rad* v tvare

$$x_{\min} = x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)} = x_{\max}.$$

Pre každú hodnotu x_j, $j = 1, 2, \ldots, k$ štatistického znaku určíme nasledujúce početnosti:

- *Absolútna početnosť* n_j hodnoty x_j štatistického znaku predstavuje počet štatistických jednotiek s rovnakou hodnotou x_j.

Tabuľka 1: Frekvenčná tabuľka (tabuľka rozdelenia početností)

j	x_j	n_j	f_j	N_j	F_j
1	x_1	n_1	f_1	n_1	f_1
2	x_2	n_2	f_2	$n_1 + n_2$	$f_1 + f_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	x_k	n_k	f_k	n	1

- *Relatívna početnosť* f_j je podiel absolútnej početnosti n_j hodnoty x_j štatistického znaku a rozsahu štatistického súboru n. Platí

$$f_j = \frac{n_j}{n}. \tag{1}$$

- *Kumulatívna absolútna početnosť* N_j hodnoty x_j štatistického znaku určuje, koľko štatistických jednotiek má hodnotu menšiu alebo rovnú ako hodnota x_j. Platí

$$N_j = n_1 + n_2 + \ldots + n_j = \sum_{i=1}^{j} n_i. \tag{2}$$

- *Kumulatívna relatívna početnosť* F_j hodnoty x_j štatistického znaku je podiel kumulatívnej absolútnej početnosti N_j hodnoty x_j štatistického znaku a rozsahu štatistického súboru n. Platí

$$F_j = \frac{N_j}{n} = \frac{n_1 + n_2 + \ldots + n_j}{n} = f_1 + f_2 + \ldots + f_j. \tag{3}$$

Relatívne početnosti sa často udávajú v percentách. Hodnota $100\, f_j\, \%$ predstavuje, koľko percent prvkov zo štatistického súboru má hodnotu x_j sledovaného štatistického znaku. Hodnota $100\, F_j\, \%$ udáva, koľko percent prvkov zo štatistického súboru má hodnotu menšiu alebo rovnú ako je hodnota x_j sledovaného štatistického znaku.

Pre početnosti platia nasledujúce vzťahy

$$\sum_{j=1}^{k} n_j = N_k = n, \qquad \sum_{j=1}^{k} f_j = F_k = 1.$$

Početnosti zapíšeme do *frekvenčnej tabuľky (tabuľky rozdelenia početností)* (Tabuľka 1).

2.1.2 Intervalové triedenie

Intervalové triedenie použijeme, ak je štatistický znak spojitý a pritom nadobúda veľké množstvo rôznych, spravidla málo alebo vôbec sa opakujúcich hodnôt. V prípade intervalového triedenia sa štatistické jednotky $x_1, x_2, \ldots, x_n$ v súbore rozdelia do k intervalov (tried), kde $k \leqq n$, n je rozsah štatistického súboru.

Označme I_j ako j-ty triedny interval, kde t_{j-1} je jeho dolná hranica a t_j je horná hranica intervalu. Pri určovaní intervalov je potrebné stanoviť, ktorá z hraníc bude patriť do intervalu. Teda, či budeme uvažovať typ intervalu $(t_{j-1}, t_j\rangle$ alebo $\langle t_{j-1}, t_j)$.

Každý triedny interval I_j bude reprezentovaný tzv. *triednym znakom* z_j, ktorý určíme ako stred daného triedneho intervalu, t. j.platí

$$z_j = \frac{t_{j-1} + t_j}{2}, \quad j = 1, 2, \ldots, k, \tag{4}$$

kde k je počet triednych intervalov.

Pri intervalovom triedení štatistických údajov je dôležité zvoliť správny počet triednych intervalov (tried) a ich šírku. Počet tried k nesmie byť príliš malý, aby to neviedlo k veľmi zjednodušenému pohľadu na vlastnosti súboru a tým by sa mohli zastrieť aj nejaké jeho charakteristické črty. Naopak nemal by byť ani príliš veľký, pretože by sa mohlo stať, že sa spracovanie stane neprehľadným a opäť môžu zaniknúť niektoré zákonitosti charakteristické pre daný súbor. Na stanovenie počtu tried neexistuje jednotný názor ani všeobecný predpis. Existuje však niekoľko odporúčaných postupov. Niektoré si uvedieme:

- $k \approx 1 + 3,322 \log n$ (Sturgessovo pravidlo),
- $k \approx \sqrt{n}$,
- $0,55n^{0,4} \leqq k \leqq 1,25n^{0,4}$, kde n je počet štatistických jednotiek súboru.

Šírku h triedneho intervalu môžeme približne vypočítať podľa nasledujúceho vzťahu

$$h \doteq \frac{R_V}{k}, \tag{5}$$

kde k je počet triednych intervalov a R_V je *variačné rozpätie*, pre ktoré platí

$$R_V = x_{(n)} - x_{(1)} = x_{\max} - x_{\min}. \tag{6}$$

Na záver intervalového triedenia opäť pre každý triedny interval vypočítame jednotlivé početnosti podobne ako v prípade jednoduchého triedenia. Výsledky zapíšeme prehľadne vo forme frekvenčnej tabuľky, kde hodnotu x_j nahradíme hodnotou triedneho znaku z_j.

2.2 Grafické zobrazenie

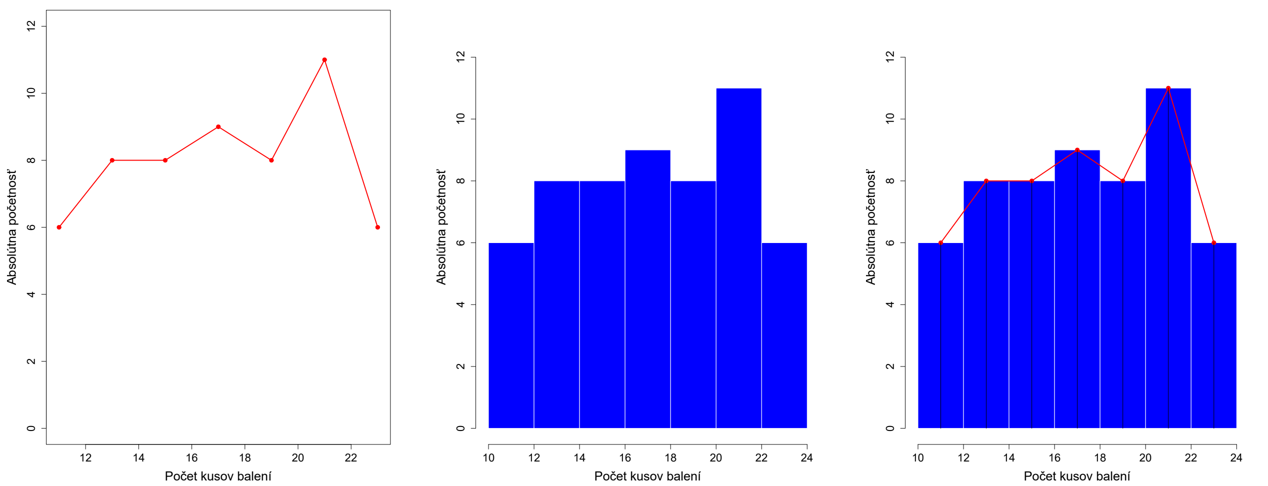
Grafické zobrazenie štatistického súboru je určitá prezentácia spracovania štatistických údajov v sledovanom štatistickom súbore. Poskytuje rýchlejšiu a názornejšiu predstavu o tendenciách, súvislostiach a charakteristických črtách analyzovaných údajov. V štatistike sa používajú rôzne druhy grafov. Uvedieme tie najčastejšie používané: bodový, spojnicový, stĺpcový a kruhový graf.

Bodový graf znázorňuje namerané hodnoty pomocou bodov v pravouhlej súradnicovej sústave. Súradnice týchto bodov tvoria príslušné hodnoty štatistického znaku na osi x a odpovedajúce absolútne, resp. relatívne početnosti na osi y. V prípade intervalového triedenia početností použijeme namiesto x_j hodnotu triedneho znaku z_j. Špeciálnym prípadom bodového grafu je *korelačný graf*, ktorý sa používa v regresnej analýze pri znázornení závislosti dvoch premenných. V tomto prípade sa na os x nanášajú hodnoty prvej (nezávislej) premennej a na os y hodnoty druhej (závislej) premennej.

Spojnicové grafy sú bodové grafy, v ktorých sú jednotlivé body spojené usečkou. Často slúžia na zobrazenie priebehu časového radu, ale aj na zobrazenie absolútnych alebo relatívnych početností.

Polygón početností je spojnicový graf rozdelenia početností (Obr. 1(a)).

Stĺpcový graf je veľmi rozšírený pre svoju jednoduchosť a názornosť. Stĺpce môžu byť umiestnené vertikálne aj horizontálne.

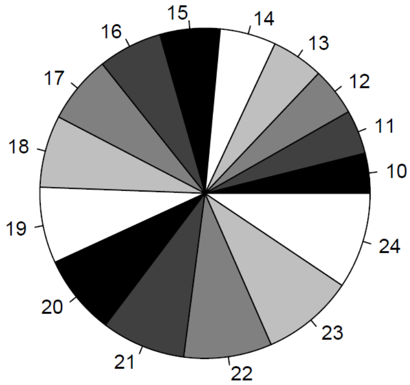


(a) (b) (c)

Obr. 1: Polygón a histogram početností

Histogram je špeciálny stĺpcový graf, ktorý slúži predovšetkým na grafické zobrazenie rozdelenia absolútnych, resp. relatívnych početností (Obr. 1(b)). Stĺpce histogramu sú vždy umiestnené vertikálne. Výška stĺpcov odpovedá hodnote príslušnej početnosti a ich šírka zase odpovedá šírke triedneho intervalu. Pospájaním stredov horných strán všetkých stĺpcov získame polygón početností (Obr. 1(c)). Pri analýze histogramu vieme určiť veľa vlastností skúma-

ného štatistického súboru, ako napríklad minimálnu a maximálnu hodnotu, z nich potom variabilitu skúmaného rozdelenia, symetriu rozdelenia.
Kruhový (koláčový) graf je tzv. plochý graf, kde plocha kruhu predstavuje celý skúmaný súbor rozdelený pomocou kruhových výsekov na jednotlivé časti (Obr. 2).



Obr. 2: Kruhový (koláčový) graf

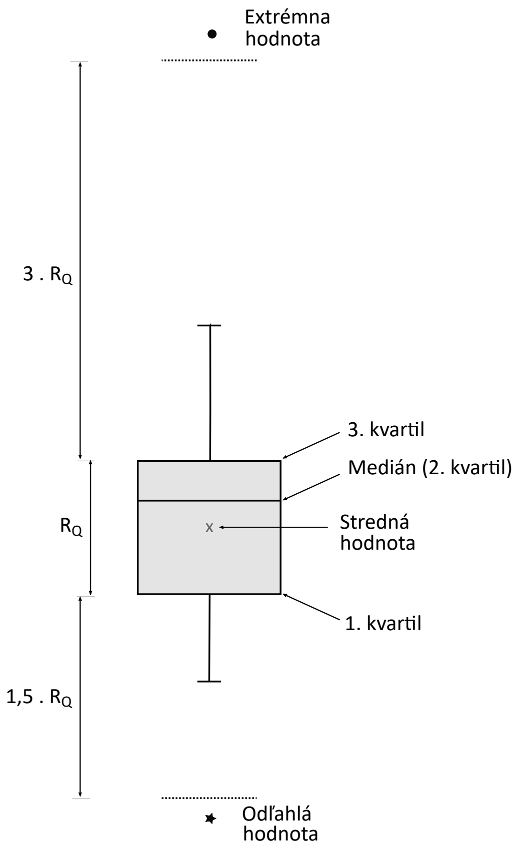
Krabicový diagram (boxplot) je graf, ktorý v jednom obrázku poskytuje informáciu o významných hodnotách (napr. o kvartiloch, o strednej hodnote) aj extrémnych hodnotách štatistického súboru (Obr. 3). Boxplot pozostáva zo základného obdĺžnika (krabice) a z dvoch čiar vychádzajúcich z obdĺžnika, ktoré sú ukončené úsečkami. Umiestnenie dolnej a hornej strany obdĺžnika zodpovedá prvému $\widetilde{x}_{25}$ a tretiemu kvartilu $\widetilde{x}_{75}$, umiestnenie vodorovnej čiary vnútri tohto obdĺžnika zodpovedá mediánu $\widetilde{x}_{50}$, t. j. druhému kvartilu. Výška obdĺžnika sa nazýva *kvartilové (medzikvartilové) rozpätie* R_Q, pre ktoré platí

$$R_Q = \widetilde{x}_{75} - \widetilde{x}_{25}. \tag{7}$$

Vo vnútri obdĺžnika sa nachádza 50% nameraných hodnôt. V najjednoduchšom prípade úsečky na konci čiar vychádzajúcich z obdĺžnika odpovedajú minimálnej a maximálnej hodnote v danom súbore. Vo všeobecnosti tieto čiary môžu mať dĺžku rovnú maximálne 1, 5–násobku výšky obdĺžnika (resp. kvartilového rozpätia). Ak niektorá hodnota presiahne túto hranicu, považuje sa za *odľahlú hodnotu.* V prípade, že presiahne až 3–násobok výšky obdĺžnika, považuje sa za *extrémnu hodnotu.* Tieto hodnoty sú znázornené špeciálnymi znakmi (napr. hviezdičkou, krížikom). Preto boxplot môže slúžiť aj ako grafický test extrémnych a odľahlých hodnôt.

2.3 Číselné charakteristiky

Štatistické spracovanie pomocou tabuliek a grafov uľahčuje analýzu údajov, poskytuje užitočnú informáciu o štruktúre skúmaného štatistického súboru, avšak nie vždy sú to posta-



Obr. 3: Boxplot

čujúce informácie. Preto na popísanie štatistického súboru zavedieme číselné charakteristiky. Patria sem charakteristiky polohy, variability, šikmosti a špicatosti.

2.3.1 Charakteristiky polohy

Charakteristiky polohy charakterizujú polohu (úroveň) znaku, okolo ktorej sú ostatné hodnoty viac, či menej koncentrované. Pri grafickom znázornení rozdelenia početností udávajú polohu znaku na osi x. Medzi charakteristiky polohy patria stredné hodnoty a kvantily. Charakteristiky polohy, ktoré sú určované pomocou všetkých nameraných hodnôt štatistického znaku, nazývame *priemery* (napr. aritmetický, geometrický, harmonický). Medzi charakteristiky polohy, ktoré sú určované pomocou vybraných hodnôt štatistického súboru, patria modus a kvantily. Z kvantilov sa najčastejšie používa medián.

Aritmetický priemer $\overline{x}$ je definovaný ako súčet nameraných hodnôt $x_1, x_2, \ldots, x_n$ delený ich počtom n, t. j.

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{8}$$

Vzorec (8) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet aritmetického priemeru môžeme použiť vzorec

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{k} x_j n_j, \tag{9}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku. Aritmetický priemer počítaný podľa vzorca (9) nazývame *vážený aritmetický priemer*. V prípade intervalového triedenia použijeme vo vzorci (9) namiesto hodnôt x_j hodnoty triedneho znaku z_j. V tomto prípade je ale potrebné si uvedomiť, že dostaneme len približnú hodnotu aritmetického priemeru. Aritmetický priemer má množstvo dôležitých vlastností. Jednou z nich je, že súčet odchýlok jednotlivých hodnôt x_i štatistického znaku od aritmetického priemeru $\overline{x}$ sa rovná nule, t. j.

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0. \tag{10}$$

Aritmetický priemer je citlivý voči extrémnym hodnotám súboru a má dôležitý význam iba vtedy, ak sú odchýlky nameraných hodnôt náhodné a v súbore sa nevyskytujú extrémne hodnoty.

Geometrický priemer $\overline{x}_G$ je definovaný ako n–ta odmocnina zo súčinu nameraných hodnôt $x_1, x_2, \ldots, x_n$, t. j.

$$\overline{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}. \tag{11}$$

Vzorec (11) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet geometrického priemeru môžeme použiť vzorec

$$\overline{x}_G = \sqrt[n]{\prod_{j=1}^{k} x_j^{n_j}} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_k^{n_k}}, \tag{12}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku. Geometrický priemer počítaný podľa vzorca (12) nazývame *vážený geometrický priemer*. Podobne ako v prípade aritmetického priemeru, ak sa jedná o intervalové triedenie, použijeme vo vzorci (12) namiesto hodnôt x_j hodnoty triedneho znaku z_j. Geometrický priemer sa používa napr. na výpočet priemernej inflácie, priemerného úroku, priemerného percentuálneho rastu medzi dvomi časovými obdobiami.

Harmonický priemer $\overline{x}_H$ je definovaný ako podiel počtu nameraných hodnôt $x_1, x_2, \ldots, x_n$ a súčinu ich prevrátených hodnôt, t. j.

$$\overline{x}_H = \frac{n}{\displaystyle\sum_{i=1}^{n} \frac{1}{x_i}}. \tag{13}$$

Analogicky ako v prechádzajúcich prípadoch, vzorec pre *vážený harmonický priemer* má tvar

$$\overline{x}_H = \frac{n}{\sum\limits_{j=1}^{k} \frac{n_j}{x_j}}, \tag{14}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku. Harmonický priemer sa používa predovšetkým pri výpočte priemernej rýchlosti, priemerného času.

Ak sú namerané hodnoty $x_1, x_2, \ldots, x_n$ kladné, platí medzi spomínanými priemermi nasledujúci vzťah

$$\overline{x}_H \leqq \overline{x}_G \leqq \overline{x}. \tag{15}$$

Modus $\widehat{x}$ ($\mathrm{mod}(x)$) je definovaný ako hodnota štatistického znaku s najväčšou početnosťou (v štatistickom súbore sa vyskytuje najčastejšie). Štatistický súbor môže mať viac ako jeden modus. V prípade intervalového rozdelenia početností pri stanovení modusu buď bude postačovať určenie modálneho (najpočetnejšieho) intervalu, alebo v rámci tohto intervalu vypočítame modus $\widehat{x}$ pomocou vzťahu

$$\widehat{x} = a + h\,\frac{d_1}{d_1 + d_2}, \tag{16}$$

kde a je začiatok modálneho intervalu, d_1 je rozdiel medzi absolútnou početnosťou modálneho a predchádzajúceho intervalu, d_2 je rozdiel medzi absolútnou početnosťou modálneho a nasledujúceho intervalu a h je dĺžka triedneho intervalu.

Kvantil $\widetilde{x}_p$ je hodnota, ktorá rozdeľuje usporiadaný súbor hodnôt určitého kvantitatívneho štatistického znaku na dve časti. Jedna časť obsahuje p% hodnôt, ktoré sú menšie a rovnaké ako daný kvantil x_p (napr. 5%, 15%, 50%, 95%). Druhá časť naopak obsahuje $(1-p)$% hodnôt, ktoré sú väčšie a rovnaké ako kvantil x_p (t. j. 95%, 85%, 50%, 5%). V prípade intervalového rozdelenia početností štatistického znaku môžeme získať len približné hodnoty kvantilov. Medzi najčastejšie používané kvantily patria 50% kvantil, kvartily, decily a percentily.

Medián $\widetilde{x}$ je 50% kvantil, teda hodnota štatistického znaku zo štatistického súboru usporiadaného do variačného radu, ktorá delí tento rad na dve rovnako početné časti. Je to vlastne prostredná hodnota variačného radu. Ak rozsah štatistického súboru n je nepárne číslo, tak pre medián platí

$$\widetilde{x} = x_{\left(\frac{n+1}{2}\right)}. \tag{17}$$

Ak rozsah štatistického súboru n je párne číslo, tak medián určíme pomocou vzťahu

$$\widetilde{x} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}\right). \tag{18}$$

V prípade intervalového rozdelenia početností štatistického znaku vypočítame medián podľa vzťahu

$$\widetilde{x} = a + h\,\frac{\frac{n+1}{2} - N_{j-1}}{n_j}, \tag{19}$$

kde a je začiatok triedneho intervalu obsahujúceho medián (mediánového intervalu), h je dĺžka triedneho intervalu, n_j je absolútna početnosť intervalu obsahujúceho medián, n je rozsah štatistického súboru a N_{j-1} je kumulatívna absolútna početnosť predchádzajúceho intervalu pred intervalom obsahujúcim medián.

Kvartily sú tri hodnoty, ktoré rozdeľujú štatistický súbor usporiadaný do variačného radu na štyri rovnako početné častí, pričom každá z nich obsahuje 25% hodnôt. *Prvý (dolný) kvartil* $\widetilde{x}_{25}$ oddeľuje štvrtinu (25%) najmenších hodnôt štatistického znaku od ostatných. *Druhý kvartil* $\widetilde{x}_{50}$ je vlastne medián. *Tretí (horný) kvartil* $\widetilde{x}_{75}$ oddeľuje tri štvrtiny (75%) najmenších hodnôt štatistického znaku od ostatných. V prípade intervalového rozdelenia početností štatistického znaku vypočítame prvý kvartil $\widetilde{x}_{25}$ podľa vzťahu

$$\widetilde{x}_{25} = a + h\,\frac{\frac{n}{4} - N_{j-1}}{n_j}, \tag{20}$$

kde a je začiatok triedneho intervalu obsahujúceho prvý kvartil, h je dĺžka triedneho intervalu, n_j je absolútna početnosť intervalu obsahujúceho prvý kvartil, n je rozsah štatistického súboru a N_{j-1} je kumulatívna absolútna početnosť predchádzajúceho intervalu pred intervalom obsahujúcim prvý kvartil. Pre tretí kvartil $\widetilde{x}_{75}$ v prípade intervalového rozdelenia početností štatistického znaku použijeme vzťah

$$\widetilde{x}_{75} = a + h\,\frac{\frac{3n}{4} - N_{j-1}}{n_j}, \tag{21}$$

kde a je začiatok triedneho intervalu obsahujúceho tretí kvartil, h je dĺžka triedneho intervalu, n_j je absolútna početnosť intervalu obsahujúceho tretí kvartil, n je rozsah štatistického súboru a N_{j-1} je kumulatívna absolútna početnosť predchádzajúceho intervalu pred intervalom obsahujúcim tretí kvartil.

Decily $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_9$ rozdeľujú štatistický súbor usporiadaný do variačného radu na desať rovnako početných častí, pričom každá z nich obsahuje 10% hodnôt.

Percentily $\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_{99}$ rozdeľujú štatistický súbor usporiadaný do variačného radu na sto rovnako početných častí, pričom každá z nich obsahuje 1% hodnôt.

2.3.2 Charakteristiky variability

Charakteristiky variability poskytujú informáciu o homogenite (rovnorodosti) súboru, či sú jednotlivé namerané hodnoty v analyzovanom súbore rozptýlené, či kolísajú. V ideálnom prípade, pokiaľ sú všetky hodnoty súboru rovnaké (x_i je konštanta), hovoríme o nulovej variabilite. Ku najpoužívanejším (základným) charakteristikám variability patria rozptyl (disperzia), smerodajná odchýlka, priemerná odchýlka, variačné rozpätie, kvartilové rozpätie a variačný koeficient.

Rozptyl (disperzia) s^2 je definovaný ako aritmetický priemer zo štvorcov odchýlok jednotlivých nameraných hodnôt $x_1, x_2, \ldots, x_n$ od ich aritmetického priemeru $\overline{x}$, t. j.

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 . \tag{22}$$

Vzorec (22) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet rozptylu môžeme použiť vzorec

$$s^2 = \frac{1}{n} \sum_{j=1}^{k} (x_j - \overline{x})^2 \, n_j, \tag{23}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku. V prípade intervalového triedenia použijeme vo vzorci (23) namiesto hodnôt x_j hodnoty triedneho znaku z_j.

Smerodajná odchýlka s je kladná druhá odmocnina z rozptylu s^2, t. j.

$$s = \sqrt{s^2}. \tag{24}$$

Priemerná odchýlka $\overline{d}$ je definovaná ako aritmetický priemer z absolútnych hodnôt odchýlok všetkých nameraných hodnôt $x_1, x_2, \ldots, x_n$ od aritmetického priemeru, t. j.

$$\overline{d} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \overline{x}|. \tag{25}$$

Vzorec (25) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet priemernej odchýlky môžeme použiť vzorec

$$\overline{d} = \frac{1}{n} \sum_{j=1}^{k} |x_j - \overline{x}| \, n_j, \tag{26}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku. V prípade intervalového triedenia použijeme vo vzorci (26) namiesto hodnôt x_j hodnoty triedneho znaku z_j.

Variačné rozpätie R_V je definované ako rozdiel medzi najväčšou a najmenšou hodnotou v danom súbore nameraných hodnôt $x_1, x_2, \ldots, x_n$ (6). Nevýhodou je jeho závislosť od extrémnych hodnôt a aj to, že neposkytuje informáciu o variabilite hodnôt nachádzajúcich sa medzi najväčšou a najmenšou hodnotou súboru.

Kvartilové (medzikvartilové) rozpätie R_Q je definované ako rozdiel medzi tretím (horným) $\widetilde{x}_{75}$ a prvým (dolným) kvartilom $\widetilde{x}_{25}$ (7).

Variačný koeficient V je definovaný ako pomer smerodajnej odchýlky s a aritmetického priemeru $\overline{x}$, t. j.

$$V = \frac{s}{\overline{x}}\, 100\%. \tag{27}$$

Variačný koeficient V patrí medzi relatívne miery variability. Na rozdiel od predchádzajúcich mier variabilít, umožňuje porovnávať variabilitu štatistického znaku vo viacerých súboroch a taktiež v rôznych merných jednotkách. Čím je hodnota variačného koeficientu väčšia, tým väčší podiel predstavuje smerodajná odchýlka z aritmetického priemeru. To znamená, že tým je väčšia variabilita nameraných hodnôt štatistického znaku. Ak je variačný koeficient väčší ako 40% – 50%, tak je to prejavom značnej nehomogenity štatistického súboru.

2.3.3 Charakteristiky šikmosti a špicatosti

Charakteristiky šikmosti sú založené na porovnávaní stupňa koncentrácie malých hodnôt sledovaného štatistického znaku so stupňom koncentrácie veľkých hodnôt tohto znaku.

Charakteristiky špicatosti sú založené na porovnávaní stupňa koncentrácie hodnôt okolo strednej hodnoty so stupňom koncentrácie ostatných hodnôt.

K najznámejším charakteristikám šikmosti patrí koeficient šikmosti (asymetrie) a k najznámejším charakteristikám špicatosti patrí koeficient špicatosti (excesu). Na ich definíciu potrebujeme poznať pojem centrálny moment r-tého rádu.

Centrálny moment r-tého rádu μ_r štatistického súboru definujeme vzťahom

$$\mu_r = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^r, \tag{28}$$

kde $r \in \mathbb{N} \cup \{0\}$, $\overline{x}$ je aritmetický priemer nameraných hodnôt znaku. Vzorec (28) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak môžeme použiť vzorec

$$\mu_r = \frac{1}{n} \sum_{j=1}^{k} (x_j - \overline{x})^r \, n_j, \tag{29}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku. Platí, že

$$\begin{aligned} \mu_0 &= 1, \\ \mu_1 &= 0, \\ \mu_2 &= s^2, \end{aligned} \tag{30}$$

kde s je smerodajná odchýlka. Pomocou centrálnych momentov tretieho a štvrtého rádu definujeme koeficient šikmosti a koeficient špicatosti.

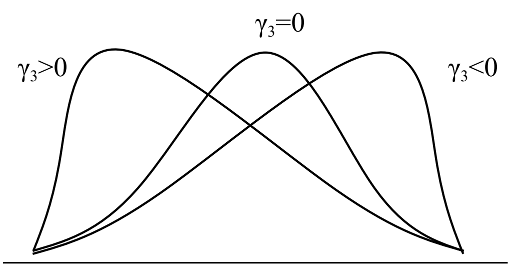
Koeficient šikmosti (asymetrie) γ_3 je definovaný vzťahom

$$\gamma_3 = \frac{\mu_3}{s^3} = \frac{1}{n\, s^3} \sum_{i=1}^{n} (x_i - \overline{x})^3 , \tag{31}$$

kde $\overline{x}$ je aritmetický priemer nameraných hodnôt a s je smerodajná odchýlka. Vzorec (31) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet koeficienta šikmosti môžeme použiť vzorec

$$\gamma_3 = \frac{1}{n\, s^3} \sum_{j=1}^{k} (x_j - \overline{x})^3 \, n_j, \tag{32}$$

kde $\overline{x}$ je aritmetický priemer nameraných hodnôt, s je smerodajná odchýlka a n_j je absolútna početnosť hodnoty x_j štatistického znaku. V prípade intervalového triedenia použijeme vo vzorci (32) namiesto hodnôt x_j hodnoty triedneho znaku z_j. Koeficient šikmosti pomáha rozhodnúť o zhode nameraných údajov s modelom normálneho rozdelenia z hľadiska symetrie okolo aritmetického priemeru $\overline{x}$.



Obr. 4: Koeficient šikmosti

- Ak $\gamma_3 = 0$, tak namerané hodnoty sú symetricky rozložené okolo aritmetického priemeru. Pre stredné hodnoty platí $\overline{x} = \widetilde{x} = \widehat{x}$ a rozdelenie početností je symetrické (Obr. 4).

- Ak $\gamma_3 < 0$, tak pre stredné hodnoty zvyčajne platí $\overline{x} < \widetilde{x} < \widehat{x}$ a rozdelenie početností je *zošikmené záporne (negatívne)* (Obr. 4).
- Ak $\gamma_3 > 0$, tak pre stredné hodnoty zvyčajne platí $\overline{x} > \widetilde{x} > \widehat{x}$ a rozdelenie početností je *zošikmené kladne (pozitívne)* (Obr. 4).

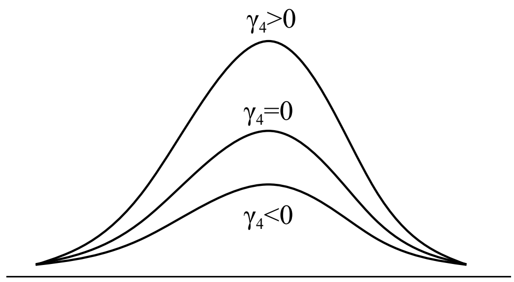
Koeficient špicatosti (excesu) γ_4 je definovaný vzťahom

$$\gamma_4 = \frac{\mu_4}{s^4} - 3 = \frac{1}{n\, s^4} \sum_{i=1}^{n} (x_i - \overline{x})^4 - 3, \tag{33}$$

kde $\overline{x}$ je aritmetický priemer nameraných hodnôt a s je smerodajná odchýlka. Vzorec (33) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet koeficienta špicatosti môžeme použiť vzorec

$$\gamma_4 = \frac{1}{n\, s^4} \sum_{j=1}^{k} (x_j - \overline{x})^4\, n_j - 3, \tag{34}$$

kde $\overline{x}$ je aritmetický priemer nameraných hodnôt, s je smerodajná odchýlka a n_j je absolútna početnosť hodnoty x_j štatistického znaku. V prípade intervalového triedenia použijeme vo vzorci (34) namiesto hodnôt x_j hodnoty triedneho znaku z_j. Koeficient špicatosti vyjadruje strmosť rozdelenia početností nameraných hodnôt štatistického znaku. Teda udáva, či koncentrácia nameraných hodnôt štatistického znaku okolo aritmetického priemeru je v rozdelení početností väčšia alebo menšia ako v súbore s normálnym rozdelením.



Obr. 5: Koeficient špicatosti

- Ak $\gamma_4 = 0$, tak krivka rozdelenia početností je rovnako špicatá ako Gaussova krivka normálneho rozdelenia (Obr. 5).
- Ak $\gamma_4 < 0$, tak krivka rozdelenia početností je plochejšia ako Gaussova krivka normálneho rozdelenia (Obr. 5).
- Ak $\gamma_4 > 0$, tak krivka rozdelenia početností je špicatejšia ako Gaussova krivka normálneho rozdelenia (Obr. 5).

Vyššia špicatosť rozdelenia tiež znamená nižšiu variabilitu hodnôt štatistického znaku.

Poznámka: Ďalšími číselnými charakteristikami sú charakteristiky štatistickej závislosti. Tie sa skúmajú v prípade, že sme v štatistickom súbore merali dva kvantitatívne znaky. Týmto charakteristikám sa budeme osobitne venovať v kapitolách 6 a 7.

Príklad 2.1 *V priebehu 40–tich pracovných dní sme zisťovali celkový počet roztrhnutých vlákien na tkáčskych stavoch s týmito získanými hodnotami [v ks]: 17, 18, 18, 22, 18, 17, 16, 17, 18, 18, 19, 18, 19, 17, 19, 19, 17, 18, 17, 18, 21, 17, 19, 18, 17, 20, 19, 17, 18, 20, 19, 15, 19, 19, 19, 19, 18, 21, 18, 18.*
a) Zostavme variačný rad a frekvenčnú tabuľku početností.
b) Vypočítajme všetky číselné charakteristiky.
c) Zostrojme bodový graf absolútnych početností, polygón absolútnych početností, stĺpcový graf kumulatívnych absolútnych početností, histogram absolútnych početností, kruhový graf a boxplot.

Riešenie:
a) Rozsah súboru nameraných hodnôt je $n = 40$. Všetky hodnoty usporiadame do neklesajúcej postupnosti, čím dostávame variačný rad: 15, 16, 17, 17, 17, 17, 17, 17, 17, 17, 17, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 20, 20, 21, 21, 22.

Do frekvenčnej tabuľky zadávame jednotlivé početnosti, ktoré vypočítame na základe vzťahov (1), (2), (3). Označíme x_i celkový počet roztrhnutých vlákien na tkáčskych stavoch zistený v i–ty pracovný deň, $i = 1, 2, \ldots, 40$ a x_j rôzne hodnoty x_i, $j = 1, 2, \ldots, 8$. Potom n_j je absolútna početnosť hodnoty x_j štatistického znaku, f_j je relatívna početnosť hodnoty x_j štatistického znaku, N_j je kumulatívna absolútna početnosť hodnoty x_j štatistického znaku a F_i je kumulatívna relatívna početnosť hodnoty x_j štatistického znaku. Potom frekvenčná tabuľka početností je uvedená v Tabuľke 2.

b) Ďalšou úlohou je vypočítať všetky číselné charakteristiky. Aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (9)

$$\overline{x} = \frac{1}{40}\left(15 \cdot 1 + 16 \cdot 1 + 17 \cdot 9 + 18 \cdot 13 + 19 \cdot 11 + 20 \cdot 2 + 21 \cdot 2 + 22 \cdot 1\right) = 18,275.$$

Táto hodnota určuje priemerný počet roztrhnutých vlákien za 1 deň. Rozsah štatistického súboru je 40, čo je párne číslo.

Medián vypočítame podľa vzťahu (18)

$$\widetilde{x} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}\right) = \frac{1}{2}\left(x_{(20)} + x_{(21)}\right) = \frac{18+18}{2} = 18.$$

Tabuľka 2: Frekvenčná tabuľka početností pre Príklad 2.1

i	x_j	n_j	N_j	f_j	F_j	f_j [v %]	F_j [v %]
1	15	1	1	0,025	0,025	2,5	2,5
2	16	1	2	0,025	0,050	2,5	5,0
3	17	9	11	0,225	0,275	22,5	27,5
4	18	13	24	0,325	0,600	32,5	60
5	19	11	35	0,275	0,875	27,5	87,5
6	20	2	37	0,050	0,925	5	92,5
7	21	2	39	0,050	0,975	5	97,5
8	22	1	40	0,025	1,00	2,5	100

Je to hodnota, ktorá delí variačný rad na dve rovnako veľké časti.

Modus $\widehat{x}$ je definovaný ako hodnota štatistického znaku s najväčšou početnosťou a v tomto prípade je $\widehat{x} = 18$.

Prvý kvartil $\widetilde{x}_{25} = 17$. Je to aritmetický priemer $x_{(10)}$ a $x_{(11)}$ hodnoty variačného radu. Oddeľuje štvrtinu najmenších hodnôt od ostatných hodnôt.

Tretí kvartil $\widetilde{x}_{75} = 19$. Je to aritmetický priemer $x_{(30)}$ a $x_{(31)}$ hodnoty variačného radu. Oddeľuje tri štvrtiny najmenších hodnôt od ostatných hodnôt.

Rozptyl s^2 vypočítame podľa vzťahu (23)

$$s^2 = \frac{1}{40}\big[1 \cdot (15 - 18,275)^2 + 1 \cdot (16 - 18,275)^2 + 9 \cdot (17 - 18,275)^2 + 13 \cdot (18 - 18,275)^2 +$$

$$+11 \cdot (19 - 18,275)^2 + 2 \cdot (20 - 18,275)^2 + 2 \cdot (21 - 18,275)^2 + 1 \cdot (22 - 18,275)^2\big] = 1,846.$$

Smerodajnú odchýlku s vypočítame podľa vzťahu (24)

$$s = \sqrt{s^2} = \sqrt{1,846} = 1,358.$$

Priemernú odchýlku $\overline{d}$ vypočítame podľa vzťahu (26)

$$\overline{d} = \frac{1}{40}\big(1 \cdot |15 - 18,275| + 1 \cdot |16 - 18,275| + 9 \cdot |17 - 18,275| + 13 \cdot |18 - 18,275| +$$

$$+11 \cdot |19 - 18,275| + 2 \cdot |20 - 18,275| + 2 \cdot |21 - 18,275| + 1 \cdot |22 - 18,275|\big) = 1,3.$$

Variačné rozpätie R_V vypočítame podľa vzťahu (6)

$$R_V = x_{\max} - x_{\min} = 22 - 15 = 7.$$

Kvartilové rozpätie R_Q vypočítame podľa vzťahu (7)

$$R_Q = \widetilde{x}_{75} - \widetilde{x}_{25} = 19 - 17 = 2.$$

Variačný koeficient V vypočítame podľa vzťahu (27)

$$V = \frac{s}{\bar{x}}\,100\% = \frac{1,358}{18,275} \cdot 100\% = 7,4\%.$$

Koeficient šikmosti γ_3 vypočítame podľa vzťahu (32)

$$\gamma_3 = \frac{1}{40 \cdot (1,358)^3}\big[1 \cdot (15 - 18,275)^3 + 1 \cdot (16 - 18,275)^3 + 9 \cdot (17 - 18,275)^3 +$$

$$+ 13 \cdot (18 - 18,275)^3 + 11 \cdot (19 - 18,275)^3 + 2 \cdot (20 - 18,275)^3 +$$

$$+ 2 \cdot (21 - 18,275)^3 + 1 \cdot (22 - 18,275)^3\big] = 0,407.$$

Pretože $\gamma_3 > 0$, rozdelenie početností je zošikmené kladne.

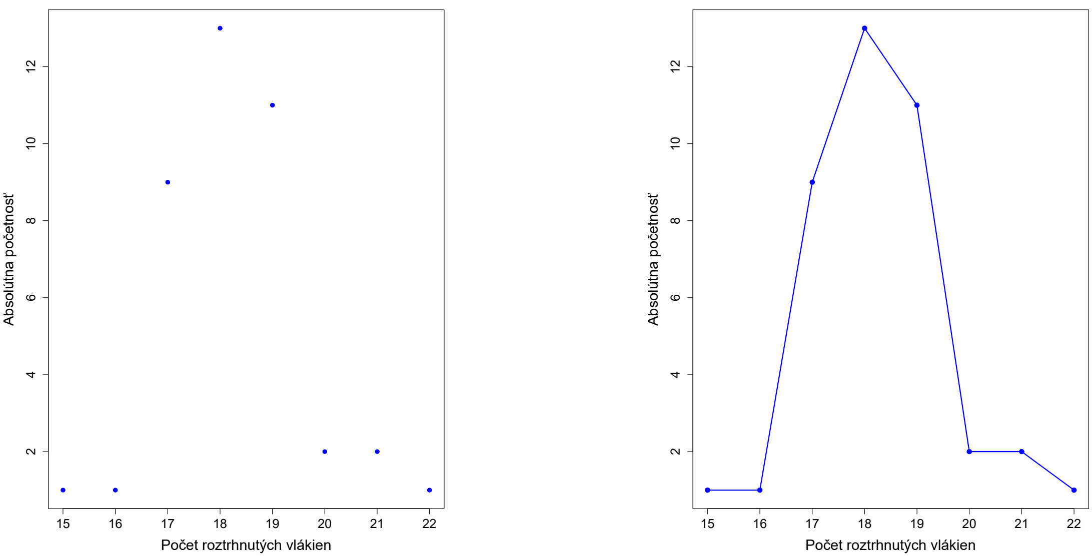
Koeficient špicatosti γ_4 vypočítame podľa vzťahu (34)

$$\gamma_4 = \frac{1}{40 \cdot (1,358)^4}\big[1 \cdot (15 - 18,275)^4 + 1 \cdot (16 - 18,275)^4 + 9 \cdot (17 - 18,275)^4 + 13 \cdot (18 - 18,275)^4 +$$

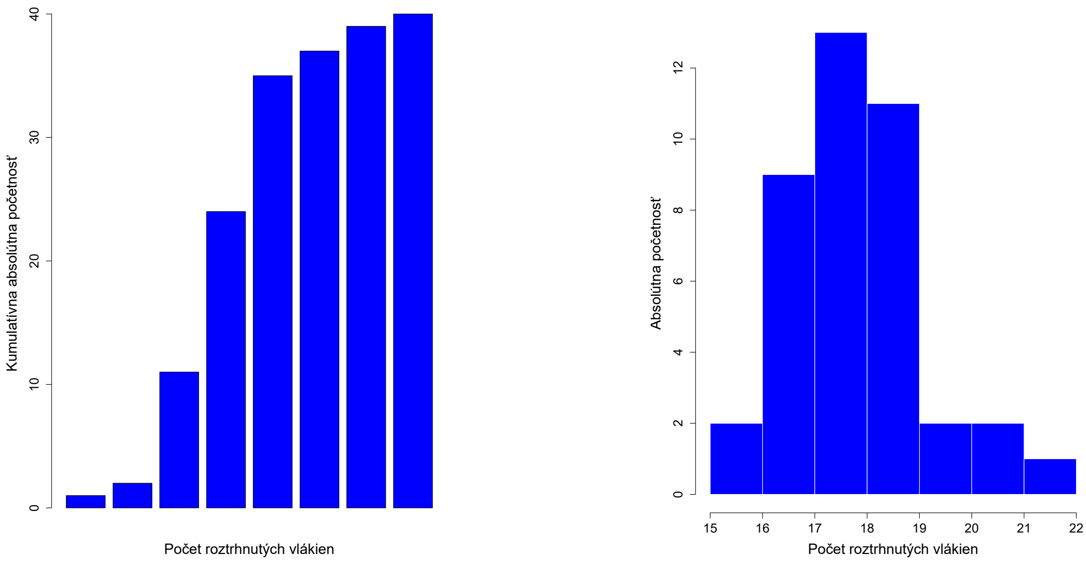
$$+ 11 \cdot (19 - 18,275)^4 + 2 \cdot (20 - 18,275)^4 + 2 \cdot (21 - 18,275)^4 + 1 \cdot (22 - 18,275)^4\big] - 3 = 0,591.$$

Pretože $\gamma_4 > 0$, krivka rozdelenia početností je špicatejšia ako Gaussova krivka normálneho rozdelenia.

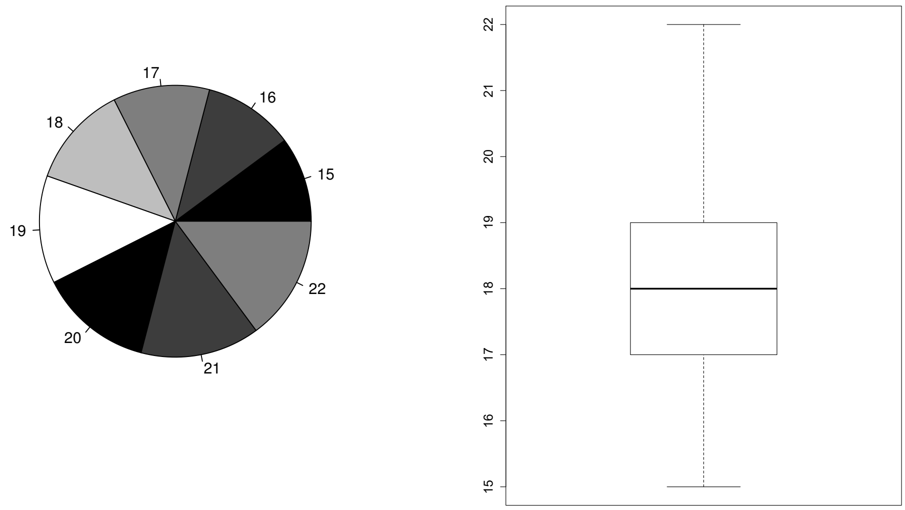
c) Jednotlivé grafy sú zobrazené na Obr. 6, Obr. 7 a Obr. 8.



Obr. 6: Bodový graf absolútnych početností a polygón absolútnych početností z Príkladu 2.1



Obr. 7: Stĺpcový graf kumulatívnych absolútnych početností a histogram absolútnych početností z Príkladu 2.1



Obr. 8: Kruhový (koláčový) graf a boxplot z Príkladu 2.1

Príklad 2.2 *Na 100 vzorkách mlieka sme zisťovali percentuálny obsah tuku v mlieku s týmito získanými hodnotami [v %]: 4,25; 4,26; 4,25; 4,07; 4,20; 4,40; 4,05; 4,00; 4,17; 3,96; 4,30; 4,22; 4,25; 4,15; 4,25; 4,25; 3,97; 4,10; 4,47; 4,10; 4,17; 4,05; 4,25; 4,18; 4,05; 4,27; 4,03; 4,05; 4,15; 4,15; 4,40; 4,27; 4,25; 4,12; 4,30; 4,17; 4,20; 4,17; 4,22; 4,32; 4,20; 4,05; 4,20; 4,37; 4,20; 4,25; 4,20; 4,15; 4,33; 4,10; 4,15; 4,25; 4,25; 4,05; 4,32; 4,27; 4,12; 4,27; 4,50;*

4,38; 4,27; 4,17; 4,17; 4,12; 4,25; 4,12; 4,30; 4,33; 4,33; 4,21; 4,12; 4,15; 4,22; 4,20; 4,25; 4,50; 4,27; 4,10; 4,15; 4,31; 4,32; 4,10; 4,10; 4,17; 4,32; 4,02; 4,20; 4,17; 4,30; 4,44; 4,17; 4,20; 4,45; 4,07; 4,27; 4,15; 4,12; 4,42; 4,21; 4,35.

a) Štatistický súbor roztrieďme do triednych intervalov a zostavme frekvenčnú tabuľku početností.

b) Vypočítajme všetky číselné charakteristiky.

c) Zostrojme polygón absolútnych početností, histogram absolútnych početností a boxplot.

Riešenie:

a) Počet triednych intervalov k zvolíme pomocou pravidla

$$k \approx \sqrt{n} = \sqrt{100} = 10.$$

Variačné rozpätie R_V vypočítame podľa vzťahu (6)

$$R_V = x_{\max} - x_{\min} = 4,50 - 3,96 = 0,54.$$

Šírku h triedneho intervalu vypočítame podľa vzťahu (5)

$$h = \frac{0,54}{10} = 0,054.$$

Ak si však zvolíme počet triednych intervalov 11, šírka h triedneho intervalu bude $h = \frac{0{,}54}{11} = 0,05$ a my budeme pracovať s „krajšími“ hodnotami. Zvolíme typ intervalu $(t_{j-1}; t_j\rangle$. Hodnotu triedneho znaku z_j jednotlivých intervalov vypočítame podľa vzťahu (4).

Podobne ako v prechádzajúcom príklade, do frekvenčnej tabuľky zadávame jednotlivé početnosti, ktoré vypočítame na základe vzťahov (1), (2), (3). Potom frekvenčná tabuľka početností je uvedená v Tabuľke 3.

b) Ďalšou úlohou je vypočítať všetky číselné charakteristiky. Vo vzťahoch, ktoré budeme používať, nahradíme hodnotu x_j hodnotou triedneho znaku z_j.

Aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (9)

$$\overline{x} = \frac{1}{100}\big(3,975 \cdot 3 + 4,025 \cdot 8 + 4,075 \cdot 8 + 4,125 \cdot 14 + 4,175 \cdot 18+$$

$$+4,225 \cdot 18 + 4,275 \cdot 12 + 4,325 \cdot 9 + 4,375 \cdot 4 + 4,425 \cdot 3 + 4,475 \cdot 3\big) = 4,212.$$

Táto hodnota určuje priemerný obsah tuku v mlieku v jednej vzorke.

Medián $\widetilde{x}$ vypočítame podľa vzťahu (19)

$$\widetilde{x} = 4,15 + 0,05\,\frac{\frac{100+1}{2} - 33}{18} = 4,2.$$

Tabuľka 3: Frekvenčná tabuľka početností pre Príklad 2.2

j	$I_j = (t_{j-1}; t_j\rangle$	z_j	n_j	N_j	f_j	F_j	f_j [v %]	F_j [v %]
1	$\langle 3,95; 4,00\rangle$	3,975	3	3	0,03	0,03	3	3
2	$(4,00; 4,05\rangle$	4,025	8	11	0,08	0,11	8	11
3	$(4,05; 4,10\rangle$	4,075	8	19	0,08	0,19	8	19
4	$(4,10; 4,15\rangle$	4,125	14	33	0,14	0,33	14	33
5	$(4,15; 4,20\rangle$	4,175	18	51	0,18	0,51	18	51
6	$(4,20; 4,25\rangle$	4,225	18	69	0,18	0,69	18	69
7	$(4,25; 4,30\rangle$	4,275	12	81	0,12	0,81	12	81
8	$(4,30; 4,35\rangle$	4,325	9	90	0,09	0,90	9	90
9	$(4,35; 4,40\rangle$	4,375	4	94	0,04	0,94	4	94
10	$(4,40; 4,45\rangle$	4,425	3	97	0,03	0,97	3	97
11	$(4,45; 4,50\rangle$	4,475	3	100	0,03	1,00	3	100

Keďže máme dva najpočetnejšie intervaly I_5 a I_6, môžeme mať dva modusy. Modus vypočítame podľa vzťahu (16)

$$\widehat{x}_1 = 4,15 + 0,05 \cdot \frac{4}{4+0} = 4,2,$$

$$\widehat{x}_2 = 4,20 + 0,05 \cdot \frac{0}{0-6} = 4,2.$$

Zistili sme, že v oboch prípadoch sa modus rovná hodnote 4,2. Avšak táto hodnota do intervalu I_6 nepatrí. Preto ostal len jeden modus $\widehat{x} = 4,2$.

Prvý kvartil $\widetilde{x}_{25}$ vypočítame podľa vzťahu (20)

$$\widetilde{x}_{25} = 4,10 + 0,05 \cdot \frac{\frac{100}{4} - 19}{14} = 4,121.$$

Je to aritmetický priemer $x_{(25)}$ a $x_{(26)}$ hodnoty variačného radu.

Tretí kvartil $\widetilde{x}_{75}$ vypočítame podľa vzťahu (21)

$$\widetilde{x}_{75} = 4,25 + 0,05 \cdot \frac{\frac{300}{4} - 69}{12} = 4,275.$$

Je to aritmetický priemer $x_{(75)}$ a $x_{(76)}$ hodnoty variačného radu.

Rozptyl s^2 vypočítame podľa vzťahu (23)

$$s^2 = \frac{1}{100}\big[3\cdot(3,975-4,212)^2 + 8\cdot(4,025-4,212)^2 + 8\cdot(4,075-4,212)^2 + 14\cdot(4,125-4,212)^2 +$$

$$+18\cdot(4,175-4,212)^2 + 18\cdot(4,225-4,212)^2 + 12\cdot(4,275-4,212)^2 + 9\cdot(4,325-4,212)^2 +$$

$$+4\cdot(4,375-4,212)^2+3\cdot(4,425-4,212)^2+3\cdot(4,475-4,212)^2\big]=0,013.$$

Smerodajnú odchýlku s vypočítame podľa vzťahu (24)

$$s=\sqrt{s^2}=\sqrt{0,013}=0,114.$$

Priemernú odchýlku $\overline{d}$ vypočítame podľa vzťahu (26)

$$\overline{d}=\frac{1}{100}\big[3\cdot|3,975-4,212|+8\cdot|4,025-4,212|+8\cdot|4,075-4,212|+14\cdot|4,125-4,212|+$$

$$+18\cdot|4,175-4,212|+18\cdot|4,225-4,212|+12\cdot|4,275-4,212|+9\cdot|4,325-4,212|+$$

$$+4\cdot|4,375-4,212|+3\cdot|4,425-4,212|+3\cdot|4,475-4,212|\big]=0,091.$$

Variačné rozpätie R_V vypočítame podľa vzťahu (6)

$$R_V=x_{\max}-x_{\min}=4,5-3,96=0,54.$$

Kvartilové rozpätie R_Q vypočítame podľa vzťahu (7)

$$R_Q=\widetilde{x}_{75}-\widetilde{x}_{25}=4,275-4,121=0,154.$$

Variačný koeficient V vypočítame podľa vzťahu (27)

$$V=\frac{s}{\overline{x}}100\%=\frac{0,114}{4,212}\cdot100\%=2,7\%.$$

Koeficient šikmosti γ_3 vypočítame podľa vzťahu (32)

$$\gamma_3=\frac{1}{100\cdot(0,114)^3}\big[3\cdot(3,975-4,212)^3+8\cdot(4,025-4,212)^3+8\cdot(4,075-4,212)^3+$$

$$+14\cdot(4,125-4,212)^3+18\cdot(4,175-4,212)^3+18\cdot(4,225-4,212)^3+12\cdot(4,275-4,212)^3+$$

$$+9\cdot(4,325-4,212)^3+4\cdot(4,375-4,212)^3+3\cdot(4,425-4,212)^3+3\cdot(4,475-4,212)^3\big]=0,265.$$

Pretože $\gamma_3>0$, rozdelenie početností je zošikmené kladne.

Koeficient špicatosti γ_4 vypočítame podľa vzťahu (34)

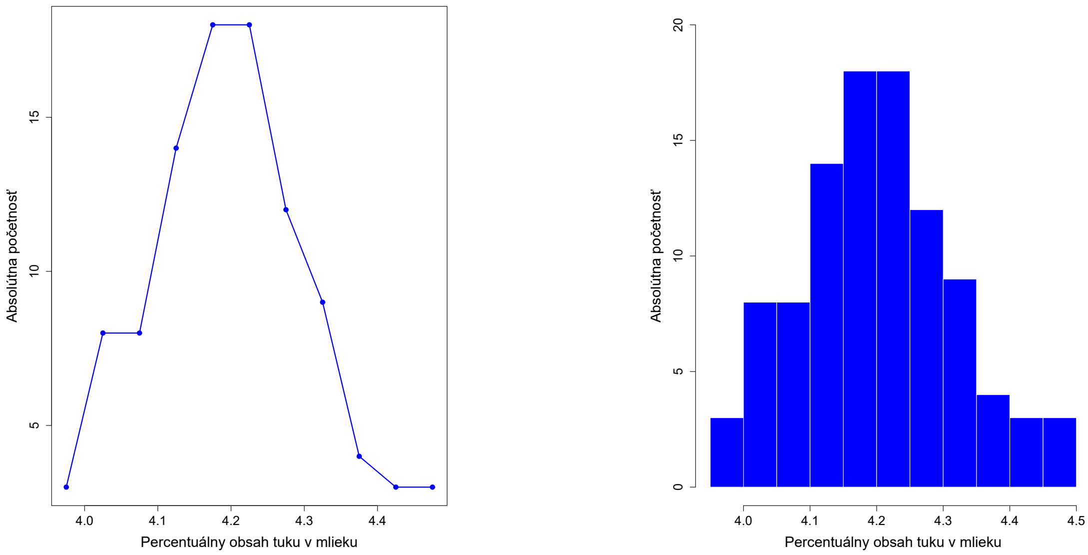
$$\gamma_4=\frac{1}{100\cdot(0,114)^4}\big[3\cdot(3,975-4,212)^4+8\cdot(4,025-4,212)^4+8\cdot(4,075-4,212)^4+$$

$$+14\cdot(4,125-4,212)^4+18\cdot(4,175-4,212)^4+18\cdot(4,225-4,212)^4+12\cdot(4,275-4,212)^4+$$

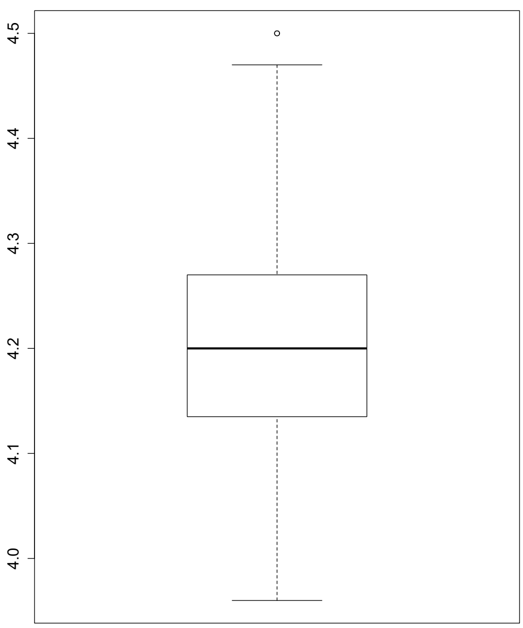
$$+9\cdot(4,325-4,212)^4+4\cdot(4,375-4,212)^4+3\cdot(4,425-4,212)^4+3\cdot(4,475-4,212)^4\big]-3=-0,134.$$

Pretože $\gamma_4<0$, krivka rozdelenia početností je plochejšia ako Gaussova krivka normálneho rozdelenia.

c) Jednotlivé grafy sú zobrazené na Obr. 9 a Obr. 10.



Obr. 9: Polygón absolútnych početností a histogram absolútnych početností z Príkladu 2.2



Obr. 10: Boxplot z Príkladu 2.2

3 Náhodná premenná a náhodný výber

Teória pravdepodobnosti je dôležitou súčasťou štatistiky a štatistických metód. K základným pojmom patrí náhodný pokus, náhodný jav a pravdepodobnosť náhodného javu. *Náhodný pokus* je každá činnosť, ktorá sa niekoľkokrát opakuje za rovnakých alebo približne rovnakých podmienok. *Náhodný jav* je každý možný výsledok náhodného pokusu, ktorý môže byť charakterizovaRný kvalitatívne (slovne) alebo kvantitatívne (číselne). V mnohých prípadoch je výhodnejšie charakterizovať výsledok kvantitatívne (číselne), k čomu používame náhodnú premennú (veličinu). *Pravdepodobnosť náhodného javu* A je číslo, ktoré kvantitatívne charakterizuje možnosť nastatia náhodného javu A. Označujeme ju $P(A)$.

3.1 Náhodná premenná a jej číselné charakteristiky

Náhodná premenná (veličina) X je zobrazenie, ktoré priraďuje každému prvku z množiny náhodných javov reálne číslo. Náhodné premenné budeme označovať veľkými písmenami X, Y, Z, ... a ich konkrétne hodnoty malými písmenami x, y, z, ...

Podľa toho, aké hodnoty nadobúda náhodná premenná, delíme ju na:

- *diskrétnu náhodnú premennú*, ktorá nadobúda konečný alebo spočítateľný počet možných hodnôt (napr. počet zákazníkov, vek respondenta),
- *spojitú náhodnú premennú*, ktorá nadobúda ľubovoľnú hodnotu z nejakého (ohraničeného alebo neohraničeného) intervalu (napr. životnosť výrobku, šírka výrobku).

Poznať náhodnú premennú znamená vedieť jej *rozdelenie pravdepodobnosti*, teda potrebujeme určiť, aké hodnoty a s akými pravdepodobnosťami ich nadobúda. To znamená, pre každé reálne číslo x vieme určiť pravdepodobnosť toho, že hodnota náhodnej premennej X je rovná číslu x (označujeme $P(X = x)$). Rozdelenie pravdepodobnosti môže byť zadané tabuľkou, funkciou alebo aj grafom. *Distribučná funkcia* priradí každému reálnemu číslu pravdepodobnosť, že náhodná premenná X nadobudne hodnotu menšiu alebo rovnú číslu x, t. j.

$$F(x) = P(X \leqq x). \tag{35}$$

Je to univerzálna forma popisu rozdelenia pravdepodobnosti, lebo sa používa pre diskrétnu aj spojitú náhodnú premennú. Pomocou distribučnej funkcie sa dá popísať aj opačná pravdepodobnosť, teda pravdepodobnosť, že náhodná premenná X nadobudne hodnotu väčšiu než číslo x, t. j.

$$P(X > x) = 1 - P(X \leqq x) = 1 - F(x).$$

Vlastnosti distribučnej funkcie:

1. Distribučná funkcia nadobúda len hodnoty z intervalu $\langle 0;1\rangle$, t. j. $0 \leqq F(x) \leqq 1$ pre každé $x \in \mathbb{R}$.

2. Distribučná funkcia je neklesajúca funkcia.

3. Distribučná funkcia je sprava spojitá funkcia na $\mathbb{R}$.

4. $\lim_{x\to-\infty} F(x) = 0$ a $\lim_{x\to\infty} F(x) = 1$.

5. Ak $a < b$, tak $P(X \in (a,b\rangle) = P(a < X \leqq b) = F(b) - F(a)$.

Najdôležitejšou charakteristikou polohy náhodnej premennej X je *stredná hodnota* $E(X)$. Jej definícia závisí od toho, či ide o diskrétnu alebo spojitú náhodnú premennú, preto ju zadefinujeme v nasledujúcich kapitolách pre tieto náhodné premenné osobitne.

Základnou charakteristikou variability je *rozptyl (disperzia)* $D(X)$. Rozptyl náhodnej premennej X je stredná hodnota štvorca odchýlky náhodnej premennej X od jej strednej hodnoty $E(X)$, t. j.

$$D(X) = E\left([X - E(X)]^2\right) = E\left(X^2\right) - [E(X)]^2. \tag{36}$$

Druhú odmocninu z rozptylu náhodnej premennej X nazývame *smerodajná odchýlka* $\sigma(X)$, t. j.

$$\sigma(X) = \sqrt{D(X)}. \tag{37}$$

3.1.1 Diskrétna náhodná premenná

Diskrétna náhodná premenná je taká náhodná premenná X, ktorá nadobúda hodnoty $x_1, x_2, \ldots, x_n$ s pravdepodobnosťami $p_1, p_2, \ldots, p_n$, t. j.

$$p_i = P(X = x_i) \text{ pre } i = 1, 2, \ldots, n, \tag{38}$$

pričom je splnená tzv. *normalizačná podmienka*

$$\sum_{i=1}^{n} p_i = 1, \tag{39}$$

kde $p_i \geqq 0$ pre $i = 1, 2, \ldots, n$.

Rozdelenie pravdepodobnosti diskrétnej náhodnej premennej X môžeme zadať pravdepodobnostnou tabuľkou alebo distribučnou funkciou diskrétnej náhodnej premennej.

Tabuľka 4: Pravdepodobnostná tabuľka (tabuľka rozdelenia pravdepodobnosti)

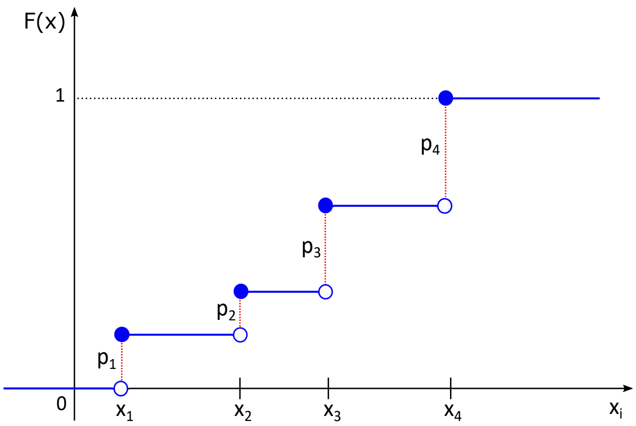
x_i	x_1	x_2	$\cdots$	x_n
p_i	p_1	p_2	$\cdots$	p_n

Pravdepodobnostná tabuľka (tabuľka rozdelenia pravdepodobnosti) diskrétnej náhodnej premennej (Tabuľka 4) je tabuľka, ktorá obsahuje možné hodnoty náhodnej premennej x_i a im príslušné pravdepodobnosti $p_i = P(X = x_i)$ pre $i = 1, 2, \ldots, n$.

Distribučná funkcia diskrétnej náhodnej premennej má tvar

$$F(x) = P(X \leqq x) = \sum_{x_i \leqq x} p_i. \tag{40}$$

Graf distribučnej funkcie diskrétnej náhodnej premennej má tzv. schodovitý tvar, kde veľkosť skoku v bode x_i sa rovná pravdepodobnosti p_i (Obr. 11).



Obr. 11: Graf distribučnej funkcie diskrétnej náhodnej premennej

Stredná hodnota $E(X)$ diskrétnej náhodnej premennej X je definovaná vzťahom

$$E(X) = \sum_{i=1}^{n} x_i p_i. \tag{41}$$

Po dosadení strednej hodnoty diskrétnej náhodnej premennej do všeobecného vzorca (36), dostávame pre *rozptyl $D(X)$ diskrétnej náhodnej premennej X* vzťah

$$D(X) = \sum_{i=1}^{n} (x_i - E(X))^2 \, p_i = \sum_{i=1}^{n} x_i^2 p_i - \left(\sum_{i=1}^{n} x_i p_i \right)^2. \tag{42}$$

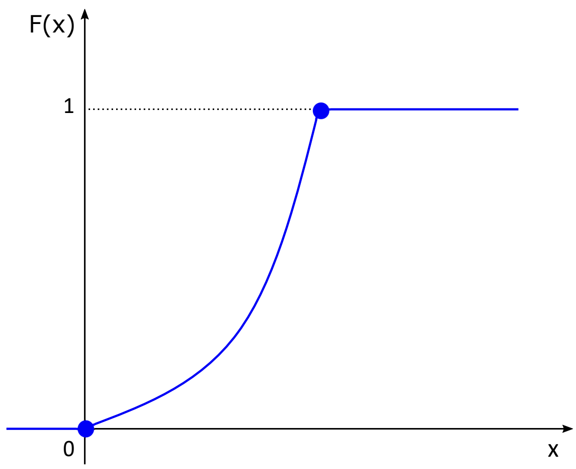
3.1.2 Spojitá náhodná premenná

Spojitá náhodná premenná je náhodná premenná X, pre ktorú existuje taká reálna, nezáporná, integrovateľná funkcia $f(x)$, že pre všetky $x \in \mathbb{R}$ možno distribučnú funkciu vyjadriť v tvare

$$F(x) = P(X \leqq x) = \int_{-\infty}^{x} f(t)\, dt, \text{ pre všetky } x \in \mathbb{R}. \tag{43}$$

Funkcia $f(x)$ sa nazýva *hustota pravdepodobnosti spojitej náhodnej premennej* a má nasledujúce vlastnosti:

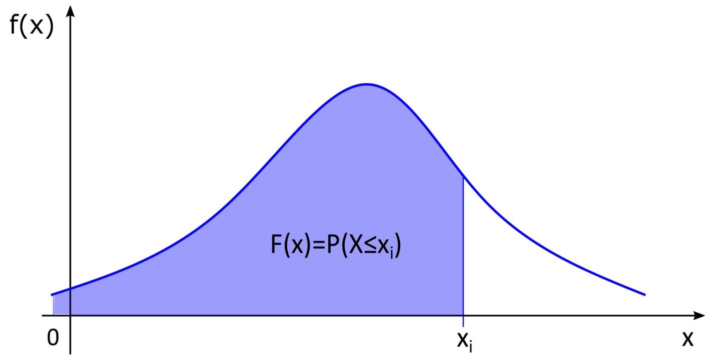
1. Ak existuje hodnota $f(x)$, tak $f(x) \geqq 0$.
2. Ak existuje $F'(x)$, tak $F'(x) = f(x)$ pre všetky $x \in \mathbb{R}$.
3. *Normalizačná podmienka*:
$$\int_{-\infty}^{\infty} f(x)\, dx = 1, \text{ pre všetky } x \in \mathbb{R}. \tag{44}$$
4. Ak $a < b$, potom
$$P(a \leqq X \leqq b) = P(a < X \leqq b) = P(a \leqq X < b) =$$
$$= P(a < X < b) = \int_{a}^{b} f(x)\, dx = F(b) - F(a).$$



Obr. 12: Graf distribučnej funkcie spojitej náhodnej premennej

Rozdelenie pravdepodobnosti spojitej náhodnej premennej X môžeme zadať hustotou pravdepodobnosti $f(x)$ spojitej náhodnej premennej X alebo distribučnou funkciou spojitej náhodnej premennej X (Obr. 12).

Vzťah hustoty pravdepodobnosti a distribučnej funkcie spojitej náhodnej premennej X, ktorý vyplýva zo vzťahu (43), je znázornený na Obr. 13.



Obr. 13: Vzťah hustoty pravdepodobnosti a distribučnej funkcie

Stredná hodnota $E(X)$ *spojitej náhodnej premennej* X je definovaná vzťahom

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx. \tag{45}$$

Po dosadení strednej hodnoty spojitej náhodnej premennej do všeobecného vzorca (36), dostávame pre *rozptyl* $D(X)$ *spojitej náhodnej premennej* X vzťah

$$D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - [E(X)]^2 . \tag{46}$$

3.2 Vybrané rozdelenia náhodných premenných

V tejto kapitole spomenieme tie rozdelenia pravdepodobnosti náhodných premenných, s ktorými sa stretneme v učebnici. Pri každom rozdelení uvedieme základné vlastnosti a charakteristiky.

3.2.1 Binomické rozdelenie

Binomické rozdelenie patrí medzi rozdelenia pravdepodobnosti diskrétnych náhodných premenných. Máme n nezávislých pokusov, z ktorých každý končí iba dvoma možnými výsledkami: jav nastane, resp. jav nenastane. Pravdepodobnosť, že jav nastane je p, pravdepodobnosť, že jav nenastane je $1 - p$, kde $0 < p < 1$ a p je v každom pokuse rovnaká.

Diskrétna náhodná premenná X má *binomické rozdelenie* pravdepodobnosti určené parametrami n a p, ak hodnoty $x = 0, 1, \dots, n$ nadobúda s pravdepodobnosťou

$$p(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x},\ \ p \in (0, 1),\ \ n = 1, 2, \dots , \tag{47}$$

kde p je pravdepodobnosť nastatia javu a n je počet nezávislých pokusov.
Zapisujeme $X \sim \mathrm{Bi}(n, p)$.
Pre číselné charakteristiky platí:

- Stredná hodnota:
$$E(X) = np, \tag{48}$$
- Disperzia:
$$D(X) = np(1-p). \tag{49}$$

3.2.2 Poissonovo rozdelenie

Poissonovo rozdelenie patrí medzi rozdelenia pravdepodobnosti diskrétnych náhodných premenných. Používa sa na modelovanie procesov, pri ktorých sledujeme počet výskytu istého javu v danom časovom intervale alebo v priestorovej oblasti (napr. počet hodín práce bez porúch, počet dopravných nehôd za mesiac, počet obslúžených zákazníkov za týždeň).

Diskrétna náhodná premenná X má *Poissonovo rozdelenie* pravdepodobnosti určené parametrom λ, $\lambda > 0$, ak hodnoty $x = 0, 1, \ldots, n$ nadobúda s pravdepodobnosťou
$$p(x) = P(X = x) = \frac{\lambda^x}{x!}\, e^{-\lambda}. \tag{50}$$
Zapisujeme $X \sim \mathrm{Po}(\lambda)$.
Pre číselné charakteristiky platí:

- Stredná hodnota:
$$E(X) = \lambda, \tag{51}$$
- Disperzia:
$$D(X) = \lambda. \tag{52}$$

3.2.3 Rovnomerné rozdelenie

Rovnomerné rozdelenie pravdepodobnosti patrí medzi rozdelenia pravdepodobnosti spojitých náhodných premenných. Spojitá náhodná premenná X nadobúda hodnoty len na intervale konkrétnej dĺžky, pričom všetky hodnoty náhodnej premennej z daného intervalu sú rovnako pravdepodobné (napr. doba čakania na autobus, príchod vlakov).

Spojitá náhodná premenná X má *rovnomerné rozdelenie* pravdepodobnosti určené parametrami a, b, $a < b$, ak jej hustota pravdepodobnosti $f(x)$ má tvar
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pre } x \in \langle a, b \rangle, \\ 0 & \text{pre } x \notin \langle a, b \rangle. \end{cases} \tag{53}$$

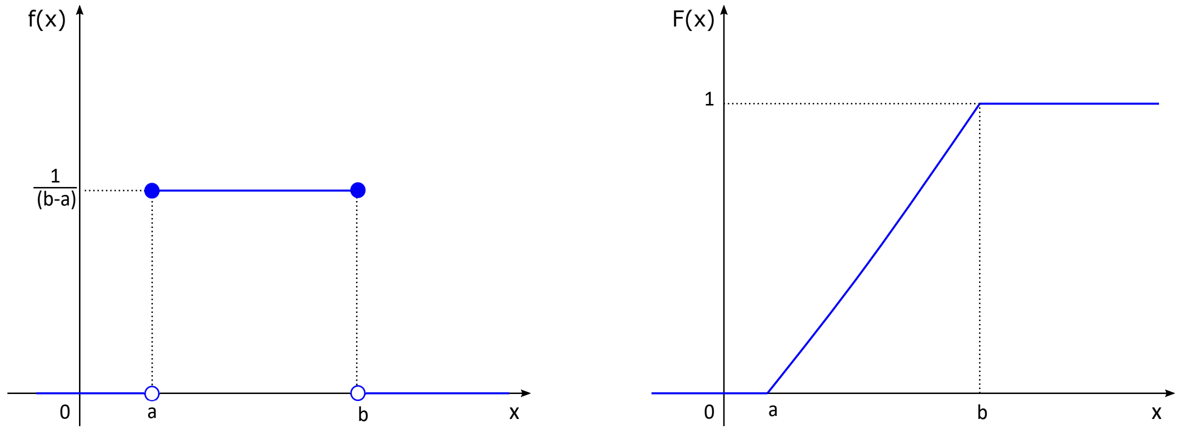
Zapisujeme $X \sim R(a, b)$.
Pre číselné charakteristiky a distribučnú funkciu platí:

- Stredná hodnota:
$$E(X) = \frac{a+b}{2}, \tag{54}$$

- Disperzia:
$$D(X) = \frac{(b-a)^2}{12}, \tag{55}$$

- Distribučná funkcia:
$$F(x) = \begin{cases} 0 & \text{pre } x < a, \\ \frac{x-a}{b-a} & \text{pre } x \in \langle a, b \rangle, \\ 1 & \text{pre } x > b. \end{cases} \tag{56}$$

Grafy hustoty pravdepodobnosti a distribučnej funkcie rovnomerného rozdelenia sú znázornené na Obr. 14.



Obr. 14: Hustota pravdepodobnosti a distribučná funkcia rovnomerného rozdelenia

3.2.4 Exponenciálne rozdelenie

Exponenciálne rozdelenie patrí medzi rozdelenia pravdepodobnosti spojitých náhodných premenných a je typické pre úlohy súvisiace so životnosťou zariadení, výrobkov, pričom stredná doba životnosti je daná parametrom λ.

Spojitá náhodná premenná X má *exponenciálne rozdelenie* pravdepodobnosti určené parametrom λ, $\lambda > 0$, ak jej hustota pravdepodobnosti $f(x)$ má tvar

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & \text{pre } x \geqq 0, \\ 0 & \text{pre } x < 0. \end{cases} \tag{57}$$

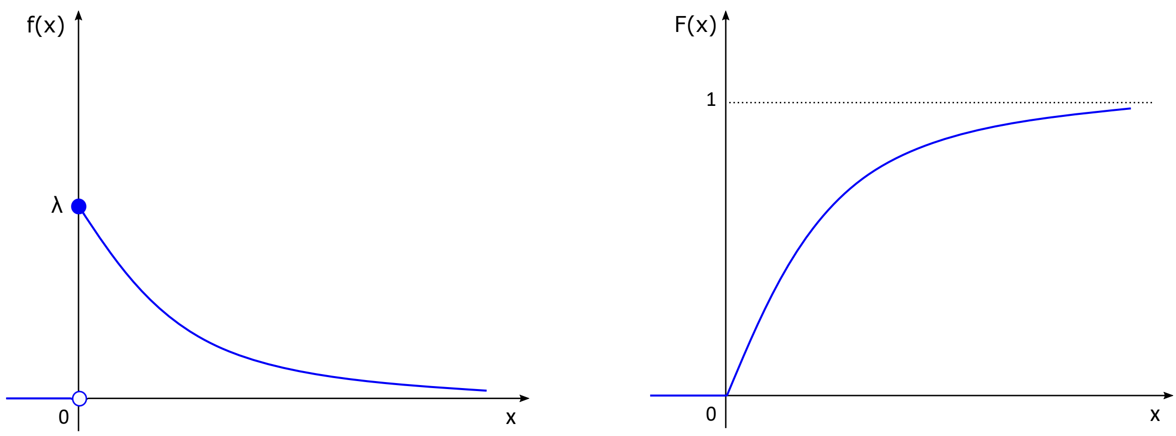
Zapisujeme $X \sim \mathrm{Ex}(\lambda)$.
Pre číselné charakteristiky a distribučnú funkciu platí:

- Stredná hodnota:
$$E(X) = \lambda, \tag{58}$$

- Disperzia:
$$D(X) = \lambda^2, \tag{59}$$

- Distribučná funkcia:
$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\lambda}} & \text{pre } x \geqq 0, \\ 0 & \text{pre } x < 0. \end{cases} \tag{60}$$

Grafy hustoty pravdepodobnosti a distribučnej funkcie exponenciálneho rozdelenia sú znázornené na Obr. 15.



Obr. 15: Hustota pravdepodobnosti a distribučná funkcia exponenciálneho rozdelenia

Poznámka: V niektorých literatúrach sa uvádza exponenciálne rozdelenie s parametrom α, pričom $\alpha = \frac{1}{\lambda}$.

3.2.5 Normálne rozdelenie

Normálne rozdelenie sa používa ako model v úlohách, kde náhodná premenná X kolíše okolo strednej hodnoty μ s rozptylom σ^2. V praxi je to jedno z najpoužívanejších rozdelení pravdepodobností (napr. chyby meraní).

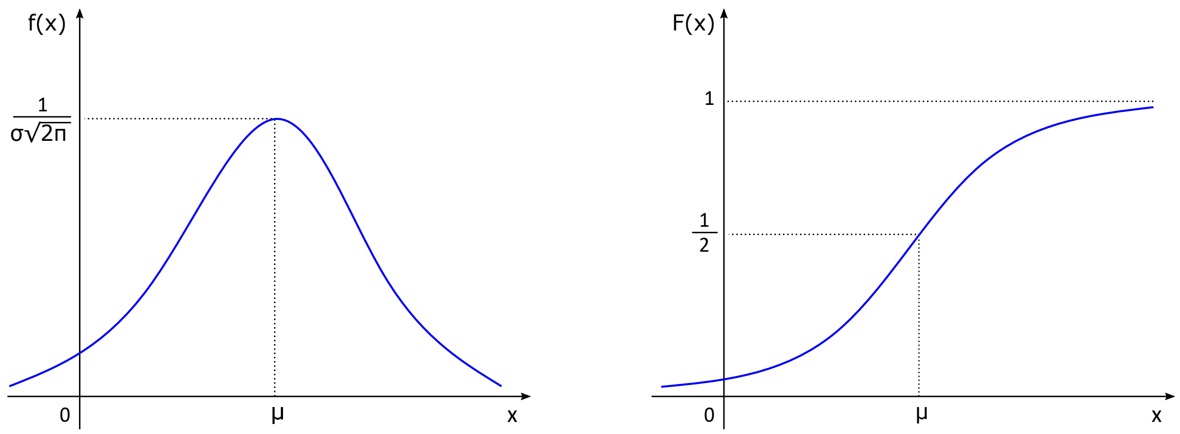
Spojitá náhodná premenná X má *normálne rozdelenie* pravdepodobnosti určené parametrami μ, σ^2, $\sigma > 0$, ak jej hustota pravdepodobnosti $f(x)$ má tvar

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{pre každé } x \in \mathbb{R}. \tag{61}$$

Zapisujeme $X \sim \mathrm{N}(\mu, \sigma^2)$.
Pre číselné charakteristiky a distribučnú funkciu platí:

- Stredná hodnota:
$$E(X) = \mu, \tag{62}$$
- Disperzia:
$$D(X) = \sigma^2, \tag{63}$$
- Distribučná funkcia:
$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\, dt. \tag{64}$$

Graf hustoty pravdepodobnosti $f(x)$ sa nazýva *Gaussova krivka* a má tzv. zvonovitý tvar symetrický vzhľadom na strednú hodnotu μ. Spolu s grafom distribučnej funkcie normálneho rozdelenia sú znázornené na Obr. 16.



Obr. 16: Hustota pravdepodobnosti a distribučná funkcia normálneho rozdelenia

Normované normálne rozdelenie

Spojitá náhodná premenná Y má *normované normálne rozdelenie* pravdepodobnosti určené parametrami $\mu = 0$, $\sigma^2 = 1$, ak jej hustota pravdepodobnosti $\varphi(y)$ má tvar
$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \quad \text{pre každé } y \in \mathbb{R}. \tag{65}$$
Zapisujeme $Y \sim \mathrm{N}(0, 1)$.
Pre číselné charakteristiky a distribučnú funkciu platí:

- Stredná hodnota:
$$E(Y) = 0, \tag{66}$$

- Disperzia:

$$D(Y) = 1, \tag{67}$$

- Distribučná funkcia:

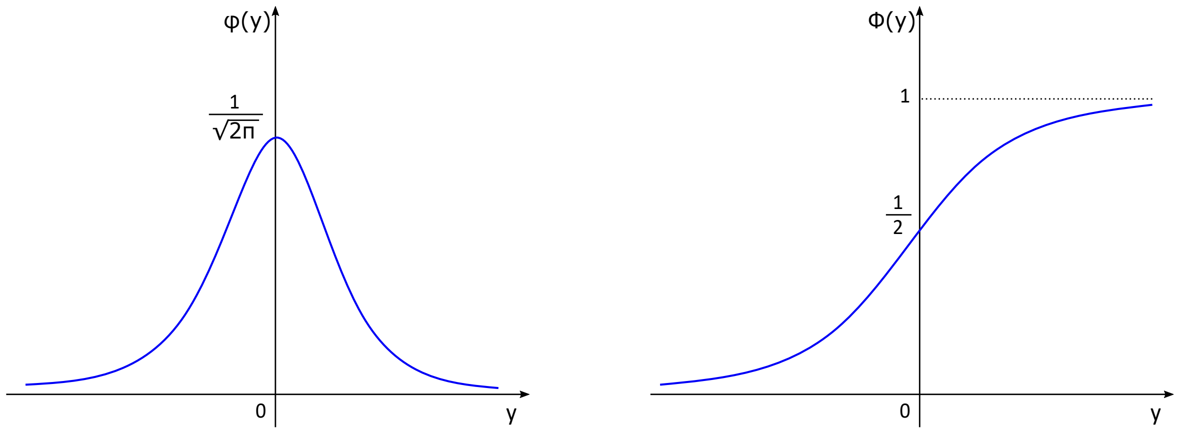
$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{y} e^{-\frac{t^2}{2}}\, dt. \tag{68}$$

Keďže funkcia $\varphi(y)$ je párna a platí pre ňu normalizačná podmienka (44), pre distribučnú funkciu $\Phi(y)$ dostávame vlastnosť symetrie, t. j.

$$\Phi(-y) = 1 - \Phi(y). \tag{69}$$

Hodnoty distribučnej funkcie $\Phi(y)$ sú tabelované len pre $y \geqq 0$ (Tabuľka 27). Na základe podmienky (69) vieme určiť hodnoty $\Phi(y)$ aj pre $y < 0$.

Grafy hustoty pravdepodobnosti a distribučnej funkcie normovaného normálneho rozdelenia sú znázornené na Obr. 17.



Obr. 17: Hustota pravdepodobnosti a distribučná funkcia normovaného normálneho rozdelenia

Vzťah medzi normovaným normálnym rozdelením a normálnym rozdelením:

1. Ak náhodná premenná X má normálne rozdelenie $\mathrm{N}(\mu, \sigma^2)$, tak $Y = \frac{X-\mu}{\sigma}$ je k nej prislušnou normovanou náhodnou premennou a má normované normálne rozdelenie $\mathrm{N}(0, 1)$.

2. Ak náhodná premenná X má normálne rozdelenie $\mathrm{N}(\mu, \sigma^2)$ s distribučnou funkciou $F(x)$, tak pre každé $x \in \mathbb{R}$ platí

$$P(X \leqq x) = F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{70}$$

3. Ak náhodná premenná X má normálne rozdelenie $\mathrm{N}(\mu, \sigma^2)$ s distribučnou funkciou $F(x)$, tak pre všetky $a, b \in \mathbb{R}$, $a < b$ platí

$$P(a \leqq X \leqq b) = P(a < X \leqq b) = P(a \leqq X < b) =$$

$$= P(a < X < b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

3.2.6 Ďalšie rozdelenia

V tejto kapitole uvedieme prehľad tých rozdelení pravdepodobnosti, ktoré sme neuviedli v predchádzajúcich kapitolách, ale budeme ich využívať v teórii odhadov a pri testovaní štatistických hypotéz.

Najskôr potrebujeme zaviesť pojem kvantilová funkcia. Nech $F(x)$ je distribučná funkcia náhodnej premennej X a $\alpha \in (0; 1)$. Potom *kvantilová funkcia* $F^{-1}(\alpha)$ je definovaná ako najmenšie číslo q, $q \in \mathbb{R}$ také, že $F(q) \geqq \alpha$. Hodnotu $q = F^{-1}(\alpha)$ nazývame *α–kvantilom* v rozdelení pravdepodobnosti danom distribučnou funkciou $F(x)$.

Studentovo t–rozdelenie

- Tvar hustoty rozdelenia závisí od parametra k, ktorý určuje počet stupňov voľnosti.
- Je symetrické okolo nuly a s rastúcim počtom stupňov voľnosti sa stále viac podobá normálnemu rozdeleniu. Pre $k > 30$ sa dá dobre aproximovať normálnym rozdelením $N(0, \frac{k}{k-2})$.
- Kvantily Studentovho t–rozdelenia $t_\alpha(k)$ sú tabelované pre rôzne stupne voľnosti (Tabuľka 29).

χ^2–rozdelenie

- Tvar hustoty rozdelenia závisí od parametra k, ktorý určuje počet stupňov voľnosti.
- Nie je symetrické pre malé k, ale pre $k > 100$ sa dá dobre aproximovať normálnym rozdelením $N(k, 2k)$.
- Kvantily χ^2–rozdelenia $\chi^2_\alpha(k)$ sú tabelované pre rôzne stupne voľnosti (Tabuľky 30 – 32).

Fisherovo F–rozdelenie

- Tvar hustoty rozdelenia závisí od dvoch parametrov k_1 a k_2.

- S rastúcimi parametrami sa hustota pravdepodobnosti stáva súmernejšou a stále viac sa podobá normálnemu rozdeleniu.

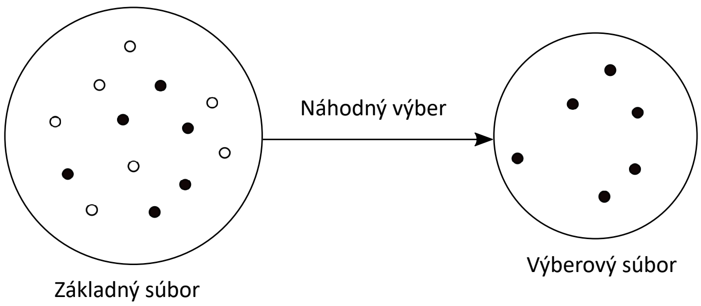
- Nie je symetrické rozdelenie.

- Kvantily Fisherovho F–rozdelenia $F_\alpha(k_1, k_2)$ sú tabelované pre rôzne stupne voľnosti (Tabuľky 33–40) a platí
$$F_\alpha(k_1, k_2) = \frac{1}{F_{1-\alpha}(k_2, k_1)}.$$

Poznámka: Kvantily normovaného normálneho rozdelenia $\mathrm{N}(0, 1)$ budeme označovať u_α a sú tabelované pre rôzne hodnoty α (Tabuľka 28).

3.3 Náhodný výber a výberové číselné charakteristiky

V praxi nemáme vždy dostupné informácie o všetkých štatistických jednotkách skúmaného súboru. Preto často svoju analýzu musíme obmedziť na menšiu časť (podmnožinu, výber) štatistického súboru. Súbor, ktorý je predmetom výskumu, nazývame *základný súbor*. Súbor štatistických jednotiek, ktoré boli z neho určitým spôsobom vybrané, nazývame *výberový súbor* (Obr. 18). Poznatky získané analýzou výberového súboru môžeme potom zovšeobecniť na základný súbor. Výberový súbor by mal byť reprezentatívny a mal by mať štruktúru podobnú štruktúre základného súboru. Najčastejším spôsobom ako dosiahnuť tieto vlastnosti, je *náhodný výber*. To znamená, že o zaradení do výberového súboru rozhoduje len náhoda. Takýmto náhodným výberom sú napr. losovanie, generátor náhodných čísel. Náhodný výber je realizovaný z netriedeného základného súboru a každá štatistická jednotka má rovnakú pravdepodobnosť výberu.



Obr. 18: Základný a výberový súbor

Označme V_n náhodný výber zo základného súboru s nameranými hodnotami $x_1, x_2, \ldots, x_n$ a rozsahom n. Ako sme ukázali v predchádzajúcej kapitole, štatistický súbor je možné opísať pomocou rôznych číselných charakteristík. Číselné charakteristiky základného súboru sa odhadujú pomocou príslušných výberových charakteristík. Medzi základné výberové číselné charakteristiky patria výberový aritmetický priemer, výberový rozptyl a výberová smerodajná odchýlka.

Výberový aritmetický priemer $\overline{x}$ náhodného výberu V_n je definovaný ako súčet hodnôt $x_1, x_2, \ldots, x_n$ delený ich počtom n, t. j.

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{71}$$

Vzorec (71) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet výberového aritmetického priemeru môžeme použiť vzorec

$$\overline{x} = \frac{1}{n} \sum_{j=1}^{k} x_j n_j, \tag{72}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku.

Výberový rozptyl (disperzia) s^2 náhodného výberu V_n je definovaný nasledovne

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2, \tag{73}$$

kde $\overline{x}$ je výberový aritmetický priemer náhodného výberu V_n. Vzorec (73) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet výberového rozptylu môžeme použiť vzorec

$$s^2 = \frac{1}{(n-1)} \sum_{j=1}^{k} (x_j - \overline{x})^2 \, n_j, \tag{74}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku.

Poznámka: V prípade, že poznáme strednú hodnotu náhodného výberu V_n, výberový rozptyl je definovaný ako aritmetický priemer zo štvorcov odchýlok jednotlivých hodnôt $x_1, x_2, \ldots, x_n$ od ich strednej hodnoty μ, t. j.

$$s_0^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2. \tag{75}$$

Vzorec (75) sa používa pre neroztriedený štatistický súbor. Ak sú hodnoty štatistického znaku roztriedené do k tried, tak na výpočet výberového rozptylu s_0^2 môžeme použiť vzorec

$$s_0^2 = \frac{1}{n} \sum_{j=1}^{k} (x_j - \mu)^2 \, n_j, \tag{76}$$

kde n_j je absolútna početnosť hodnoty x_j štatistického znaku.

Výberová smerodajná odchýlka s náhodného výberu V_n je kladná druhá odmocnina z výberového rozptylu s^2, t. j.

$$s = \sqrt{s^2}. \tag{77}$$

Príklad 3.1 *Pri vážení dvadsiatich kilogramových vreciek cukru sme zistili nasledovné hodnoty [v kg]: 1,00; 1,01; 1,05; 0,99; 0,95; 1,00; 0,98; 0,99; 1,04; 1,06; 0,93; 1,00; 1,03; 0,97; 1,00; 0,99; 1,05; 1,01; 0,94; 1,00. Pre daný výberový súbor vypočítajme*
a) výberový aritmetický priemer,
b) výberový rozptyl,
c) výberovú smerodajnú odchýlku.

Riešenie:
a) Výberový aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (71)

$$\overline{x} = \frac{1}{20}\big(1,00 + 1,01 + 1,05 + 0,99 + 0,95 + 1,00 + 0,98 + 0,99 + 1,04 + 1,06 + 0,93+$$

$$+1,00 + 1,03 + 0,97 + 1,00 + 0,99 + 1,05 + 1,01 + 0,94 + 1,00\big) = 0,9995.$$

Táto hodnota vyjadruje priemerné množstvo cukru v jednom vrecku v kilogramoch.
b) Výberový rozptyl s^2 vypočítame podľa vzťahu (73)

$$s^2 = \frac{1}{19}\big[(1,00 - 0,9995)^2 + (1,01 - 0,9995)^2 + (1,05 - 0,9995)^2 + (0,99 - 0,9995)^2+$$

$$+(0,95-0,9995)^2+(1,00-0,9995)^2+(0,98-0,9995)^2+(0,99-0,9995)^2+(1,04-0,9995)^2+$$

$$+(1,06-0,9995)^2+(0,93-0,9995)^2+(1,00-0,9995)^2+(1,03-0,9995)^2+(0,97-0,9995)^2+$$

$$+(1,00-0,9995)^2+(0,99-0,9995)^2+(1,05-0,9995)^2+(1,01-0,9995)^2+(0,94-0,9995)^2+$$

$$+(1,00 - 0,9995)^2\big] = 0,00126.$$

c) Výberovú smerodajnú odchýlku s vypočítame podľa vzťahu (77)

$$s = \sqrt{s^2} = 0,0355.$$

Túto hodnotu môžeme zhruba považovať za priemernú odchýlku nameraných hodnôt od priemernej hodnoty.

4 Teória odhadu

Úlohou teórie odhadu je čo najlepšie odhadnúť neznáme parametre rozdelenia základného súboru, z ktorého pochádza náhodný výber. Existujú dva typy odhadov:

- *Bodový odhad* je odhad jedinou hodnotou, a to nahradením neznámeho parametra hodnotou vhodnej výberovej charakteristiky.
- *Intervalový odhad* je odhad intervalom, ktorý s vopred danou pravdepodobnosťou obsahuje hodnotu neznámeho parametra.

4.1 Bodový odhad

Bodový odhad parametra základného súboru spočíva v nahradení neznámeho parametra jedným číslom, konkrétne vhodnou výberovou charakteristikou získanou z hodnôt náhodného výberu.

Nech základný súbor má normálne rozdelenie so strednou hodnotou μ a smerodajnou odchýlkou σ. Potom bodovým odhadom strednej hodnoty μ je výberový aritmetický priemer $\overline{x}$, bodovým odhadom rozptylu σ^2 je výberový rozptyl s^2 a bodovým odhadom smerodajnej odchýlky σ je výberová smerodajná odchýlka s. Podobne platí, ak základný súbor má exponenciálne rozdelenie so strednou hodnotou λ, potom bodovým odhadom strednej hodnoty λ je výberový aritmetický priemer $\overline{x}$.

Každý kvalitný bodový odhad musí spĺňať nasledujúce vlastnosti:

- *nestrannosť (nevychýlenosť)* – pri opakovaných výberoch odhad parametra kolíše okolo teoretickej hodnoty symetricky na obe strany,
- *konzistentnosť* – zvyšovaním počtu štatistických jednotiek náhodného výberu sa stále viac koncentruje okolo odhadovanej hodnoty,
- *výdatnosť* – pri opakovaných výberoch je rozptyl odhadu zo všetkých možných odhadov minimálny,
- *robustnosť* – odľahlé hodnoty spôsobené napríklad hrubou chybou merania nemajú veľký vplyv na hodnotu odhadu.

4.2 Intervalový odhad

Intervalový odhad parametra základného súboru spočíva v nájdení takého intervalu spoľahlivosti, ktorý s vopred danou pravdepodobnosťou γ obsahuje hodnotu neznámeho parametra.

Táto pravdepodobnosť γ sa nazýva *koeficient spoľahlivosti* odhadu. Číslo $\alpha = 1 - \gamma$ sa nazýva *hladina významnosti* a udáva, akú chybovosť sme ochotní pri intervalovom odhade parametra pripustiť, teda v koľkých percentách sa skutočná hodnota parametra nebude nachádzať. Najčastejšie sa za hladinu významnosti volia hladiny $0,01$ a $0,05$. Získaný interval potom nazývame $100(1-\alpha)\%$*–ný interval spoľahlivosti* a v tých najčastejších prípadoch je to 99%–ný, resp. 95%–ný interval spoľahlivosti.

Interval spoľahlivosti môže mať tri tvary. Môže byť obojstranný (ohraničený sprava aj zľava) alebo jednostranný (ohraničený len z jednej strany, zľava alebo sprava).
Obojstranným intervalom spoľahlivosti pre parameter Q základného súboru nazývame odhad pomocou číselného intervalu $\langle Q_d; Q_h\rangle$, pričom pre pravdepodobnosť platí

$$P(Q_d \leqq Q \leqq Q_h) = 1 - \alpha, \text{ kde } \alpha \in (0;1).$$

Ľavostranným intervalom spoľahlivosti pre parameter Q základného súboru nazývame odhad pomocou číselného intervalu $\langle Q_d; \infty)$, pričom pre pravdepodobnosť platí

$$P(Q_d \leqq Q) = 1 - \alpha, \text{ kde } \alpha \in (0;1).$$

Pravostranným intervalom spoľahlivosti pre parameter Q základného súboru nazývame odhad pomocou číselného intervalu $(-\infty; Q_h\rangle$, pričom pre pravdepodobnosť platí

$$P(Q \leqq Q_h) = 1 - \alpha, \text{ kde } \alpha \in (0;1).$$

4.2.1 Interval spoľahlivosti na odhad parametrov normálneho rozdelenia

Uvažujme náhodný výber V_n, ktorý pochádza zo základného súboru s normálnym rozdelením so strednou hodnotou μ a smerodajnou odchýlkou σ. Zapisujeme $V_n \sim N(\mu, \sigma^2)$.

Pri určovaní intervalu spoľahlivosti jednotlivých parametrov budeme rozlišovať nasledujúce prípady:

- odhad strednej hodnoty μ, ak poznáme rozptyl σ^2,
- odhad strednej hodnoty μ, ak nepoznáme rozptyl σ^2,
- odhad rozptylu σ^2, ak poznáme strednú hodnotu μ,
- odhad rozptylu σ^2, ak nepoznáme strednú hodnotu μ.

Pri konštrukcii intervalu spoľahlivosti budeme využívať pojem *vhodná štatistika. Štatistikou* $g(x)$ nazývame takú funkciu náhodného výberu, rozdelenie ktorej nezávisí od parametrov rozdelenia, z ktorého náhodný výber pochádza.

Postup pri konštrukcii intervalu spoľahlivosti pre parameter Q je nasledovný:

1. Určíme bodový odhad parametra Q.

2. Doplníme tento odhad na vhodnú štatistiku $g(x)$.

3. Určíme čísla g_1, g_2, $g_1 < g_2$ tak, aby platilo $P(g \leqq g_1) = \alpha_1$, $P(g_2 \leqq g) = \alpha_2$, kde $\alpha_1 + \alpha_2 = \alpha$, α_1, $\alpha_2 \in (0; 1)$. V prípade, že distribučná funkcia $F(x)$ je spojitá a rastúca, potom g_1, g_2 sú kvantily príslušného rozdelenia, t. j. $g_1 = F^{-1}(\alpha_1)$, $g_2 = F^{-1}(1 - \alpha_2)$.

4. Sčítaním rovníc $P(g \leqq g_1) = \alpha_1$, $P(g_2 \leqq g) = \alpha_2$ a následnými ekvivalentnými úpravami dostaneme vhodný interval spoľahlivosti.

Interval spoľahlivosti na odhad strednej hodnoty μ, ak poznáme rozptyl σ^2

V prípade obojstranného intervalu spoľahlivosti budeme postupovať nasledovne:

1. Bodový odhad strednej hodnoty μ je výberový aritmetický priemer (71), resp. (72).

2. Vhodná štatistika je
$$g(x) = \frac{\overline{x} - \mu}{\sigma} \sqrt{n} \sim N(0, 1).$$

3. Čísla $g_1 = u_{\frac{\alpha}{2}}$, $g_2 = u_{1-\frac{\alpha}{2}}$ sú kvantily normovaného normálneho rozdelenia (Tabuľka 28).

4. Keďže pre obojstranný interval spoľahlivosti má platiť $P(g_1 \leqq g \leqq g_2) = 1 - \alpha$, dostávame
$$u_{\frac{\alpha}{2}} \leqq \frac{\overline{x} - \mu}{\sigma} \sqrt{n} \leqq u_{1-\frac{\alpha}{2}}.$$
Ekvivalentnými úpravami spolu s faktom, že normálne rozdelenie je symetrické, t. j. platí $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$, získame nasledujúci $100(1 - \alpha)\%$–ný obojstranný interval spoľahlivosti na odhad strednej hodnoty μ, ak poznáme rozptyl σ^2
$$\mu \in \left\langle \overline{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \overline{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\rangle. \tag{78}$$

Podobne postupujeme pri určovaní ľavostranného a pravostranného intervalu spoľahlivosti s pravdepodobnosťou $1 - \alpha$. Teda $100(1 - \alpha)\%$–ný ľavostranný interval spoľahlivosti pre strednú hodnotu μ, ak poznáme rozptyl σ^2, má tvar
$$\mu \in \left\langle \overline{x} - u_{1-\alpha} \frac{\sigma}{\sqrt{n}}; \infty \right) \tag{79}$$

a $100(1-\alpha)\%$–ný pravostranný interval spoľahlivosti pre strednú hodnotu μ, ak poznáme rozptyl σ^2, má tvar

$$\mu \in \left(-\infty; \overline{x} + u_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right\rangle. \tag{80}$$

Príklad 4.1 *Skúmajme náhodnú premennú s rozptylom σ^2=7,4, ktorá pochádza zo základného súboru s normálnym rozdelením. Získané hodnoty náhodného výberu sú dané v nasledujúcej tabuľke:*

x_j	9	11	12	14	15	16	17	18	20	21
n_j	1	2	3	4	7	5	4	3	2	1

Určme

a) 99%–ný obojstranný interval spoľahlivosti pre neznámu strednú hodnotu,

b) 90%–ný ľavostranný interval spoľahlivosti pre neznámu strednú hodnotu,

c) 95%–ný pravostranný interval spoľahlivosti pre neznámu strednú hodnotu.

Riešenie: Na základe zadania je zrejmé, že ideme určovať intervaly spoľahlivosti na „odhad strednej hodnoty, ak poznáme rozptyl“. Rozsah súboru je $n = 32$. Najprv potrebujeme vhodný bodový odhad strednej hodnoty μ, ktorým je výberový aritmetický priemer $\overline{x}$ vypočítaný podľa vzťahu (72)

$$\overline{x} = \frac{1}{32}(9\cdot 1 + 11\cdot 2 + 12\cdot 3 + 14\cdot 4 + 15\cdot 7 + 16\cdot 5 + 17\cdot 4 + 18\cdot 3 + 20\cdot 2 + 21\cdot 1) = 15,344.$$

a) K určeniu obojstranného intervalu spoľahlivosti použijeme vzťah (78). V ňom je potrebné určiť $\alpha = 0,01$, $\sigma = \sqrt{7,4} = 2,72$ a kvantil normálneho rozdelenia $u_{1-\frac{\alpha}{2}} = u_{0,995} = 2,576$ (Tabuľka 28). Dosadením do vzťahu (78) dostaneme 99%–ný obojstranný interval spoľahlivosti pre neznámu strednú hodnotu μ

$$\mu \in \left\langle 15,544 - 2,576\cdot\frac{2,72}{\sqrt{32}}; 15,544 + 2,576\cdot\frac{2,72}{\sqrt{32}}\right\rangle \Rightarrow \mu \in \langle 14,11; 16,58\rangle.$$

b) K určeniu ľavostranného intervalu spoľahlivosti použijeme vzťah (79). V ňom je potrebné určiť $\alpha = 0,1$, $\sigma = \sqrt{7,4} = 2,72$ a kvantil normálneho rozdelenia $u_{1-\alpha} = u_{0,9} = 1,282$ (Tabuľka 28). Dosadením do vzťahu (79) dostaneme 90%–ný ľavostranný interval spoľahlivosti pre neznámu strednú hodnotu μ

$$\mu \in \left\langle 15,544 - 1,282\cdot\frac{2,72}{\sqrt{32}}; \infty\right) \Rightarrow \mu \in \langle 14,727; \infty).$$

c) K určeniu pravostranného intervalu spoľahlivosti použijeme vzťah (80). V ňom je potrebné určiť $\alpha = 0,05$, $\sigma = \sqrt{7,4} = 2,72$ a kvantil normálneho rozdelenia $u_{1-\alpha} = u_{0,95} = 1,645$ (Tabuľka 28). Dosadením do vzťahu (80) dostaneme 95%–ný pravostranný interval spoľahlivosti pre neznámu strednú hodnotu μ

$$\mu \in \left(-\infty; 15,544 + 1,645 \cdot \frac{2,72}{\sqrt{32}}\right\rangle \Rightarrow \mu \in (-\infty; 16,135\rangle\,.$$

Interval spoľahlivosti na odhad strednej hodnoty μ, ak nepoznáme rozptyl σ^2

V prípade obojstranného intervalu spoľahlivosti, ak rozsah náhodného výberu je malý ($n < 30$), budeme postupovať nasledovne:

1. Bodový odhad strednej hodnoty μ je výberový aritmetický priemer (71), resp. (72).
2. Vhodná štatistika je
$$g(x) = \frac{\overline{x} - \mu}{s}\sqrt{n} \sim t(n-1).$$
3. Čísla $g_1 = t_{\frac{\alpha}{2}}(n-1)$, $g_2 = t_{1-\frac{\alpha}{2}}(n-1)$ sú kvantily Studentovho t–rozdelenia (Tabuľka 29).
4. Keďže pre obojstranný interval spoľahlivosti má platiť $P(g_1 \leqq g \leqq g_2) = 1 - \alpha$, dostávame
$$t_{\frac{\alpha}{2}}(n-1) \leqq \frac{\overline{x} - \mu}{s} \cdot \sqrt{n} \leqq t_{1-\frac{\alpha}{2}}(n-1).$$
Ekvivalentnými úpravami spolu s faktom, že Studentovo t–rozdelenie je symetrické, t. j. platí $t_{\frac{\alpha}{2}}(n-1) = -t_{1-\frac{\alpha}{2}}(n-1)$, získame nasledujúci $100(1-\alpha)$%–ný obojstranný interval spoľahlivosti na odhad strednej hodnoty μ, ak nepoznáme rozptyl σ^2
$$\mu \in \left\langle \overline{x} - t_{1-\frac{\alpha}{2}}(n-1)\,\frac{s}{\sqrt{n}}; \overline{x} + t_{1-\frac{\alpha}{2}}(n-1)\,\frac{s}{\sqrt{n}} \right\rangle. \tag{81}$$

Podobne postupujeme pri určovaní ľavostranného a pravostranného intervalu spoľahlivosti s pravdepodobnosťou $1 - \alpha$. Teda $100(1-\alpha)$%–ný ľavostranný interval spoľahlivosti pre strednú hodnotu μ, ak nepoznáme rozptyl σ^2, má tvar

$$\mu \in \left\langle \overline{x} - t_{1-\alpha}(n-1)\,\frac{s}{\sqrt{n}}; \infty \right) \tag{82}$$

a $100(1-\alpha)$%–ný pravostranný interval spoľahlivosti pre strednú hodnotu μ, ak nepoznáme rozptyl σ^2, má tvar

$$\mu \in \left(-\infty; \overline{x} + t_{1-\alpha}(n-1)\,\frac{s}{\sqrt{n}}\right\rangle. \tag{83}$$

V prípade, že rozsah náhodného výberu je dostatočne veľký ($n \geqq 30$), budeme postupovať rovnako, pričom štatistika $g(x) = \frac{\overline{x}-\mu}{s}\sqrt{n}$ bude mať normálne rozdelenie $N(0,1)$ ako je to v prvom prípade. Potom $100(1-\alpha)\%$–ný obojstranný interval spoľahlivosti na odhad strednej hodnoty μ, ak nepoznáme rozptyl σ^2 má tvar

$$\mu \in \left\langle \overline{x} - u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \overline{x} + u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right\rangle. \tag{84}$$

$100(1-\alpha)\%$–ný ľavostranný interval spoľahlivosti pre strednú hodnotu μ, ak nepoznáme rozptyl σ^2, má tvar

$$\mu \in \left\langle \overline{x} - u_{1-\alpha} \frac{s}{\sqrt{n}}; \infty \right) \tag{85}$$

a $100(1-\alpha)\%$–ný pravostranný interval spoľahlivosti pre strednú hodnotu μ, ak nepoznáme rozptyl σ^2, má tvar

$$\mu \in \left(-\infty; \overline{x} + u_{1-\alpha} \frac{s}{\sqrt{n}} \right\rangle. \tag{86}$$

Príklad 4.2 *Skúmajme náhodnú premennú, ktorá pochádza zo základného súboru s normálnym rozdelením. Získané hodnoty náhodného výberu sú dané v nasledujúcej tabuľke:*

x_j	9	11	12	14	15	16	17	18	20	21
n_j	1	2	3	3	6	4	4	3	2	1

Určme

a) 90%–ný obojstranný interval spoľahlivosti pre neznámu strednú hodnotu,

b) 99%–ný ľavostranný interval spoľahlivosti pre neznámu strednú hodnotu,

c) 95%–ný pravostranný interval spoľahlivosti pre neznámu strednú hodnotu.

Riešenie: Na základe zadania je zrejmé, že ideme určovať intervaly spoľahlivosti na „odhad strednej hodnoty, ak nepoznáme rozptyl“. Rozsah súboru je $n = 29$. Najprv potrebujeme vhodný bodový odhad strednej hodnoty μ, ktorým je výberový aritmetický priemer $\overline{x}$ vypočítaný podľa vzťahu (72)

$$\overline{x} = \frac{1}{29}(9\cdot 1 + 11\cdot 2 + 12\cdot 3 + 14\cdot 3 + 15\cdot 6 + 16\cdot 4 + 17\cdot 4 + 18\cdot 3 + 20\cdot 2 + 21\cdot 1) = 15,379.$$

Keďže nepoznáme rozptyl σ^2, budeme potrebovať výberový rozptyl s^2 a výberovú smerodajnú odchýlku s, ktoré vypočítame podľa vzťahov (74) a (77)

$$s^2 = \frac{1}{28}\left[1\cdot(9-15,379)^2 + 2\cdot(11-15,379)^2 + 3\cdot(12-15,379)^2 + 3\cdot(14-15,379)^2 + \right.$$

$$+6\cdot(15-15,379)^2 + 4\cdot(16-15,379)^2 + 4\cdot(17-15,379)^2 + 3\cdot(18-15,379)^2 +$$

$$+2\cdot(20-15,379)^2+1\cdot(21-15,379)^2\big] = 8,101 \Rightarrow s=\sqrt{s^2}=2,846.$$

Pretože rozsah súboru je $n < 30$, k určeniu intervalov spoľahlivosti použijeme vzťahy (81), (82) a (83).

a) V prípade obojstranného intervalu spoľahlivosti potrebujeme určiť $\alpha = 0,1$ a kvantil Studentovho t–rozdelenia $t_{1-\frac{\alpha}{2}}(n-1) = t_{0,95}(28) = 1,701$ (Tabuľka 29). Dosadením do vzťahu (81) dostaneme 90%–ný obojstranný interval spoľahlivosti pre neznámu strednú hodnotu μ

$$\mu \in \left\langle 15,379 - 1,701\cdot\frac{2,846}{\sqrt{29}}; 15,379 + 1,701\cdot\frac{2,846}{\sqrt{29}}\right\rangle \Rightarrow \mu \in \langle 14,480; 16,278\rangle .$$

b) V prípade ľavostranného intervalu spoľahlivosti potrebujeme určiť $\alpha = 0,01$ a kvantil Studentovho t–rozdelenia $t_{1-\alpha}(n-1) = t_{0,99}(28) = 2,467$ (Tabuľka 29). Dosadením do vzťahu (82) dostaneme 99%–ný ľavostranný interval spoľahlivosti pre neznámu strednú hodnotu μ

$$\mu \in \left\langle 15,379 - 2,467\cdot\frac{2,846}{\sqrt{29}}; \infty\right) \Rightarrow \mu \in \langle 14,075; \infty) .$$

c) V prípade pravostranného intervalu spoľahlivosti potrebujeme určiť $\alpha = 0,05$ a kvantil Studentovho t–rozdelenia $t_{1-\alpha}(n-1) = t_{0,95}(28) = 1,701$ (Tabuľka 29). Dosadením do vzťahu (83) dostaneme 95%–ný pravostranný interval spoľahlivosti pre neznámu strednú hodnotu μ

$$\mu \in \left(-\infty; 15,379 + 1,701\cdot\frac{2,846}{\sqrt{29}}\right\rangle \Rightarrow \mu \in (-\infty; 16,278\rangle .$$

Interval spoľahlivosti na odhad rozptylu σ^2, ak poznáme strednú hodnotu μ

V prípade obojstranného intervalu spoľahlivosti budeme postupovať nasledovne:

1. Bodový odhad rozptylu σ^2 je výberový rozptyl s_0^2 (75), resp. (76).

2. Vhodná štatistika je
$$g(x) = \frac{n\, s_0^2}{\sigma^2} \sim \chi^2(n).$$

3. Čísla $g_1 = \chi^2_{\frac{\alpha}{2}}(n)$, $g_2 = \chi^2_{1-\frac{\alpha}{2}}(n)$ sú kvantily χ^2–rozdelenia (Tabuľky 30 – 32).

4. Keďže pre obojstranný interval spoľahlivosti má platiť $P(g_1 \leqq g \leqq g_2) = 1-\alpha$, dostávame
$$\chi^2_{\frac{\alpha}{2}}(n) \leqq \frac{n\, s_0^2}{\sigma^2} \leqq \chi^2_{1-\frac{\alpha}{2}}(n).$$

Ekvivalentnými úpravami získame nasledujúci $100(1-\alpha)\%$–ný obojstranný interval spoľahlivosti na odhad rozptylu σ^2, ak poznáme strednú hodnotu μ

$$\sigma^2 \in \left\langle \frac{n\, s_0^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}; \frac{n\, s_0^2}{\chi^2_{\frac{\alpha}{2}}(n)} \right\rangle . \tag{87}$$

Podobne postupujeme pri určovaní ľavostranného a pravostranného intervalu spoľahlivosti s pravdepodobnosťou $1-\alpha$. Teda $100(1-\alpha)\%$–ný ľavostranný interval spoľahlivosti pre rozptyl σ^2, ak poznáme strednú hodnotu μ, má tvar

$$\sigma^2 \in \left\langle \frac{n\, s_0^2}{\chi^2_{1-\alpha}(n)}; \infty \right) \tag{88}$$

a $100(1-\alpha)\%$–ný pravostranný interval spoľahlivosti pre rozptyl σ^2, ak poznáme strednú hodnotu μ, má tvar

$$\sigma^2 \in \left(0; \frac{n\, s_0^2}{\chi^2_{\alpha}(n)} \right\rangle . \tag{89}$$

Príklad 4.3 *Skúmajme náhodnú premennú, ktorá pochádza zo základného súboru s normálnym rozdelením a strednou hodnotou μ=15. Získané hodnoty náhodného výberu sú dané v nasledujúcej tabuľke:*

x_j	9	11	12	14	15	16	17	18	20	21
n_j	1	2	3	4	7	5	4	3	2	1

Určme

a) 99%–ný obojstranný interval spoľahlivosti pre rozptyl,

b) 95%–ný ľavostranný interval spoľahlivosti pre rozptyl,

c) 95%–ný pravostranný interval spoľahlivosti pre rozptyl.

Riešenie: Na základe zadania je zrejmé, že ideme určovať intervaly spoľahlivosti na „odhad rozptylu, ak poznáme strednú hodnotu“. Rozsah súboru je $n = 32$. Najprv potrebujeme vhodný bodový odhad rozptylu σ^2, ktorým je výberový rozptyl s_0^2. Keďže poznáme strednú hodnotu μ, použijeme vzťah (76)

$$s_0^2 = \frac{1}{32}\big[1\cdot(9-15)^2+2\cdot(11-15)^2+3\cdot(12-15)^2+4\cdot(14-15)^2+7\cdot(15-15)^2+5\cdot(16-15)^2+$$

$$+4\cdot(17-15)^2+3\cdot(18-15)^2+2\cdot(20-15)^2+1\cdot(21-15)^2\big] = 7,281.$$

a) K určeniu obojstranného intervalu spoľahlivosti použijeme vzťah (87). V ňom je potrebné určiť $\alpha = 0,01$ a kvantily χ^2–rozdelenia $\chi^2_{\frac{\alpha}{2}}(n) = \chi^2_{0,005}(32) = 15,13$; $\chi^2_{1-\frac{\alpha}{2}}(n) =$

$\chi^2_{0,995}(32) = 56,33$ (Tabuľka 30). Dosadením do vzťahu (87) dostaneme 99%–ný obojstranný interval spoľahlivosti pre rozptyl σ^2

$$\sigma^2 \in \left\langle \frac{32 \cdot 7,281}{56,33}; \frac{32 \cdot 7,281}{15,13} \right\rangle \Rightarrow \sigma^2 \in \langle 4,136; 15,396 \rangle .$$

b) K určeniu ľavostranného intervalu spoľahlivosti použijeme vzťah (88). V ňom je potrebné určiť $\alpha = 0,05$ a kvantil χ^2–rozdelenia $\chi^2_{1-\alpha}(n) = \chi^2_{0,95}(32) = 46,19$ (Tabuľka 30). Dosadením do vzťahu (88) dostaneme 95%–ný ľavostranný interval spoľahlivosti pre rozptyl σ^2

$$\sigma^2 \in \left\langle \frac{32 \cdot 7,281}{46,19}; \infty \right) \Rightarrow \sigma^2 \in \langle 5,044; \infty) .$$

c) K určeniu pravostranného intervalu spoľahlivosti použijeme vzťah (89). V ňom je potrebné určiť $\alpha = 0,05$ a kvantil χ^2–rozdelenia $\chi^2_{\alpha}(n) = \chi^2_{0,05}(32) = 20,07$ (Tabuľka 30). Dosadením do vzťahu (89) dostaneme 95%–ný pravostranný interval spoľahlivosti pre rozptyl σ^2

$$\sigma^2 \in \left(0; \frac{32 \cdot 7,281}{20,07} \right\rangle \Rightarrow \sigma^2 \in (0; 11,608\rangle .$$

Interval spoľahlivosti na odhad rozptylu σ^2, ak nepoznáme strednú hodnotu μ

V prípade obojstranného intervalu spoľahlivosti budeme postupovať nasledovne:

1. Bodový odhad rozptylu σ^2 je výberový rozptyl s^2 (73), resp. (74).

2. Vhodná štatistika je
$$g(x) = \frac{(n-1)\, s^2}{\sigma^2} \sim \chi^2(n-1).$$

3. Čísla $g_1 = \chi^2_{\frac{\alpha}{2}}(n-1)$, $g_2 = \chi^2_{1-\frac{\alpha}{2}}(n-1)$ sú kvantily χ^2–rozdelenia (Tabuľky 30 – 32).

4. Keďže pre obojstranný interval spoľahlivosti má platiť $P(g_1 \leqq g \leqq g_2) = 1 - \alpha$, dostávame
$$\chi^2_{\frac{\alpha}{2}}(n-1) \leqq \frac{(n-1)\, s^2}{\sigma^2} \leqq \chi^2_{1-\frac{\alpha}{2}}(n-1).$$

   Ekvivalentnými úpravami získame nasledujúci $100(1-\alpha)$%–ný obojstranný interval spoľahlivosti na odhad rozptylu σ^2, ak nepoznáme strednú hodnotu μ
$$\sigma^2 \in \left\langle \frac{(n-1)\, s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}; \frac{(n-1)\, s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right\rangle . \tag{90}$$

Podobne postupujeme pri určovaní ľavostranného a pravostranného intervalu spoľahlivosti s pravdepodobnosťou $1-\alpha$. Teda $100(1-\alpha)\%$–ný ľavostranný interval spoľahlivosti pre rozptyl σ^2, ak nepoznáme strednú hodnotu μ, má tvar

$$\sigma^2 \in \left\langle \frac{(n-1)\,s^2}{\chi^2_{1-\alpha}(n-1)}; \infty \right) \tag{91}$$

a $100(1-\alpha)\%$–ný pravostranný interval spoľahlivosti pre rozptyl σ^2, ak nepoznáme strednú hodnotu μ, má tvar

$$\sigma^2 \in \left(0; \frac{(n-1)\,s^2}{\chi^2_{\alpha}(n-1)} \right\rangle. \tag{92}$$

Príklad 4.4 *Skúmajme náhodnú premennú, ktorá pochádza zo základného súboru s normálnym rozdelením. Získané hodnoty náhodného výberu sú dané v nasledujúcej tabuľke:*

x_j	9	11	12	14	15	16	17	18	20	21
n_j	1	2	3	4	7	5	4	3	2	1

Určme

a) 99%–ný obojstranný interval spoľahlivosti pre rozptyl,

b) 95%–ný ľavostranný interval spoľahlivosti pre rozptyl,

c) 95%–ný pravostranný interval spoľahlivosti pre rozptyl.

Riešenie: Na základe zadania je zrejmé, že ideme určovať intervaly spoľahlivosti na „odhad rozptylu, ak nepoznáme strednú hodnotu“. Rozsah súboru je $n = 32$. Najprv potrebujeme vhodný bodový odhad rozptylu σ^2, ktorým je výberový rozptyl s^2. Keďže nepoznáme strednú hodnotu μ, použijeme vzťah (74), kde výberový aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (72)

$$\overline{x} = \frac{1}{32}(9 \cdot 1 + 11 \cdot 2 + 12 \cdot 3 + 14 \cdot 4 + 15.7 + 16 \cdot 5 + 17 \cdot 4 + 18 \cdot 3 + 20 \cdot 2 + 21 \cdot 1) = 15,344.$$

Potom pre výberový rozptyl s^2 na základe vzťahu (74) platí

$$s^2 = \frac{1}{31}\big[1 \cdot (9 - 15,344)^2 + 2 \cdot (11 - 15,344)^2 + 3 \cdot (12 - 15,344)^2 + 4 \cdot (14 - 15,344)^2 +$$

$$+7 \cdot (15 - 15,344)^2 + 5 \cdot (16 - 15,344)^2 + 4 \cdot (17 - 15,344)^2 + 3 \cdot (18 - 15,344)^2 +$$

$$+2 \cdot (20 - 15,344)^2 + 1 \cdot (21 - 15,344)^2\big] = 7,394.$$

a) K určeniu obojstranného intervalu spoľahlivosti použijeme vzťah (90). V ňom je potrebné určiť $\alpha = 0,01$ a kvantily χ^2–rozdelenia $\chi^2_{\frac{\alpha}{2}}(n-1) = \chi^2_{0,005}(31) = 14,46$; $\chi^2_{1-\frac{\alpha}{2}}(n-1) =$

$\chi^2_{0,995}(31) = 55,00$ (Tabuľka 30). Dosadením do vzťahu (90) dostaneme 99%–ný obojstranný interval spoľahlivosti pre rozptyl σ^2

$$\sigma^2 \in \left\langle \frac{31 \cdot 7,394}{55,00}; \frac{31 \cdot 7,394}{14,46} \right\rangle \Rightarrow \sigma^2 \in \langle 4,1674; 15,8544 \rangle .$$

b) K určeniu ľavostranného intervalu spoľahlivosti použijeme vzťah (91). V ňom je potrebné určiť $\alpha = 0,05$ a kvantil χ^2–rozdelenia $\chi^2_{1-\alpha}(n-1) = \chi^2_{0,95}(31) = 44,99$ (Tabuľka 30). Dosadením do vzťahu (91) dostaneme 95%–ný ľavostranný interval spoľahlivosti pre rozptyl σ^2

$$\sigma^2 \in \left\langle \frac{31 \cdot 7,394}{44,99}; \infty \right) \Rightarrow \sigma^2 \in \langle 5,0954; \infty) .$$

c) K určeniu pravostranného intervalu spoľahlivosti použijeme vzťah (92). V ňom je potrebné určiť $\alpha = 0,05$ a kvantil χ^2–rozdelenia $\chi^2_{\alpha}(n-1) = \chi^2_{0,05}(31) = 19,28$ (Tabuľka 30). Dosadením do vzťahu (92) dostaneme 95%–ný pravostranný interval spoľahlivosti pre rozptyl σ^2

$$\sigma^2 \in \left(0; \frac{31 \cdot 7,394}{19,28} \right\rangle \Rightarrow \sigma^2 \in (0; 11,8886\rangle .$$

4.2.2 Interval spoľahlivosti na odhad parametra exponenciálneho rozdelenia

Uvažujme náhodný výber V_n, ktorý pochádza zo základného súboru s exponenciálnym rozdelením s parametrom λ, $\lambda > 0$. Zapisujeme $V_n \sim \mathrm{Ex}(\lambda)$. V prípade obojstranného intervalu spoľahlivosti budeme postupovať nasledovne:

1. Bodový odhad parametra λ je výberový aritmetický priemer (71), resp. (72).

2. Vhodná štatistika je
$$g(x) = \frac{2n\overline{x}}{\lambda} \sim \chi^2(2n).$$

3. Čísla $g_1 = \chi^2_{\frac{\alpha}{2}}(2n)$, $g_2 = \chi^2_{1-\frac{\alpha}{2}}(2n)$ sú kvantily χ^2–rozdelenia (Tabuľky 30 – 32).

4. Keďže pre obojstranný interval spoľahlivosti má platiť $P(g_1 \leqq g \leqq g_2) = 1 - \alpha$, dostávame
$$\chi^2_{\frac{\alpha}{2}}(2n) \leqq \frac{2n\overline{x}}{\lambda} \leqq \chi^2_{1-\frac{\alpha}{2}}(2n).$$
Ekvivalentnými úpravami získame nasledujúci $100(1-\alpha)$%–ný obojstranný interval spoľahlivosti pre parameter λ
$$\lambda \in \left\langle \frac{2n\overline{x}}{\chi^2_{1-\frac{\alpha}{2}}(2n)}; \frac{2n\overline{x}}{\chi^2_{\frac{\alpha}{2}}(2n)} \right\rangle . \tag{93}$$

Podobne postupujeme pri určovaní ľavostranného a pravostranného intervalu spoľahlivosti s pravdepodobnosťou $1-\alpha$. Teda $100(1-\alpha)\%$–ný ľavostranný interval spoľahlivosti pre parameter λ má tvar

$$\lambda \in \left\langle \frac{2n\overline{x}}{\chi^2_{1-\alpha}(2n)}; \infty \right) \tag{94}$$

a $100(1-\alpha)\%$–ný pravostranný interval spoľahlivosti pre parameter λ má tvar

$$\lambda \in \left(0; \frac{2n\overline{x}}{\chi^2_{\alpha}(2n)} \right\rangle. \tag{95}$$

Príklad 4.5 *U 100 náhodne vybraných cestujúcich bola zisťovaná doba ich čakania na autobus [v min.]. Výsledky sú v nasledujúcej tabuľke:*

x_j	0,5	2,5	4,5	6,5	8,5
n_j	60	20	11	5	4

Predpokladáme, že doba čakania má exponenciálne rozdelenie s parametrom λ. Určme

a) 95%–ný obojstranný interval spoľahlivosti pre parameter λ,

b) 99%–ný ľavostranný interval spoľahlivosti pre parameter λ,

c) 99%–ný pravostranný interval spoľahlivosti pre parameter λ.

Riešenie: Na základe zadania je zrejmé, že ideme určovať interval spoľahlivosti na „odhad parametra exponenciálneho rozdelenia“. Najprv potrebujeme vhodný bodový odhad strednej hodnoty μ, ktorým je výberový aritmetický priemer $\overline{x}$ vypočítaný podľa vzťahu (72)

$$\overline{x} = \frac{1}{100}\big(0,5\cdot 60 + 2,5\cdot 20 + 4,5\cdot 11 + 6,5\cdot 5 + 8,5\cdot 4\big) = 1,96,$$

teda bodový odhad parametra λ je $1,96$.

a) K určeniu obojstranného intervalu spoľahlivosti použijeme vzťah (93). V ňom je potrebné určiť $\alpha = 0,05$ a kvantily χ^2–rozdelenia $\chi^2_{\frac{\alpha}{2}}(2n) = \chi^2_{0,025}(200) = 162,73$, $\chi^2_{1-\frac{\alpha}{2}}(2n) = \chi^2_{0,975}(200) = 241,06$ (Tabuľka 32). Dosadením do vzťahu (93) dostaneme 95%–ný obojstranný interval spoľahlivosti pre parameter λ

$$\lambda \in \left\langle \frac{200\cdot 1,96}{241,06}; \frac{200\cdot 1,96}{162,73} \right\rangle \Rightarrow \langle 1,626; 2,409\rangle .$$

b) K určeniu ľavostranného intervalu spoľahlivosti použijeme vzťah (94). V ňom je potrebné určiť $\alpha = 0,01$ a kvantil χ^2–rozdelenia $\chi^2_{1-\alpha}(2n) = \chi^2_{0,99}(200) = 249,45$ (Tabuľka 32). Dosadením do vzťahu (94) dostaneme 99%–ný ľavostranný interval spoľahlivosti pre parameter λ

$$\lambda \in \left\langle \frac{200\cdot 1,96}{249,45}; \infty \right) \Rightarrow \lambda \in \langle 1,572; \infty) .$$

c) K určeniu pravostranného intervalu spoľahlivosti použijeme vzťah (95). V ňom je potrebné určiť $\alpha = 0,01$ a kvantil χ^2–rozdelenia $\chi^2_\alpha(2n) = \chi^2_{0,99}(200) = 156,43$ (Tabuľka 32). Dosadením do vzťahu (95) dostaneme 99%–ný pravostranný interval spoľahlivosti pre parameter λ

$$\lambda \in \left(0; \frac{200 \cdot 1,96}{156,43}\right\rangle \Rightarrow \lambda \in (0; 2,506\rangle\,.$$

5 Testovanie štatistických hypotéz

V praxi je testovanie štatistických hypotéz jednou z najdôležitejších štatistických metód. *Štatistická hypotéza* je predpoklad (tvrdenie), týkajúci sa náhodného výberu V_n, resp. rozdelenia, z ktorého náhodný výber pochádza. *Testovanie štatistickej hypotézy* je postup, pomocou ktorého overujeme platnosť (správnosť) danej hypotézy (predpokladu). *Nulová hypotéza* H_0 je tvrdenie, ktorého pravdivosť overujeme. V tom najjednoduchšom prípade má tvar $H_0 : Q = Q_0$ resp. $H_0 : Q - Q_0 = 0$, pričom Q je parameter základného súboru a Q_0 je konkrétna konštanta. Proti nulovej hypotéze staviame *alternatívnu hypotézu* H_1. V praxi má alternatívna hypotéza tri tvary:

- *Obojstranná alternatíva:* $H_1 : Q \neq Q_0$, ktorá popiera platnosť nulovej hypotézy.
- *Ľavostranná alternatíva:* $H_1 : Q < Q_0$, ktorá vymedzuje obor hodnôt parametra Q naľavo od hodnoty Q_0.
- *Pravostranná alternatíva:* $H_1 : Q > Q_0$, ktorá vymedzuje obor hodnôt parametra Q napravo od hodnoty Q_0.

Pri testovaní štatistických hypotéz, t. j. pri rozhodovaní o nezamietnutí alebo zamietnutí nulovej hypotézy H_0 v prospech alternatívnej hypotézy H_1, sa dopúšťame dvoch druhov chýb (Tabuľka 5):

- *Chyba prvého druhu* nastáva vtedy, ak zamietneme nulovú hypotézu H_0, aj keď je správna. Pravdepodobnosť výskytu tejto chyby označujeme α a nazývame ju *hladina významnosti* testu.
- *Chyba druhého druhu* nastáva vtedy, ak nezamietneme nulovú hypotézu H_0, hoci je nesprávna. Pravdepodobnosť výskytu tejto chyby označujeme β a hodnotu $1 - \beta$ nazývame *sila testu*. Sila testu predstavuje vlastne pravdepodobnosť zamietnutia nulovej hypotézy.

Tabuľka 5: Chyby pri testovaní štatistických hypotéz

	Skutočnosť	
Záver	Platí H_0	Neplatí H_0
Zamietame H_0	Chyba prvého druhu	Správne rozhodnutie
Nezamietame H_0	Správne rozhodnutie	Chyba druhého druhu

Najideálnejšie by bolo, keby α aj β boli minimálne súčasne. Keďže sa to nedá dosiahnuť, tak sa v praxi pri pevnom α (najčastejšie sa volí $0,05$; $0,01$; $0,1$) minimalizuje β.

Testovacie kritérium (testovacia charakteristika, štatistika) je hodnota vypočítaná z výberového súboru, ktorá nám pomáha rozhodovať o nezamietnutí, resp. zamietnutí nulovej hypotézy. Obor (oblasť) zamietnutia nulovej hypotézy nazývame *kritický obor (oblasť)* K_α a je závislý od zvolenej hladiny významnosti α. Ak hodnota testovacieho kritéria padne do kritickej oblasti, tak nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. V opačnom prípade nulovú hypotézu H_0 nezamietame.

S rozvojom výpočtovej techniky sa pri rozhodovaní o nezamietnutí alebo zamietnutí nulovej hypotézy H_0 používa p–hodnota (z ang. p–value). *P–hodnotou* (p) nazývame najmenšiu možnú hladinu významnosti α, na ktorej je ešte možné pri daných nameraných hodnotách náhodného výberu zamietnuť nulovú hypotézu H_0. Platí:

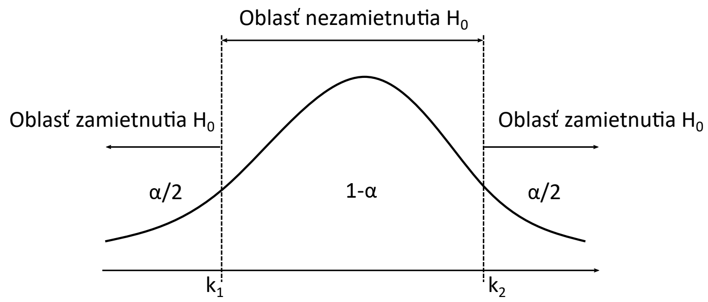
- ak $p < \alpha$, tak nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1 na hladine významnosti α,
- ak $p \geqq \alpha$, tak nulovú hypotézu H_0 nezamietame na hladine významnosti α.

Postup pri testovaní hypotéz je nasledovný:

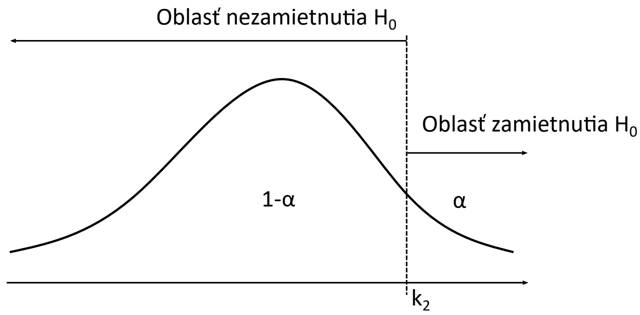
1. Formulácia nulovej hypotézy H_0 a alternatívnej hypotézy H_1.
2. Voľba hladiny významnosti α.
3. Výpočet hodnoty testovacieho kritéria na základe daných hodnôt náhodného výberu.
4. Určenie kritickej oblasti K_α.
5. Vyhodnotenie testu, ktoré spočíva buď v zamietnutí nulovej hypotézy H_0 a prijatí alternatívnej hypotézy H_1, alebo v nezamietnutí nulovej hypotézy H_0.

Oblasť nezamietnutia a zamietnutia nulovej hypotézy H_0 pri obojstrannom teste je znázornená na Obr. 19, pri pravostrannom teste je znázornená na Obr. 20 a pri ľavostrannom teste je znázornená na Obr. 21. Hodnoty k_1 a k_2, ktoré oddeľujú oblasť nezamietnutia nulovej hypotézy H_0 od oblasti zamietnutia nulovej hypotézy H_0, nazývame *kritické hodnoty.*

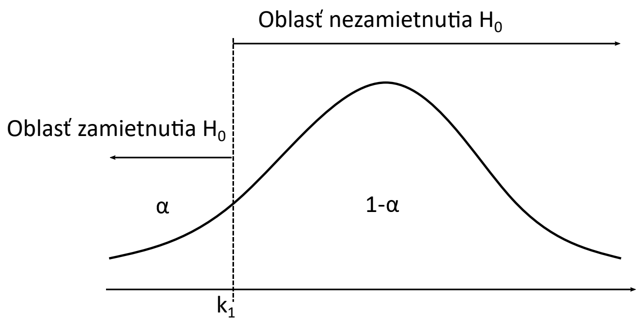
Existuje veľa štatistických testov, ktoré vieme rozdeliť podľa rôznych kritérií. Základným delením štatistických testov je delenie na parametrické a neparametrické testy. *Parametrické testy* testujú neznámy parameter rozdelenia, z ktorého náhodný výber V_n pochádza. *Neparametrické testy* testujú typ rozdelenia alebo niektorú jeho vlastnosť. Medzi najčastejšie používané neparametrické testy patria *testy symetrie* (napr. znamienkový test, Wilcoxonov



Obr. 19: Oblasť nezamietnutia a zamietnutia nulovej hypotézy pri obojstrannom teste



Obr. 20: Oblasť nezamietnutia a zamietnutia nulovej hypotézy pri pravostrannom teste



Obr. 21: Oblasť nezamietnutia a zamietnutia nulovej hypotézy pri ľavostrannom teste

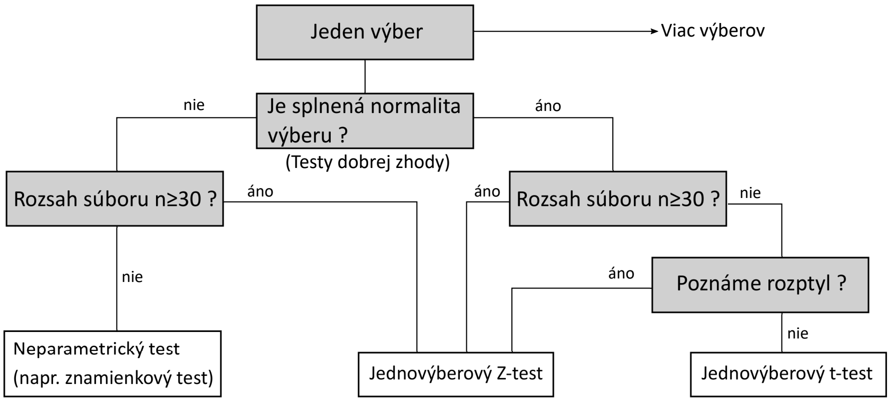
dvojvýberový test), *testy dobrej zhody* (napr. Pearsonov test, Kolmogorovov–Smirnovov test, Andersonov–Darlingov test, Shapiro–Wilkov test normality) a *testy extrémnych hodnôt* (napr. Dixonov test, Grubbsov test). Všetky uvedené testy si zadefinujeme v nasledujúcich kapitolách.

5.1 Jednovýberové parametrické testy o parametroch normálneho rozdelenia

Uvažujme náhodný výber V_n, ktorý pochádza zo základného súboru s normálnym rozdelením so strednou hodnotou μ a smerodajnou odchýlkou σ. Zapisujeme $V_n \sim N(\mu, \sigma^2)$. V tejto kapitole sa budeme venovať nasledujúcim testom:

- test strednej hodnoty μ, ak
  - poznáme rozptyl σ^2 (jednovýberový Z–test),
  - nepoznáme rozptyl σ^2 (jednovýberový t–test),
- test rozptylu σ^2, ak
  - poznáme strednú hodnotu μ,
  - nepoznáme strednú hodnotu μ.

Všeobecný postup výberu testu strednej hodnoty pre jeden výber je zobrazený na Obr. 22.



Obr. 22: Postup výberu testu strednej hodnoty pre jeden výber

5.1.1 Jednovýberový test strednej hodnoty μ, pričom poznáme rozptyl σ^2 (jednovýberový Z–test)

1. Hypotézy:

$$H_0 : \mu = \mu_0 \quad (\mu_0 \text{ je známa konštanta}),$$
$$H_1 : \mu \neq \mu_0 \quad (\text{resp. } \mu < \mu_0, \text{ resp. } \mu > \mu_0).$$

2. Testovacie kritérium je

$$Z = \frac{\overline{x} - \mu_0}{\sigma} \sqrt{n} \sim N(0,1), \tag{96}$$

kde $\overline{x}$ je výberový aritmetický priemer a n je rozsah náhodného výberu.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; \infty\right)$
$\mu = \mu_0$	$\mu > \mu_0$	$(u_{1-\alpha}; \infty)$
$\mu = \mu_0$	$\mu < \mu_0$	$(-\infty; -u_{1-\alpha})$

Príklad 5.1 *Dĺžka vyrobenej súčiastky má mať hodnotu 10 cm. Náhodne bolo vybraných 9 súčiastok, ktorých dĺžky [v cm] boli 10,24; 10,12; 9,91; 10,19; 9,78; 10,14; 9,86; 10,17; 10,05. Predpokladáme, že náhodný výber bol realizovaný zo základného súboru s normálnym rozdelením, ktorého rozptyl je 0,0016.*

a) Na hladine významnosti α=0,05 otestujme, či dĺžka súčiastok sa významne líši od požadovanej dĺžky.

b) Na hladine významnosti α=0,01 otestujme, či dĺžka súčiastok je menšia ako požadovaná dĺžka.

c) Na hladine významnosti α=0,01 otestujme, či dĺžka súčiastok je väčšia ako požadovaná dĺžka.

Riešenie: Na testovanie použijeme jednovýberový Z–test, pretože poznáme rozptyl náhodného výberu. Rozsah súboru je $n = 9$ a smerodajná odchýlka $\sigma = \sqrt{0,0016} = 0,04$.

a) Budeme testovať

$$H_0 : \mu = 10,$$
$$H_1 : \mu \neq 10.$$

Nulová hypotéza hovorí, že dĺžky súčiastok sa významne nelíšia od požadovanej dĺžky. Proti alternatívnej hypotéze, ktorá hovorí, že dĺžky súčiastok sa významne líšia od požadovanej dĺžky. Hodnotu testovacieho kritéria Z určíme podľa vzťahu (96)

$$Z = \frac{\overline{x} - 10}{0,04} \cdot \sqrt{9},$$

kde výberový aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (71)

$$\overline{x} = \frac{1}{9}(10,24 + 10,12 + 9,91 + 10,19 + 9,78 + 10,14 + 9,86 + 10,17 + 10,05) = 10,0511.$$

Potom hodnota testovacieho kritéria je $Z = 3,8325$. Kvantil normovaného normálneho rozdelenia je $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$ (Tabuľka 28). Kritická oblasť je

$$K_{0,05} = (-\infty; -1,96) \cup (1,96; \infty) \Rightarrow Z \in K_{0,05}.$$

Hodnota testovacieho kritéria $Z = 3,8325$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti α=0,05 sa dĺžky súčiastok významne líšia od požadovanej hodnoty.

b) Budeme testovať

$$H_0 : \mu = 10,$$
$$H_1 : \mu < 10.$$

Nulová hypotéza hovorí, že dĺžky súčiastok sa významne nelíšia od požadovanej dĺžky. Proti alternatívnej hypotéze, ktorá hovorí, že dĺžky súčiastok sú menšie ako požadovaná dĺžka. Hodnotu testovacieho kritéria Z vypočítame rovnako ako v prípade a) a teda $Z = 3,8325$. Kvantil normovaného normálneho rozdelenia je $u_{1-\alpha} = u_{0,99} = 2,3264$ (Tabuľka 28). Kritická oblasť je

$$K_{0,01} = (-\infty; -2,3264) \Rightarrow Z \notin K_{0,01}.$$

Hodnota testovacieho kritéria $Z = 3,8325$ nepatrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti α=0,01 dĺžky súčiastok nie sú menšie ako požadovaná hodnota.

c) Budeme testovať

$$H_0 : \mu = 10,$$
$$H_1 : \mu > 10.$$

Nulová hypotéza hovorí, že dĺžky súčiastok sa významne nelíšia od požadovanej dĺžky. Proti alternatívnej hypotéze, ktorá hovorí, že dĺžky súčiastok sú väčšie ako požadovaná dĺžka. Hodnotu testovacieho kritéria Z vypočítame rovnako ako v prípade a) a teda $Z = 3,8325$. Kvantil normovaného normálneho rozdelenia je $u_{1-\alpha} = u_{0,99} = 2,3264$ (Tabuľka 28). Kritická oblasť je

$$K_{0,01} = (2,3264; \infty) \Rightarrow Z \in K_{0,01}.$$

Hodnota testovacieho kritéria $Z = 3,8325$ patrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti α=0,01 dĺžky súčiastok sú väčšie ako požadovaná hodnota.

5.1.2 Jednovýberový test strednej hodnoty μ, pričom nepoznáme rozptyl σ^2 (jednovýberový t–test)

1. Hypotézy:

$$H_0 : \mu = \mu_0 \quad (\mu_0 \text{ je známa konštanta}),$$
$$H_1 : \mu \neq \mu_0 \quad (\text{resp. } \mu < \mu_0, \text{ resp. } \mu > \mu_0).$$

2. Testovacie kritérium je

$$t = \frac{\overline{x} - \mu_0}{s} \sqrt{n} \sim t(n-1), \tag{97}$$

kde $\overline{x}$ je výberový aritmetický priemer, s je výberová smerodajná odchýlka a n je rozsah náhodného výberu.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\mu = \mu_0$	$\mu \neq \mu_0$	$\left(-\infty; -t_{1-\frac{\alpha}{2}}(n-1)\right) \cup \left(t_{1-\frac{\alpha}{2}}(n-1); \infty\right)$
$\mu = \mu_0$	$\mu > \mu_0$	$(t_{1-\alpha}(n-1); \infty)$
$\mu = \mu_0$	$\mu < \mu_0$	$(-\infty; -t_{1-\alpha}(n-1))$

Poznámka: V prípade, že rozsah náhodného výberu je dostatočne veľký ($n \geqq 30$), budeme postupovať rovnako, avšak Studentovo t–rozdelenie aproximujeme normovaným normálnym rozdelením ako je to pri predchádzajúcom teste (jednovýberový Z–test).

Príklad 5.2 *Dĺžka vyrobenej súčiastky má mať hodnotu 10 cm. Náhodne bolo vybraných 9 súčiastok, ktorých dĺžky [v cm] boli 10,24; 10,12; 9,91; 10,19; 9,78; 10,14; 9,86; 10,17; 10,05. Predpokladáme, že náhodný výber bol realizovaný zo základného súboru s normálnym rozdelením.*

a) Na hladine významnosti α=0,05 otestujme, či dĺžky súčiastok sa významne líšia od požadovanej dĺžky.

b) Na hladine významnosti α=0,01 otestujme, či dĺžky súčiastok sú menšie ako požadovaná dĺžka.

c) Na hladine významnosti α=0,01 otestujme, či dĺžky súčiastok sú väčšie ako požadovaná dĺžka.

Riešenie: Na testovanie použijeme jednovýberový t–test, pretože nepoznáme rozptyl náhodného výberu. Rozsah súboru je $n = 9$.

a) Budeme testovať

$$H_0 : \mu = 10,$$

$$H_1 : \mu \neq 10.$$

Nulová hypotéza hovorí, že dĺžky súčiastok sa významne nelíšia od požadovanej dĺžky. Proti alternatívnej hypotéze, ktorá hovorí, že dĺžky súčiastok sa významne líšia od požadovanej dĺžky. Hodnotu testovacieho kritéria t určíme podľa vzťahu (97)

$$t = \frac{\overline{x} - 10}{s} \cdot \sqrt{9},$$

kde výberový aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (71)

$$\overline{x} = \frac{1}{9}\,(10,24 + 10,12 + 9,91 + 10,19 + 9,78 + 10,14 + 9,86 + 10,17 + 10,05) = 10,0511$$

a výberovú smerodajnú odchýlku s vypočítame podľa vzťahu (77), pričom pre výberový rozptyl s^2 platí (73)

$$s^2 = \frac{1}{8}\big[(10,24 - 10,0511)^2 + (10,12 - 10,0511)^2 + (9,91 - 10,0511)^2 + (10,19 - 10,0511)^2 +$$
$$+(9,78 - 10,0511)^2 + (10,14 - 10,0511)^2 + (9,86 - 10,0511)^2 + (10,17 - 10,0511)^2 +$$
$$+(10,05 - 10,0511)^2\big] = 0,0265 \Rightarrow s = \sqrt{s^2} = 0,1627.$$

Potom hodnota testovacieho kritéria je $t = 0,9426$. Kvantil Studentovho t–rozdelenia je $t_{1-\frac{\alpha}{2}}(n-1) = t_{0,975}(8) = 2,306$ (Tabuľka 29). Kritická oblasť je

$$K_{0,05} = (-\infty; -2,306) \cup (2,306; \infty) \Rightarrow t \notin K_{0,05}.$$

Hodnota testovacieho kritéria $t = 0,9426$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti α=0,05 sa dĺžky súčiastok významne nelíšia od požadovanej hodnoty.

b) Budeme testovať

$$H_0 : \mu = 10,$$
$$H_1 : \mu < 10.$$

Nulová hypotéza hovorí, že dĺžky súčiastok sa významne nelíšia od požadovanej dĺžky. Proti alternatívnej hypotéze, ktorá hovorí, že dĺžky súčiastok sú menšie ako požadovaná dĺžka. Hodnotu testovacieho kritéria t vypočítame rovnako ako v prípade a) a teda $t = 0,9426$. Kvantil Studentovho t–rozdelenia je $t_{1-\alpha}(n-1) = t_{0,99}(8) = 2,896$ (Tabuľka 29). Kritická oblasť je

$$K_{0,01} = (-\infty; -2,896) \Rightarrow t \notin K_{0,01}.$$

Hodnota testovacieho kritéria $t = 0,9426$ nepatrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti α=0,01 dĺžky súčiastok nie sú menšie ako požadovaná hodnota.

c) Budeme testovať

$$H_0 : \mu = 10,$$
$$H_1 : \mu > 10.$$

Nulová hypotéza hovorí, že dĺžky súčiastok sa významne nelíšia od požadovanej dĺžky. Proti alternatívnej hypotéze, ktorá hovorí, že dĺžky súčiastok sú väčšie ako požadovaná dĺžka. Hodnotu testovacieho kritéria t vypočítame rovnako ako v prípade a) a teda $t = 0,9426$. Kvantil Studentovho t–rozdelenia je $t_{1-\alpha}(n-1) = t_{0,99}(8) = 2,896$ (Tabuľka 29). Kritická oblasť je

$$K_{0,01} = (2,896; \infty) \Rightarrow t \notin K_{0,01}.$$

Hodnota testovacieho kritéria $t = 0,9426$ nepatrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,01$ dĺžky súčiastok nie sú väčšie ako požadovaná hodnota.

5.1.3 Jednovýberový test rozptylu σ^2, pričom poznáme strednú hodnotu μ

1. Hypotézy:

$$H_0 : \sigma^2 = \sigma_0^2 \quad (\sigma_0^2 \text{ je známa konštanta}),$$
$$H_1 : \sigma^2 \neq \sigma_0^2 \quad (\text{resp. } \sigma^2 < \sigma_0^2, \text{ resp. } \sigma^2 > \sigma_0^2).$$

2. Testovacie kritérium je

$$\chi^2 = \frac{n\, s_0^2}{\sigma_0^2} \sim \chi^2(n), \tag{98}$$

kde n je rozsah náhodného výberu a s_0^2 je výberový rozptyl.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\left(0; \chi^2_{\frac{\alpha}{2}}(n)\right) \cup \left(\chi^2_{1-\frac{\alpha}{2}}(n); \infty\right)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\left(\chi^2_{1-\alpha}(n); \infty\right)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$(0; \chi^2_\alpha(n))$

Príklad 5.3 *Dĺžka vyrobenej súčiastky má mať hodnotu 10 cm. Náhodne bolo vybraných 9 súčiastok, ktorých dĺžky [v cm] boli 10,24; 10,12; 9,91; 10,19; 9,78; 10,14; 9,86; 10,17; 10,05. Predpokladáme, že náhodný výber bol realizovaný zo základného súboru s normálnym rozdelením, ktorého variabilita dĺžky súčiastok je vyjadrená smerodajnou odchýlkou 0,1 cm.*

a) Na hladine významnosti α=0,05 otestujme, či variabilita dĺžky súčiastok sa významne líši od udávanej variability.

b) Na hladine významnosti α=0,05 otestujme, či variabilita dĺžky súčiastok je menšia ako udávaná variabilita.

c) Na hladine významnosti α=0,05 otestujme, či variabilita dĺžky súčiastok je väčšia ako udávaná variabilita.

Riešenie: Na testovanie použijeme jednovýberový test rozptylu σ^2, pričom poznáme strednú hodnotu $\mu = 10$. Rozsah súboru je $n = 9$. Máme zadanú smerodajnú odchýlku $\sigma_0 = 0,1$ cm, teda $\sigma_0^2 = 0,1^2 = 0,01$.

a) Budeme testovať

$$H_0 : \sigma^2 = 0,01,$$
$$H_1 : \sigma^2 \neq 0,01.$$

Nulová hypotéza hovorí, že variabilita dĺžky súčiastok sa významne nelíši od požadovanej variability. Proti alternatívnej hypotéze, ktorá hovorí, že variabilita dĺžky súčiastok sa významne líši od udávanej hodnoty variability. Hodnotu testovacieho kritéria χ^2 určíme podľa vzťahu (98)

$$\chi^2 = \frac{9 \cdot s_0^2}{0,01},$$

kde výberový rozptyl s_0^2 vypočítame podľa vzťahu (75)

$$s_0^2 = \frac{1}{9}\big[(10,24-10)^2 + (10,12-10)^2 + (9,91-10)^2 + (10,19-10)^2 + (9,78-10)^2 +$$
$$+ (10,14-10)^2 + (9,86-10)^2 + (10,17-10)^2 + (10,05-10)^2\big] = 0,0235.$$

Potom hodnota testovacieho kritéria je $\chi^2 = 21,159$. Kvantily χ^2–rozdelenia sú $\chi^2_{1-\frac{\alpha}{2}}(n) = \chi^2_{0,975}(9) = 19,023$ a $\chi^2_{\frac{\alpha}{2}}(n) = \chi^2_{0,025}(9) = 2,7$ (Tabuľka 30). Kritická oblasť je

$$K_{0,05} = (0; 2,7) \cup (19,023; \infty) \Rightarrow \chi^2 \in K_{0,05}.$$

Hodnota testovacieho kritéria $\chi^2 = 21,159$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ sa variabilita dĺžky súčiastok významne líši od udávanej hodnoty.

b) Budeme testovať

$$H_0 : \sigma^2 = 0,01,$$
$$H_1 : \sigma^2 < 0,01.$$

Nulová hypotéza hovorí, že variabilita dĺžky súčiastok sa významne nelíši od požadovanej variability. Proti alternatívnej hypotéze, ktorá hovorí, že variabilita dĺžky súčiastok je menšia ako udávaná hodnota variability. Hodnotu testovacieho kritéria χ^2 vypočítame rovnako ako v prípade a) a teda $\chi^2 = 21,159$. Kvantil χ^2–rozdelenia je $\chi^2_\alpha(n) = \chi^2_{0,05}(9) = 3,33$ (Tabuľka 30). Kritická oblasť je

$$K_{0,05} = (0; 3,33) \Rightarrow \chi^2 \notin K_{0,05}.$$

Hodnota testovacieho kritéria $\chi^2 = 21,159$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ nie je variabilita dĺžky súčiastok menšia ako udávaná hodnota variability.

c) Budeme testovať

$$H_0 : \sigma^2 = 0,01,$$
$$H_1 : \sigma^2 > 0,01.$$

Nulová hypotéza hovorí, že dĺžky súčiastok sa významne nelíšia od požadovanej dĺžky. Proti alternatívnej hypotéze, ktorá hovorí, že dĺžky súčiastok sú väčšie ako požadovaná dĺžka. Hodnotu testovacieho kritéria χ^2 vypočítame rovnako ako v prípade a) a teda $\chi^2 = 21,159$. Kvantil χ^2–rozdelenia je $\chi^2_{1-\alpha}(n) = \chi^2_{0,95}(9) = 16,92$ (Tabuľka 30). Kritická oblasť je

$$K_{0,05} = (16,92; \infty) \Rightarrow \chi^2 \in K_{0,05}.$$

Hodnota testovacieho kritéria $\chi^2 = 21,159$ patrí do kritickej oblasti $K_{0,05}$, nulovú hypotézu zamietame H_0 a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ je variabilita dĺžky súčiastok väčšia ako udávaná hodnota variability.

5.1.4 Jednovýberový test rozptylu σ^2, pričom nepoznáme strednú hodnotu μ

1. Hypotézy:

$$H_0 : \sigma^2 = \sigma_0^2 \quad (\sigma_0^2 \text{ je známa konštanta}),$$
$$H_1 : \sigma^2 \neq \sigma_0^2 \quad (\text{resp. } \sigma^2 < \sigma_0^2, \text{ resp. } \sigma^2 > \sigma_0^2).$$

2. Testovacie kritérium je

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2} \sim \chi^2(n-1), \tag{99}$$

kde n je rozsah náhodného výberu a s^2 je výberový rozptyl.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\Big(0; \chi^2_{\frac{\alpha}{2}}(n-1)\Big) \cup \Big(\chi^2_{1-\frac{\alpha}{2}}(n-1); \infty\Big)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\big(\chi^2_{1-\alpha}(n-1); \infty\big)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$(0; \chi^2_{\alpha}(n-1))$

Príklad 5.4 *Z krabice bolo náhodne vybraných 9 súčiastok, ktorých dĺžky [v cm] boli 10,24; 10,12; 9,91; 10,19; 9,78; 10,14; 9,86; 10,17; 10,05. Predpokladáme, že náhodný výber bol realizovaný zo základného súboru s normálnym rozdelením, ktorého variabilita dĺžky súčiastok je vyjadrená smerodajnou odchýlkou 0,1 cm.*

a) Na hladine významnosti α=0,05 otestujme, či variabilita dĺžky súčiastok sa významne líši od udávanej variability.

b) Na hladine významnosti α=0,05 otestujme, či variabilita dĺžky súčiastok je menšia ako udávaná variabilita.

c) Na hladine významnosti α=0,05 otestujme, či variabilita dĺžky súčiastok je väčšia ako udávaná variabilita.

Riešenie: Na testovanie použijeme jednovýberový test rozptylu σ^2, pričom nepoznáme strednú hodnotu μ. Rozsah súboru je $n = 9$. Máme zadanú smerodajnú odchýlku $\sigma_0 = 0,1$ cm, teda $\sigma_0^2 = 0,1^2 = 0,01$.

a) Budeme testovať

$$H_0 : \sigma^2 = 0,01,$$
$$H_1 : \sigma^2 \neq 0,01.$$

Nulová hypotéza hovorí, že variabilita dĺžky súčiastok sa významne nelíši od požadovanej variability. Proti alternatívnej hypotéze, ktorá hovorí, že variabilita dĺžky súčiastok sa významne líši od udávanej hodnoty variability. Hodnotu testovacieho kritéria χ^2 určíme podľa vzťahu (99)

$$\chi^2 = \frac{8 \cdot s^2}{0,01},$$

kde výberový aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (71)

$$\overline{x} = \frac{1}{9}(10,24 + 10,12 + 9,91 + 10,19 + 9,78 + 10,14 + 9,86 + 10,17 + 10,05) = 10,0511$$

a výberový rozptyl s^2 vypočítame podľa vzťahu (73)

$$s^2 = \frac{1}{8}\big[(10,24 - 10,0511)^2 + (10,12 - 10,0511)^2 + (9,91 - 10,0511)^2 + (10,19 - 10,0511)^2 +$$
$$+ (9,78 - 10,0511)^2 + (10,14 - 10,0511)^2 + (9,86 - 10,0511)^2 + (10,17 - 10,0511)^2 +$$
$$+ (10,05 - 10,0511)^2\big] = 0,0265.$$

Potom hodnota testovacieho kritéria je $\chi^2 = 21,1689$. Kvantily χ^2–rozdelenia sú $\chi^2_{1-\frac{\alpha}{2}}(n-1) = \chi^2_{0,975}(8) = 17,53$ a $\chi^2_{\frac{\alpha}{2}}(n-1) = \chi^2_{0,025}(8) = 2,18$ (Tabuľka 30). Kritická oblasť je

$$K_{0,05} = (0; 2,18) \cup (17,53; \infty) \Rightarrow \chi^2 \in K_{0,05}.$$

Hodnota testovacieho kritéria $\chi^2 = 21,1689$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ sa variabilita dĺžky súčiastok významne líši od udávanej hodnoty.

b) Budeme testovať

$$H_0 : \sigma^2 = 0,01,$$

$$H_1 : \sigma^2 < 0,01.$$

Nulová hypotéza hovorí, že variabilita dĺžky súčiastok sa významne nelíši od požadovanej variability. Proti alternatívnej hypotéze, ktorá hovorí, že variabilita dĺžky súčiastok je menšia ako udávaná hodnota variability. Hodnotu testovacieho kritéria χ^2 vypočítame rovnako ako v prípade a) a teda $\chi^2 = 21,1689$. Kvantil χ^2–rozdelenia je $\chi^2_\alpha(n-1) = \chi^2_{0,05}(8) = 2,73$ (Tabuľka 30). Kritická oblasť je

$$K_{0,05} = (0; 2,73) \Rightarrow \chi^2 \notin K_{0,05}.$$

Hodnota testovacieho kritéria $\chi^2 = 21,1689$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ nie je variabilita dĺžky súčiastok menšia ako udávaná hodnota variability.

c) Budeme testovať

$$H_0 : \sigma^2 = 0,01,$$

$$H_1 : \sigma^2 > 0,01.$$

Nulová hypotéza hovorí, že dĺžky súčiastok sa významne nelíšia od požadovanej dĺžky. Proti alternatívnej hypotéze, ktorá hovorí, že dĺžky súčiastok sú väčšie ako požadovaná dĺžka. Hodnotu testovacieho kritéria χ^2 vypočítame rovnako ako v prípade a) a teda $\chi^2 = 21,1689$. Kvantil χ^2–rozdelenia je $\chi^2_{1-\alpha}(n-1) = \chi^2_{0,95}(8) = 15,51$ (Tabuľka 30). Kritická oblasť je

$$K_{0,05} = (15,51; \infty) \Rightarrow \chi^2 \in K_{0,05}.$$

Hodnota testovacieho kritéria $\chi^2 = 21,1689$ patrí do kritickej oblasti $K_{0,05}$, nulovú hypotézu zamietame H_0 a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ je variabilita dĺžky súčiastok väčšia ako udávaná hodnota variability.

5.2 Jednovýberový parametrický test o parametri exponenciálneho rozdelenia

Uvažujme náhodný výber V_n, ktorý pochádza zo základného súboru s exponenciálnym rozdelením s parametrom λ, $\lambda > 0$.

1. Hypotézy:
$$H_0 : \lambda = \lambda_0 \quad (\lambda_0 \text{ je známa konštanta}),$$
$$H_1 : \lambda \neq \lambda_0 \quad (\text{resp. } \lambda < \lambda_0, \text{ resp. } \lambda > \lambda_0).$$

2. Testovacie kritérium je
$$\chi^2 = \frac{2n\overline{x}}{\lambda_0} \sim \chi^2(2n), \qquad (100)$$
kde $\overline{x}$ je výberový aritmetický priemer a n je rozsah náhodného výberu.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\lambda = \lambda_0$	$\lambda \neq \lambda_0$	$\left(0; \chi^2_{\frac{\alpha}{2}}(2n)\right) \cup \left(\chi^2_{1-\frac{\alpha}{2}}(2n); \infty\right)$
$\lambda = \lambda_0$	$\lambda > \lambda_0$	$\left(\chi^2_{1-\alpha}(2n); \infty\right)$
$\lambda = \lambda_0$	$\lambda < \lambda_0$	$(0; \chi^2_\alpha(2n))$

Príklad 5.5 *U 100 náhodne vybraných cestujúcich bola zisťovaná doba ich čakania na autobus [v min.]. Výsledky sú uvedené v tabuľke:*

priemerná doba čakania x_j [v min.]	0,5	2,5	4,5	6,5	8,5
počet ľudí n_j	60	20	11	5	4

Predpokladáme, že doba čakania má exponenciálne rozdelenie s parametrom λ. Dopravca tvrdí, že priemerná doba čakania na autobus je dvojminútová. Rozhodnime, či hovorí pravdu.

a) Na hladine významnosti α=0,05 otestujme, či priemerná doba čakania vybraných cestujúcich sa významne líši od doby čakania danej dopravcom.

b) Na hladine významnosti α=0,01 otestujme, či priemerná doba čakania vybraných cestujúcich je väčšia ako doba čakania daná dopravcom.

c) Na hladine významnosti α=0,01 otestujme, či priemerná doba čakania vybraných cestujúcich je menšia ako doba čakania daná dopravcom.

Riešenie: Na testovanie použijeme jednovýberový parametrický test o parametri exponenciálneho rozdelenia. Rozsah súboru je $n = 100$.

a) Budeme testovať

$$H_0 : \lambda = 2,$$
$$H_1 : \lambda \neq 2.$$

Nulová hypotéza hovorí, že priemerná doba čakania λ sa významne nelíši od doby čakania danej dopravcom. Proti alternatívnej hypotéze, ktorá hovorí, že priemerná doba čakania sa významne líši od doby čakania danej dopravcom. Hodnotu testovacieho kritéria χ^2 určíme podľa vzťahu (100)

$$\chi^2 = \frac{200 \cdot \overline{x}}{2},$$

kde výberový aritmetický priemer vypočítame podľa vzťahu (72)

$$\overline{x} = \frac{1}{100}\big(0,5 \cdot 60 + 2,5 \cdot 20 + 4,5 \cdot 11 + 6,5 \cdot 5 + 8,5 \cdot 4\big) = 1,96.$$

Potom hodnota testovacieho kritéria je $\chi^2 = 196$. Kvantily χ^2–rozdelenia sú $\chi^2_{1-\frac{\alpha}{2}}(2n) = \chi^2_{0,975}(200) = 241,06$ a $\chi^2_{\frac{\alpha}{2}}(2n) = \chi^2_{0,025}(200) = 162,73$ (Tabuľka 32). Kritická oblasť je

$$K_{0,05} = (0; 162,73) \cup (241,06; \infty) \Rightarrow \chi^2 \notin K_{0,05}.$$

Hodnota testovacieho kritéria $\chi^2 = 196$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ priemerná doba čakania sa významne nelíši od doby čakania danej dopravcom.

b) Budeme testovať

$$H_0 : \lambda = 2,$$
$$H_1 : \lambda > 2.$$

Nulová hypotéza hovorí, že priemerná doba čakania λ sa významne nelíši od požadovanej doby. Proti alternatívnej hypotéze, ktorá hovorí, že priemerná doba čakania je väčšia ako doba čakania daná dopravcom. Hodnotu testovacieho kritéria χ^2 vypočítame rovnako ako v prípade a) a teda $\chi^2 = 196$. Kvantil χ^2–rozdelenia je $\chi^2_{1-\alpha}(2n) = \chi^2_{0,99}(200) = 249,45$ (Tabuľka 32). Kritická oblasť je

$$K_{0,01} = (249,45; \infty) \Rightarrow \chi^2 \notin K_{0,01}.$$

Hodnota testovacieho kritéria $\chi^2 = 196$ nepatrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,01$ priemerná doba čakania nie je väčšia ako doba čakania daná dopravcom.

c) Budeme testovať

$$H_0 : \lambda = 2,$$
$$H_1 : \lambda < 2.$$

Nulová hypotéza hovorí, že priemerná doba čakania λ sa významne nelíši od požadovanej doby. Proti alternatívnej hypotéze, ktorá hovorí, že priemerná doba čakania je menšia ako doba čakania daná dopravcom. Hodnotu testovacieho kritéria χ^2 vypočítame rovnako ako v prípade a) a teda $\chi^2 = 196$. Kvantil χ^2–rozdelenia je $\chi^2_\alpha(2n) = \chi^2_{0,01}(200) = 156,43$ (Tabuľka 32). Kritická oblasť je

$$K_{0,01} = (0; 156,43) \Rightarrow \chi^2 \notin K_{0,01}.$$

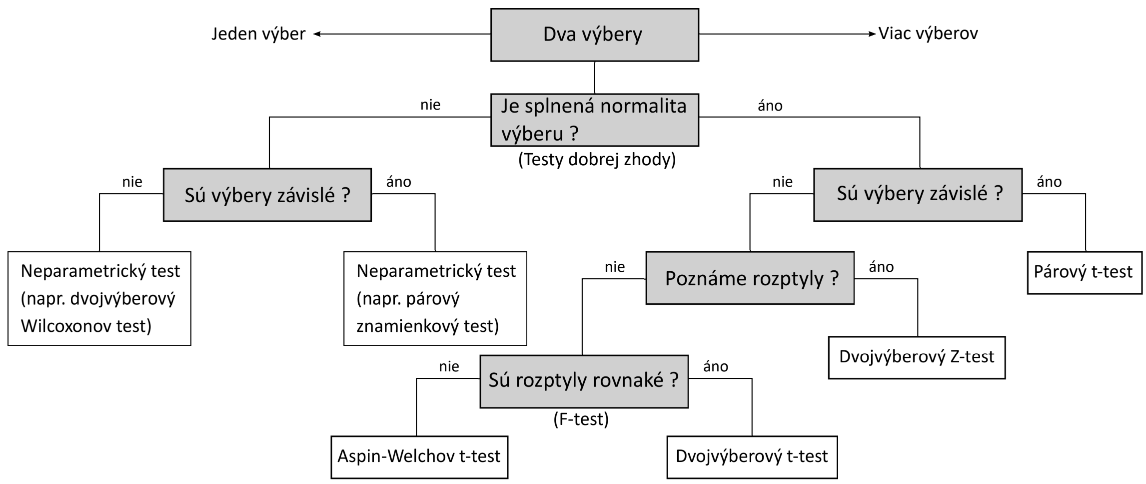
Hodnota testovacieho kritéria $\chi^2 = 196$ nepatrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu nezamietame H_0. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,01$ priemerná doba čakania nie je menšia ako doba čakania daná dopravcom.

5.3 Dvojvýberové parametrické testy o parametroch normálneho rozdelenia

Uvažujme dva náhodné výbery V_{n_1} a V_{n_2}. Náhodný výber V_{n_1} má rozsah n_1 a pochádza zo základného súboru s normálnym rozdelením so strednou hodnotou μ_1 a smerodajnou odchýlkou σ_1. Zapisujeme $V_{n_1} \sim N(\mu_1, \sigma_1^2)$. Náhodný výber V_{n_2} má rozsah n_2 a pochádza zo základného súboru s normálnym rozdelením so strednou hodnotou μ_2 a smerodajnou odchýlkou σ_2. Zapisujeme $V_{n_2} \sim N(\mu_2, \sigma_2^2)$.

Pri testovaní musíme zohľadniť, či náhodné výbery pochádzajú zo závislých alebo nezávislých súborov. V prípade *nezávislých súborov* predpokladáme, že výber z jedného základného súboru nezávisí od výberu z druhého základného súboru. Naopak, pri *závislých súboroch* predpokladáme, že výber z jedného základného súboru závisí od výberu z druhého základného súboru.

Ak nulovú hypotézu o zhode daných parametrov nezamietame, medzi testovanými parametrami nie je štatisticky významný rozdiel. Všeobecný postup výberu testu stredných hodnôt pre dva výbery je zobrazený na Obr. 23.



Obr. 23: Postup výberu testu stredných hodnôt pre dva výbery

5.3.1 Test zhody rozptylov dvoch nezávislých súborov (Fisherov F–test)

Predpokladáme, že máme dva nezávislé náhodné výbery V_{n_1} a V_{n_2}, kde $V_{n_1} \sim N(\mu_1, \sigma_1^2)$ a $V_{n_2} \sim N(\mu_2, \sigma_2^2)$. Náhodný výber V_{n_1} má rozsah n_1 a náhodný výber V_{n_2} má rozsah n_2.

1. Hypotézy:

$$H_0 : \sigma_1^2 = \sigma_2^2,$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad (\text{resp. } \sigma_1^2 < \sigma_2^2, \text{ resp. } \sigma_1^2 > \sigma_2^2).$$

2. Testovacie kritérium je

$$F = \frac{s_1^2}{s_2^2} \sim F(n_1 - 1, n_2 - 1), \tag{101}$$

kde s_1^2, s_2^2 sú výberové rozptyly, pre ktoré platí

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \overline{x}_1)^2, \qquad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{2i} - \overline{x}_2)^2,$$

kde $\overline{x}_1$ a $\overline{x}_2$ sú výberové aritmetické priemery daných náhodných výberov.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\big(0; F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)\big) \cup \big(F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1); \infty\big)$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$(F_{1-\alpha}(n_1 - 1, n_2 - 1); \infty)$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$(0; F_\alpha(n_1 - 1, n_2 - 1))$

5.3.2 Test zhody stredných hodnôt dvoch nezávislých súborov

Dvojvýberový Z–test

Dvojvýberový Z–test použijeme, ak poznáme rozptyly základných súborov σ_1^2, σ_2^2.

1. Hypotézy:

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2 \quad (\text{resp. } \mu_1 < \mu_2, \text{ resp. } \mu_1 > \mu_2).$$

2. Testovacie kritérium je

$$Z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1), \tag{102}$$

kde $\overline{x}_1$ a $\overline{x}_2$ sú výberové aritmetické priemery daných náhodných výberov.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; \infty\right)$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$(u_{1-\alpha}; \infty)$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$(-\infty; -u_{1-\alpha})$

Príklad 5.6 *Dvoma druhmi ochranného náteru sme natreli 24 predmetov. Prvým náterom sme natreli 10 predmetov a druhým 14 predmetov. Po istej dobe sa na každom predmete zistil stupeň opotrebovania. Zistili sme, že priemerný stupeň opotrebovania prvého druhu náteru bol 62, pričom σ_1^2 bolo 16 a priemerný stupeň opotrebovania druhého druhu náteru bol 60, pričom σ_2^2 bolo 15. Predpokladáme normálne rozdelenie oboch základných súborov, z ktorých boli realizované náhodné výbery.*

a) Overme predpoklad o rovnakej účinnosti obidvoch druhov ochranných náterov na hladine významnosti α=0,05.

b) Overme predpoklad o vyššej účinnosti (nižší stupeň opotrebovania) prvého druhu ochranného náteru na hladine významnosti α=0,01.

c) Overme predpoklad o nižšej účinnosti (vyšší stupeň opotrebovania) prvého druhu ochranného náteru na hladine významnosti α=0,1.

Riešenie: Testujeme zhodu stredných hodnôt dvoch nezávislých súborov. Na testovanie použijeme dvojvýberový Z–test (poznáme rozptyly základných súborov). Rozsah prvého náhodného výberu je $n_1 = 10$ a rozsah druhého náhodného výberu je $n_2 = 14$.

a) Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že ochranné nátery nemajú rovnakú účinnosť. Zo zadania úlohy vyplýva, že výberové aritmetické priemery sú $\overline{x}_1 = 62$, $\overline{x}_2 = 60$ a výberové rozptyly základných súborov sú $\sigma_1^2 = 16$, $\sigma_2^2 = 15$. Hodnota testovacieho kritéria Z je podľa vzťahu (102)

$$Z = \frac{62 - 60}{\sqrt{\frac{16}{10} + \frac{15}{14}}} = 1,2237.$$

Kvantil normovaného normálneho rozdelenia je $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$ (Tabuľka 28). Kritická oblasť je

$$K_{0,05} = (-\infty; -1,96) \cup (1,96; \infty) \Rightarrow Z \notin K_{0,05}.$$

Hodnota testovacieho kritéria $Z = 1,2237$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ ochranné nátery majú rovnakú účinnosť.

b) Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 < \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že prvý druh ochranného náteru má nižší stupeň opotrebovania, teda vyššiu účinnosť. Hodnotu testovacieho kritéria Z vypočítame rovnako ako v prípade a) a teda $Z = 1,2237$. Kvantil normovaného normálneho rozdelenia je $u_{1-\alpha} = u_{0,99} = 2,3264$ (Tabuľka 28). Kritická oblasť je

$$K_{0,01} = (-\infty; -2,3264) \Rightarrow Z \notin K_{0,01}.$$

Hodnota testovacieho kritéria $Z = 1,2237$ nepatrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,01$ ochranný náter prvého druhu nemá vyššiu účinnosť ako náter druhého druhu.

c) Budeme testovať

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu_1 > \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že prvý druh ochranného náteru má vyšší stupeň opotrebovania, teda nižšiu účinnosť. Hodnotu testovacieho kritéria Z vypočítame rovnako ako v prípade a) a teda $Z = 1,2237$. Kvantil normovaného normálneho rozdelenia je $u_{1-\alpha} = u_{0,9} = 1,2816$ (Tabuľka 28). Kritická oblasť je

$$K_{0,1} = (1,2816; \infty) \Rightarrow Z \notin K_{0,1}.$$

Hodnota testovacieho kritéria $Z = 1,2237$ nepatrí do kritickej oblasti $K_{0,1}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,1$ ochranný náter prvého druhu nemá nižšiu účinnosť ako náter druhého druhu.

Dvojvýberový t–test

Dvojvýberový t–test použijeme, ak nepoznáme rozptyly základných súborov σ_1^2, σ_2^2, ale môžeme predpokladať, že $\sigma_1^2 = \sigma_2^2$.

1. Hypotézy:

$$H_0: \mu_1 = \mu_2,$$
$$H_1: \mu_1 \neq \mu_2 \quad (\text{resp. } \mu_1 < \mu_2, \text{ resp. } \mu_1 > \mu_2).$$

2. Testovacie kritérium je

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{(n_1-1)s_1^2 + (n_2-1)s_2^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} \sim t(n_1+n_2-2), \tag{103}$$

kde s_1^2, s_2^2 sú výberové rozptyly, pre ktoré platí

$$s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_{1i}-\overline{x}_1)^2, \qquad s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(x_{2i}-\overline{x}_2)^2,$$

kde $\overline{x}_1$ a $\overline{x}_2$ sú výberové aritmetické priemery daných náhodných výberov.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\big(-\infty; -t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\big) \cup \big(t_{1-\frac{\alpha}{2}}(n_1+n_2-2); \infty\big)$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$(t_{1-\alpha}(n_1+n_2-2); \infty)$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$(-\infty; -t_{1-\alpha}(n_1+n_2-2))$

Aspin-Welchov test

Aspin-Welchov test použijeme, ak nepoznáme rozptyly základných súborov σ_1^2, σ_2^2, ale môžeme predpokladať, že $\sigma_1^2 \neq \sigma_2^2$.

1. Hypotézy:

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2 \quad (\text{resp. } \mu_1 < \mu_2, \text{ resp. } \mu_1 > \mu_2).$$

2. Testovacie kritérium je

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(m), \tag{104}$$

pričom

$$m = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}, \tag{105}$$

kde s_1^2, s_2^2 sú výberové rozptyly, pre ktoré platí

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \overline{x}_1)^2, \qquad s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{2i} - \overline{x}_2)^2,$$

kde $\overline{x}_1$ a $\overline{x}_2$ sú výberové aritmetické priemery daných náhodných výberov.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\big(-\infty; -t_{1-\frac{\alpha}{2}}(m)\big) \cup \big(t_{1-\frac{\alpha}{2}}(m); \infty\big)$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$(t_{1-\alpha}(m); \infty)$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$(-\infty; -t_{1-\alpha}(m))$

Príklad 5.7 *Dvoma druhmi ochranného náteru sme natreli 24 predmetov. Prvým náterom sme natreli 10 predmetov a druhým 14 predmetov. Po istej dobe sa na každom predmete zistil stupeň opotrebovania. Zistili sme, že priemerný stupeň opotrebovania prvého druhu náteru bol 62, pričom* s_1^2 *bolo 16 a priemerný stupeň opotrebovania druhého druhu náteru bol 60, pričom* s_2^2 *bolo 15. Predpokladáme normálne rozdelenie oboch základných súborov, z ktorých boli realizované náhodné výbery.*

a) Overme predpoklad o rovnakej účinnosti obidvoch druhov ochranných náterov na hladine významnosti α*=0,05.*

b) Overme predpoklad o vyššej účinnosti (nižší stupeň opotrebovania) prvého druhu ochranného náteru na hladine významnosti α=0,01.

c) Overme predpoklad o nižšej účinnosti (vyšší stupeň opotrebovania) prvého druhu ochranného náteru na hladine významnosti α=0,1.

Riešenie: Testujeme zhodu stredných hodnôt dvoch nezávislých súborov, pričom nepoznáme rozptyly základných súborov. Rozsah prvého náhodného výberu je $n_1 = 10$ a rozsah druhého náhodného výberu je $n_2 = 14$. Najprv musíme overiť, či môžeme predpokladať, že rozptyly oboch súborov sú rovnaké alebo rôzne. Podľa toho budeme potom rozhodovať, ktorý z testov (dvojvýberový t–test alebo Aspin-Welchov test) použijeme na otestovanie zhody stredných hodnôt dvoch nezávislých súborov.

a) Na overenie, či rozptyly oboch súborov sú rovnaké alebo rôzne, použijeme test zhody rozptylov dvoch nezávislých súborov (Fisherov F–test). Budeme testovať

$$H_0 : \sigma_1^2 = \sigma_2^2,$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Nulová hypotéza hovorí, že je splnený predpoklad rovnosti rozptylov. Proti alternatívnej hypotéze, ktorá hovorí, že nie je splnený predpoklad rovnosti rozptylov. Hodnota testovacieho kritéria F je podľa vzťahu (101)

$$F = \frac{16}{15} = 1,067.$$

Kvantily Fisherovho F–rozdelenia sú $F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0{,}975}(9, 13) = 3,312$ a $F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0{,}025}(9, 13) = 0,2611$ (Tabuľka 38). Kritická oblasť je

$$K_{0{,}05} = (0; 0,2611) \cup (3,312; \infty) \Rightarrow F \notin K_{0{,}05}.$$

Hodnota testovacieho kritéria $F = 1,067$ nepatrí do kritickej oblasti $K_{0{,}05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ nie sú rozdiely medzi rozptylmi oboch súborov štatisticky významné. Z toho vyplýva, že na testovanie zhody stredných hodnôt dvoch nezávislých súborov použijeme dvojvýberový t–test.

Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že ochranné nátery nemajú rovnakú účinnosť. Zo zadania úlohy vyplýva, že výberové aritmetické priemery sú $\overline{x}_1 = 62$, $\overline{x}_2 = 60$ a výberové rozptyly sú $s_1^2 = 16$, $s_2^2 = 15$. Hodnota testovacieho kritéria t je podľa vzťahu (103)

$$t = \frac{62 - 60}{\sqrt{(10-1)\cdot 16 + (14-1)\cdot 15}} \cdot \sqrt{\frac{10 \cdot 14 \cdot (10 + 14 - 2)}{10 + 14}} = 1,2306.$$

Kvantil Studentovho t–rozdelenia je $t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2) = t_{0,975}(22) = 2,0739$ (Tabuľka 29). Kritická oblasť je

$$K_{0,05} = (-\infty; -2,0739) \cup (2,0739; \infty) \Rightarrow t \notin K_{0,05}.$$

Hodnota testovacieho kritéria $t = 1,2306$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ ochranné nátery majú rovnakú účinnosť.

b) Na overenie, či rozptyly oboch súborov sú rovnaké alebo rôzne, použijeme test zhody rozptylov dvoch nezávislých súborov (Fisherov F–test). Budeme testovať

$$H_0 : \sigma_1^2 = \sigma_2^2,$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Nulová hypotéza hovorí, že je splnený predpoklad rovnosti rozptylov. Proti alternatívnej hypotéze, ktorá hovorí, že nie je splnený predpoklad rovnosti rozptylov. Hodnotu testovacieho kritéria F vypočítame rovnako ako v prípade a) a teda $F = 1,067$. Kvantily Fisherovho F–rozdelenia sú $F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0,995}(9,13) = 4,9351$ a $F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0,005}(9,13) = 0,1625$ (Tabuľka 42). Kritická oblasť je

$$K_{0,01} = (0; 0,1625) \cup (4,9351; \infty) \Rightarrow F \notin K_{0,01}.$$

Hodnota testovacieho kritéria $F = 1,067$ nepatrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,01$ nie sú rozdiely medzi rozptylmi oboch súborov štatisticky významné. Z toho vyplýva, že na testovanie zhody stredných hodnôt dvoch nezávislých súborov použijeme dvojvýberový t–test.

Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 < \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že prvý druh ochranného náteru má nižší stupeň opotrebovania, teda vyššiu účinnosť. Hodnotu testovacieho kritéria t vypočítame rovnako ako v prípade a) a teda

$t = 1,2306$. Kvantil Studentovho t–rozdelenia je $t_{1-\alpha}(n_1 + n_2 - 2) = t_{0,999}(22) = 2,5083$ (Tabuľka 29). Kritická oblasť je

$$K_{0,01} = (-\infty; -2,5083) \Rightarrow t \notin K_{0,01}.$$

Hodnota testovacieho kritéria $t = 1,2306$ nepatrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,01$ ochranný náter prvého druhu nemá vyššiu účinnosť ako ochranný náter druhého druhu.

c) Na overenie, či rozptyly oboch súborov sú rovnaké alebo rôzne, použijeme test zhody rozptylov dvoch nezávislých súborov (Fisherov F–test). Budeme testovať

$$H_0 : \sigma_1^2 = \sigma_2^2,$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Nulová hypotéza hovorí, že je splnený predpoklad rovnosti rozptylov. Proti alternatívnej hypotéze, ktorá hovorí, že nie je splnený predpoklad rovnosti rozptylov. Hodnotu testovacieho kritéria F vypočítame rovnako ako v prípade a) a teda $F = 1,067$. Kvantily Fisherovho F–rozdelenia sú $F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1) = F_{0,95}(9,13) = 2,7144$ a $F_{\frac{\alpha}{2}}(n_1-1, n_2-1) = F_{0,05}(9,13) = 0,3281$ (Tabuľka 34). Kritická oblasť je

$$K_{0,1} = (0; 0,3281) \cup (2,7144; \infty) \Rightarrow F \notin K_{0,1}.$$

Hodnota testovacieho kritéria $F = 1,067$ nepatrí do kritickej oblasti $K_{0,1}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,1$ nie sú rozdiely medzi rozptylmi oboch súborov štatisticky významné. Z toho vyplýva, že na testovanie zhody stredných hodnôt dvoch nezávislých súborov použijeme dvojvýberový t–test.

Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 > \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že prvý druh ochranného náteru má vyšší stupeň opotrebovania, teda nižšiu účinnosť. Hodnotu testovacieho kritéria t vypočítame rovnako ako v prípade a) a teda $t = 1,2306$. Kvantil Studentovho t–rozdelenia je $t_{1-\alpha}(n_1 + n_2 - 2) = t_{0,9}(22) = 1,3212$ (Tabuľka 29). Kritická oblasť je

$$K_{0,1} = (1,3212; \infty) \Rightarrow t \notin K_{0,1}.$$

Hodnota testovacieho kritéria $t = 1,2306$ nepatrí do kritickej oblasti $K_{0,1}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,1$ ochranný náter prvého druhu nemá nižšiu účinnosť ako ochranný náter druhého druhu.

Príklad 5.8 *Dvoma druhmi ochranného náteru sme natreli 24 predmetov. Prvým náterom sme natreli 10 predmetov a druhým 14 predmetov. Po istej dobe sa na každom predmete zistil stupeň opotrebovania. Zistili sme, že priemerný stupeň opotrebovania prvého druhu náteru bol 62, pričom* s_1^2 *bolo 4 a priemerný stupeň opotrebovania druhého druhu náteru bol 60, pričom* s_2^2 *bolo 16. Predpokladáme normálne rozdelenie oboch základných súborov, z ktorých boli realizované náhodné výbery.*

a) Overme predpoklad o rovnakej účinnosti obidvoch druhov ochranných náterov na hladine významnosti α*=0,05.*

b) Overme predpoklad o vyššej účinnosti (nižší stupeň opotrebovania) prvého druhu ochranného náteru na hladine významnosti α*=0,1.*

c) Overme predpoklad o nižšej účinnosti (vyšší stupeň opotrebovania) prvého druhu ochranného náteru na hladine významnosti α*=0,05.*

Riešenie: Testujeme zhodu stredných hodnôt dvoch nezávislých súborov, pričom nepoznáme rozptyly základných súborov. Rozsah prvého náhodného výberu je $n_1 = 10$ a rozsah druhého náhodného výberu je $n_2 = 14$. Rovnako ako v predchádzajúcom príklade, musíme overiť, či môžeme predpokladať, že rozptyly oboch súborov sú rovnaké alebo rôzne.

a) Na overenie, či rozptyly oboch súborov sú rovnaké alebo rôzne, použijeme test zhody rozptylov dvoch nezávislých súborov (Fisherov F–test). Budeme testovať

$$H_0 : \sigma_1^2 = \sigma_2^2,$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Nulová hypotéza hovorí, že je splnený predpoklad rovnosti rozptylov. Proti alternatívnej hypotéze, ktorá hovorí, že nie je splnený predpoklad rovnosti rozptylov. Hodnotu testovacieho kritéria F určíme podľa vzťahu (101)

$$F = \frac{4}{16} = 0,25.$$

Kvantily Fisherovho F–rozdelenia sú $F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0,975}(9,13) = 3,312$ a $F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) = F_{0,025}(9,13) = 0,2611$ (Tabuľka 38). Kritická oblasť je

$$K_{0,05} = (0; 0,2611) \cup (3,312; \infty) \Rightarrow F \in K_{0,05}.$$

Hodnota testovacieho kritéria $F = 0,25$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ sú rozdiely medzi rozptylmi oboch súborov štatisticky významné. Z toho vyplýva, že na testovanie zhody stredných hodnôt dvoch nezávislých súborov použijeme Aspin-Welchov test.

Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 \neq \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že ochranné nátery nemajú rovnakú účinnosť. Zo zadania úlohy vyplýva, že výberové aritmetické priemery sú $\overline{x}_1 = 62$, $\overline{x}_2 = 60$ a výberové rozptyly sú $s_1^2 = 4$, $s_2^2 = 16$. Hodnotu testovacieho kritéria t určíme podľa vzťahu (104)

$$t = \frac{62 - 60}{\sqrt{\frac{4}{10} + \frac{16}{14}}} = 1,6102.$$

Kvantil Studentovho t–rozdelenia je $t_{1-\frac{\alpha}{2}}(m) = t_{0,975}(20) = 2,0851$ (Tabuľka 29), kde m vypočítame podľa vzťahu (105)

$$m = \frac{\left(\frac{4}{10} + \frac{16}{14}\right)^2}{\frac{\left(\frac{4}{10}\right)^2}{10-1} + \frac{\left(\frac{16}{14}\right)^2}{14-1}} = 20,1305 \approx 20.$$

Kritická oblasť je

$$K_{0,05} = (-\infty; -2,0851) \cup (2,0851; \infty) \Rightarrow t \notin K_{0,05}.$$

Hodnota testovacieho kritéria $t = 1,6102$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ ochranné nátery majú rovnakú účinnosť.

b) Na overenie, či rozptyly oboch súborov sú rovnaké alebo rôzne, použijeme test zhody rozptylov dvoch nezávislých súborov (Fisherov F–test). Budeme testovať

$$H_0 : \sigma_1^2 = \sigma_2^2,$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Nulová hypotéza hovorí, že je splnený predpoklad rovnosti rozptylov. Proti alternatívnej hypotéze, ktorá hovorí, že nie je splnený predpoklad rovnosti rozptylov. Hodnotu testovacieho kritéria F vypočítame rovnako ako v prípade a) a teda $F = 0,25$. Kvantily Fisherovho F–rozdelenia sú $F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1) = F_{0,95}(9, 13) = 2,7144$ a $F_{\frac{\alpha}{2}}(n_1-1, n_2-1) = F_{0,05}(9, 13) = 0,3281$ (Tabuľka 34). Kritická oblasť je

$$K_{0,1} = (0; 0,3281) \cup (2,7144; \infty) \Rightarrow F \in K_{0,1}.$$

Hodnota testovacieho kritéria $F = 0,25$ patrí do kritickej oblasti $K_{0,1}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,1$ sú rozdiely medzi rozptylmi oboch súborov štatisticky významné. Z toho vyplýva, že na testovanie zhody stredných hodnôt dvoch nezávislých súborov použijeme Aspin-Welchov test.

Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 < \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že prvý druh ochranného náteru má nižší stupeň opotrebovania, teda vyššiu účinnosť. Hodnotu testovacieho kritéria t vypočítame rovnako ako v prípade a) a teda $t = 1,6102$. Kvantil Studentovho t–rozdelenia je $t_{1-\alpha}(m) = t_{0,9}(20) = 1,3251$ (Tabuľka 29), kde m vypočítame rovnako ako v prípade a). Kritická oblasť je

$$K_{0,1} = (-\infty; -1,3251) \Rightarrow t \notin K_{0,1}.$$

Hodnota testovacieho kritéria $t = 1,6102$ nepatrí do kritickej oblasti $K_{0,1}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,1$ ochranné nátery majú rovnakú účinnosť.

c) Z úlohy a) sme zistili, že na hladine významnosti $\alpha = 0,05$ môžeme predpokladať, že sú rozdiely medzi rozptyly oboch súborov štatisticky významné. Z toho vyplýva, že na testovanie zhodu stredných hodnôt dvoch nezávislých súborov použijeme Aspin-Welchov test.

Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 > \mu_2.$$

Nulová hypotéza hovorí, že obidva ochranné nátery majú rovnakú účinnosť. Proti alternatívnej hypotéze, ktorá hovorí, že prvý druh ochranného náteru má vyšší stupeň opotrebovania, teda nižšiu účinnosť. Hodnotu testovacieho kritéria t vypočítame rovnako ako v prípade a) a teda $t = 1,6102$. Kvantil Studentovho t–rozdelenia je $t_{1-\alpha}(m) = t_{0,95}(20) = 1,7242$ (Tabuľka 29), kde m vypočítame rovnako ako v prípade a). Kritická oblasť je

$$K_{0,05} = (1,7242; \infty) \Rightarrow t \notin K_{0,05}.$$

Hodnota testovacieho kritéria $t = 1,6102$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ ochranné nátery majú rovnakú účinnosť.

5.3.3 Test zhody stredných hodnôt dvoch závislých súborov

Ak obidva náhodné výbery V_{n_1} a V_{n_2} pochádzajú zo závislých základných súborov a majú rovnaký rozsah (t. j. $n_1 = n_2 = n$), môžeme použiť *párový t–test* (napr. ak meriame nejaký súbor, objekt dvakrát: pred nejakou udalosťou a po nej).

1. Hypotézy:

$$H_0 : \mu_1 = \mu_2,$$
$$H_1 : \mu_1 \neq \mu_2 \quad (\text{resp. } \mu_1 < \mu_2, \text{ resp. } \mu_1 > \mu_2).$$

2. Testovacie kritérium je

$$t = \frac{\overline{d}}{s_d}\sqrt{n} \sim t(n-1), \tag{106}$$

kde $\overline{d}$ je výberový aritmeticý priemer odchýlok

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{i=1}^{n}(x_{1i} - x_{2i}) \tag{107}$$

a s_d^2 je výberový rozptyl odchýlok

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \overline{d}\right)^2. \tag{108}$$

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left(-\infty; -t_{1-\frac{\alpha}{2}}(n-1)\right) \cup \left(t_{1-\frac{\alpha}{2}}(n-1); \infty\right)$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$(t_{1-\alpha}(n-1); \infty)$
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	$(-\infty; -t_{1-\alpha}(n-1))$

Poznámka: Tieto vzťahy sa používajú v prípade malých rozsahov náhodných výberov (podobne ako to bolo v kapitole 4.2), t. j. ak $n_1 \leqq 30$, $n_2 \leqq 30$. V opačnom prípade Studentovo t–rozdelenie aproximujeme normovaným normálnym rozdelením.

Príklad 5.9 *Skúmali sme 8 vzoriek chemickej látky. Každú vzorku sme analyzovali titračnou a polarografickou metódou. Výsledky sú uvedené v tabuľke. Predpokladáme normálne rozdelenie oboch základných súborov, z ktorých boli realizované náhodné výbery. Overme rovnocennosť obidvoch metód na hladine významnosti α=0,01.*

vzorka	*Polarografická metóda*	*Titračná metóda*
1	*18,60*	*18,58*
2	*27,60*	*27,37*
3	*27,50*	*27,27*
4	*25,00*	*24,64*
5	*24,50*	*24,10*
6	*26,80*	*26,33*
7	*29,70*	*29,33*
8	*26,63*	*26,50*

Riešenie: Obidva náhodné výbery pochádzajú zo závislých základných súborov a majú rovnaký rozsah $n_1 = n_2 = n$. Môžeme použiť párový t–test.

Budeme testovať

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 \neq \mu_2.$$

Nulová hypotéza hovorí, že metódy sú rovnocenné. Proti alternatívnej hypotéze, ktorá hovorí, že metódy nie sú rovnocenné. Zadanú tabuľku doplníme o stĺpec odchýlok (Tabuľka 6).

Tabuľka 6

vzorka (i)	Polarografická metóda (x_{1i})	Titračná metóda (x_{2i})	Odchýlky ($d_i = x_{1i} - x_{2i}$)
1	$18,60$	$18,58$	$0,02$
2	$27,60$	$27,37$	$0,23$
3	$27,50$	$27,27$	$0,23$
4	$25,00$	$24,64$	$0,36$
5	$24,50$	$24,10$	$0,40$
6	$26,80$	$26,33$	$0,47$
7	$29,70$	$29,33$	$0,37$
8	$26,63$	$26,50$	$0,13$

Hodnotu testovacieho kritéria t určíme podľa vzťahu (106)

$$t = \frac{\overline{d}}{s_d}\sqrt{8},$$

kde výberový aritmetický priemer odchýlok $\overline{d}$ vypočítame podľa vzťahu (107)

$$\overline{d} = \frac{1}{8}(0,02 + 0,23 + 0,23 + 0,36 + 0,40 + 0,47 + 0,37 + 0,13) = 0,2763$$

a výberový rozptyl odchýlok s_d^2 vypočítame podľa vzťahu (108)

$$s_d^2 = \frac{1}{7}\big[(0,02 - 0,2763)^2 + (0,23 - 0,2763)^2 + (0,23 - 0,2763)^2 +$$

$$+(0,36 - 0,2763)^2 + (0,40 - 0,2763)^2 + (0,47 - 0,2763)^2 + (0,37 - 0,2763)^2 +$$

$$+(0,13 - 0,2763)^2\big] = 0,0229 \Rightarrow s_d = \sqrt{s_d^2} = \sqrt{0,0229} = 0,1513.$$

Potom hodnota testovacieho kritéria je $t = 5,1684$. Kvantil Studentovho t–rozdelenia $t_{1-\frac{\alpha}{2}}(n-1) = t_{0,995}(7) = 3,4995$ (Tabuľka 29). Kritická oblasť je

$$K_{0,01} = (-\infty; -3,4995) \cup (3,4995; \infty) \Rightarrow t \in K_{0,01}.$$

Hodnota testovacieho kritéria $t = 5,1684$ patrí do kritickej oblasti $K_{0,01}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,01$ nie sú použité metódy rovnocenné.

5.4 Testy symetrie

Testy symetrie patria medzi neparametrické testy. Existuje niekoľko takých testov, ale uvedieme znamienkový test a Wilcoxonov dvojvýberový test.

5.4.1 Znamienkový test

Znamienkový test (test mediánu) patrí medzi najstaršie a najjednoduchšie neparametrické testy, ktoré testujú symetriu rozdelenia štatistického znaku okolo konštanty x_0 ako mediánu.

Uvažujme náhodný výber V_n so spojitým rozdelením s neznámym mediánom $\widetilde{x}$. Rozsah výberu je n.

1. Hypotézy:

$$H_0 : \widetilde{x} = x_0,$$

$$H_1 : \widetilde{x} \neq x_0.$$

2. Testovacie kritérium je

$$T = \min\{Z^+, Z^-\} \sim \mathrm{Bi}\left(n_0, \frac{1}{2}\right), \tag{109}$$

kde Z^+ je počet hodnôt $d_i = x_i - x_0$ s kladným znamienkom a Z^- je počet hodnôt $d_i = x_i - x_0$ so záporným znamienkom, n_0 je počet nenulových rozdielov $d_i = x_i - x_0$.

3. Kritická oblasť K_α je

$$K_\alpha = (-\infty; k_1(\alpha; n_0)) \cup (k_2(\alpha; n_0); \infty),$$

kde $k_1(\alpha; n_0)$, $k_2(\alpha; n_0)$ sú kritické hodnoty znamienkového testu tabelované pre $n_0 \leqq 100$ pre $\alpha = 0,05$ a $\alpha = 0,01$ (Tabuľka 45).

Poznámka: Keďže binomické rozdelenie $\mathrm{Bi}\left(n_0, \frac{1}{2}\right)$ je symetrické, platí $k_2 = n_0 - k_1$.

V prípade veľkého rozsahu náhodného výberu (stačí $n \geqq 20$), môžeme použiť nasledujúci test:

1. Hypotézy:

$$H_0 : \widetilde{x} = x_0,$$
$$H_1 : \widetilde{x} \neq x_0.$$

2. Testovacie kritérium je

$$Z = \frac{2T - n_0}{\sqrt{n_0}} \sim N(0, 1), \tag{110}$$

kde $T = \min\{Z^+, Z^-\}$, Z^+ je počet hodnôt $d_i = x_i - x_0$ s kladným znamienkom a Z^- je počet hodnôt $d_i = x_i - x_0$ so záporným znamienkom, n_0 je počet nenulových hodnôt $d_i = x_i - x_0$.

3. Kritická oblasť K_α je

$$K_\alpha = \left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; \infty\right).$$

Poznámka: Znamienkový test sa často používa aj ako *párový znamienkový test*, ktorý je podobný parametrickému párovému t–testu. V tomto prípade rozdiely d_i vypočítame ako rozdiely každého páru pozorovaní, t. j. $d_i = x_i - y_i$ pre $i = 1, 2, \ldots, n$, kde x_i sú hodnoty z prvého náhodného výberu a y_i sú hodnoty z druhého náhodného výberu. Ďalej postupujeme rovnako ako pri znamienkovom teste pre jeden výber.

Príklad 5.10 *Zo skupiny výrobkov sme na kontrolu hmotnosti [v g] vybrali 12 kusov: 21,3; 18,5; 19,7; 26,5; 16,7; 23,1; 19,3; 17,5; 19,1; 24,3; 15,9; 20,8. Na hladine významnosti* α*=0,05 otestujme, či priemerná hmotnosť výrobku je 19 gramov.*

Riešenie: Pretože nepoznáme rozdelenie základného súboru, použijeme znamienkový test. Budeme testovať

$$H_0 : \widetilde{x} = 19,$$

$$H_1 : \widetilde{x} \neq 19.$$

V Tabuľke 7 sú uvedené namerané hodnoty a rozdiely $d_i = x_i - 19$.

Tabuľka 7

x_i	$21,3$	$18,5$	$19,7$	$26,5$	$16,7$	$23,1$	$19,3$	$17,5$	$19,1$	$24,3$	$15,9$	$20,8$
d_i	$2,3$	$-0,5$	$0,7$	$7,5$	$-2,3$	$4,1$	$0,3$	$-1,5$	$0,1$	$5,3$	$-3,1$	$1,8$
znamienko	$+$	$-$	$+$	$+$	$-$	$+$	$+$	$-$	$+$	$+$	$-$	$+$

Počet rozdielov Z^+ s kladným znamienkom je 8, počet rozdielov Z^- so záporným znamienkom je 4. Potom hodnotu testovacieho kritéria T určíme podľa vzťahu (109) $T = \min\{Z^+, Z^-\} = 4$. Keďže počet nenulových hodnôt je $n_0 = 12$ a $\alpha = 0,05$, potom $k_1(\alpha; n_0) = k_1(0,05; 12) = 2$ (Tabuľka 45) a $k_2(\alpha; n_0) = k_2(0,05; 12) = 10$. Kritická oblasť je

$$K_{0,05} = (-\infty; 2) \cup (10; \infty) \Rightarrow T \notin K_{0,05}.$$

Hodnota testovacieho kritéria $T = 4$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ medián hmotnosti výrobkov je 19 gramov.

5.4.2 Wilcoxonov dvojvýberový test

Wilcoxonov dvojvýberový test (test výberových poradí, Mann-Whitneyov U-test) predstavuje neparametrickú analógiu dvojvýberového t–testu pre nezávislé výbery. Používa sa namiesto uvedeného testu vtedy, ak nejde o výbery z normálneho rozdelenia.

Uvažujme dva nezávislé náhodné výbery V_{n_1} a V_{n_2}. Náhodný výber V_{n_1} pochádza zo základného súboru s distribučnou funkciou $F(x)$ s mediánom $\widetilde{x}_1$ a náhodný výber V_{n_2} pochádza zo základného súboru s distribučnou funkciou $G(x)$ s mediánom $\widetilde{x}_2$. Výbery zvolíme tak, aby bola splnená podmienka $n_1 \leqq n_2$.

Najprv všetkých $n_1 + n_2$ hodnôt náhodných výberov usporiadame do neklesajúcej postupnosti, ktorá bude tvoriť tzv. *združený variačný rad.* Potom jednotlivým hodnotám v združenom variačnom rade priradíme poradové čísla, t. j. hodnoty očíslujeme od najmenšej po najväčšiu prirodzenými číslami, pričom rovnako veľkým hodnotám priradíme rovnaké priemerné poradie (aritmetický priemer ich poradia).

V prípade náhodného výberu s malým rozsahom, môžeme použiť nasledujúci test:

1. Hypotézy:

$$H_0 : \widetilde{x}_1 = \widetilde{x}_2,$$
$$H_1 : \widetilde{x}_1 \neq \widetilde{x}_2,$$

resp.

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. Testovacie kritérium je

$$U = \min\{U_1; U_2\}, \tag{111}$$

kde

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T_1, \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - T_2, \tag{112}$$

pričom T_1 je súčet poradí hodnôt z výberu V_{n_1} a T_2 je súčet poradí hodnôt z výberu V_{n_2}. Na overenie môžeme použiť vzťahy

$$U_1 + U_2 = n_1\, n_2 \quad \text{a} \quad T_1 + T_2 = \frac{1}{2}(n_1+n_2)(n_1+n_2+1).$$

3. Kritická oblasť K_α je

$$K_\alpha = (-\infty; U_\alpha(n_1, n_2))\,,$$

kde $U_\alpha(n_1, n_2)$ je tabuľková kritická hodnota Wilcoxonovho dvojvýberového testu (Tabuľka 46).

V prípade náhodného výberu s veľkým rozsahom, môžeme použiť nasledujúci test:

1. Hypotézy:

$$H_0 : \widetilde{x}_1 = \widetilde{x}_2,$$
$$H_1 : \widetilde{x}_1 \neq \widetilde{x}_2,$$

resp.

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. Testovacie kritérium je

$$U_0 = \frac{U_1 - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} \sim N(0,1), \tag{113}$$

kde

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T_1, \tag{114}$$

pričom T_1 je súčet poradí hodnôt z výberu V_{n_1}.

3. Kritická oblasť K_α je

$$K_\alpha = \left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; \infty\right).$$

Príklad 5.11 *Farmaceutická spoločnosť porovnáva nežiadúce vedľajšie účinky dvoch vyrábaných liekov. Sleduje sa [v %] počet pacientov, u ktorých boli tieto účinky zaznamenané. Získané hodnoty sú v tabuľke. Dvojvýberovým Wilcoxonovým testom otestujme na hladine významnosti α=0,05, či majú obidva lieky percentuálne rovnaké nežiadúce účinky.*

Liek A	*31*	*33*	*29*	*35*		
Liek B	*25*	*34*	*28*	*24*	*33*	*40*

Riešenie: Budeme testovať

$$H_0 : \widetilde{x}_1 = \widetilde{x}_2,$$
$$H_1 : \widetilde{x}_1 \neq \widetilde{x}_2.$$

Nulová hypotéza hovorí, že percentuálne množstvo pacientov, u ktorých sa objavili nežiadúce účinky je rovnaké u oboch liekov. Proti obojstrannej alternatívnej hypotéze, ktorá hovorí, že percentuálne množstvo pacientov, u ktorých sa objavili nežiadúce účinky je rôzne u oboch liekov.

Rozsah prvého súboru je $n_1 = 4$ a druhého $n_2 = 6$. Všetky hodnoty, bez ohľadu na to, do ktorej skupiny patria, usporiadame do neklesajúcej postupnosti a každej hodnote priradíme poradie. Ak máme rovnaké hodnoty, priradíme im rovnaké priemerné poradie (Tabuľka 8).

Tabuľka 8

Hodnoty	24	25	28	29	31	33	33	34	35	40
Liek	B	B	B	A	A	A	B	B	A	B
Poradie	1	2	3	4	5	6,5	6,5	8	9	10

Vypočítame súčty poradí T_1, T_2 a hodnoty U_1, U_2 určíme podľa vzťahov (112).

$$T_1 = 4 + 5 + 6,5 + 9 = 24,5,$$

$$T_2 = 1 + 2 + 3 + 6,5 + 8 + 10 = 30,5,$$

$$U_1 = 4 \cdot 6 + \frac{4(4+1)}{2} - 24,5 = 9,5,$$

$$U_2 = 4 \cdot 6 + \frac{6(6+1)}{2} - 30,5 = 14,5.$$

Hodnotu testovacieho kritéria U určíme podľa vzťahu (111)

$$U = \min\{9,5; 14,5\} = 9,5.$$

Kritická hodnota Wicoxonovho dvojvýberového testu je $U_\alpha(n_1, n_2) = U_{0,05}(4, 6) = 2$ (Tabuľka 46). Kritická oblasť je

$$K_{0,05} = (-\infty; 2) \Rightarrow U \notin K_{0,05}.$$

Hodnota testovacieho kritéria $U = 9,5$ nepatrí do kritickej oblasti $K_{0,05}$, preto na hladine významnosti $\alpha = 0,05$ nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že percentuálne množstvo pacientov, u ktorých sa objavili nežiadúce účinky je rovnaké u oboch liekov.

5.5 Testy dobrej zhody

Vo väčšine testov, ktoré sme doteraz spomínali, sme predpokladali, že základný súbor sa riadi určitým rozdelením. V praxi takéto tvrdenie musíme overiť. K tomu slúžia tzv. *testy dobrej zhody*. Existuje veľa takých testov, my uvedieme najčastejšie používané: Pearsonov test dobrej zhody, Kolmogorovov–Smirnovov test, Andersonov–Darlingov test a Shapiro–Wilkov test normality.

Uvažujme náhodný výber V_n, ktorý pochádza zo základného súboru s rozdelením pravdepodobnosti s neznámou distribučnou funkciou $F(x)$. Daná je tiež známa distribučná funkcia $G(x)$ teoretického rozdelenia pravdepodobnosti.

5.5.1 Pearsonov test dobrej zhody

Pearsonov test dobrej zhody (χ^2–test dobrej zhody) je jeden z najstarších štatistických testov (pochádza z roku 1900). Je to univerzálny test, ktorý sa môže použiť pre rozdelenia diskrétnych aj spojitých náhodných premenných. Pritom predpokladáme, že hodnoty náhodného výberu sú roztriedené do k tried (intervalov).

1. Hypotézy:

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. Testovacie kritérium je

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j} \sim \chi^2(k - r - 1), \tag{115}$$

kde n_j je počet hodnôt sledovaného náhodného výberu v j-tej triede (intervale), p_j je pravdepodobnostná miera j-tej triedy (intervalu) určená distribučnou funkciou G, np_j je teoretická početnosť v j-tej triede (intervale), r je počet odhadovaných parametrov pri určovaní G, k je počet tried a n je rozsah náhodného výberu (aspoň $n \geqq 30$).

3. Kritická oblasť K_α je

$$K_\alpha = \left(\chi^2_{1-\alpha}(k - r - 1); \infty\right).$$

Pri praktickom používaní testu sa požaduje splnenie tzv. *Cochranovho pravidla*, t. j. $np_j \geqq 5$ pre $j = 1, 2, \ldots, k$. V prípade, že táto podmienka nie je splnená, t. j. $np_j < 5$, zlúčime susedné triedy tak, aby podmienka bola dodatočne splnená. Jej prísne dodržiavanie je však potrebné len pri malom počte stupňov voľnosti testovacieho kritéria χ^2. Pri väčšom počte stupňov voľnosti testovacieho kritéria $k - r - 1 \geqq 3$ stačí požadovať podmienku $np_j \geqq 4$ a pri $k - r - 1 \geqq 6$ stačí požadovať podmienku $np_j \geqq 1$ pre $j = 1, 2, \ldots, k$.

Príklad 5.12 *Na hladine významnosti α=0,05 zistime, či náhodná premenná daná tabuľkou pochádza zo základného súboru s exponenciálnym rozdelením. Použime pritom Pearsonov test dobrej zhody.*

I_j	$\langle 0;1 \rangle$	$(1;2 \rangle$	$(2;3 \rangle$	$(3;4 \rangle$	$(4;5 \rangle$
n_j	*60*	*20*	*11*	*5*	*4*

Riešenie: Rozsah súboru je $n = 100$. Hodnoty náhodného výberu sú rozdelené do $k = 5$ tried (intervalov). Počet odhadovaných parametrov pri určovaní $G(x)$ je $r = 1$ (exponenciálne rozdelenie má jeden parameter λ). Na odhad parametra λ použijeme výberový aritmetický priemer $\overline{x}$, ktorý vypočítame podľa vzťahu (72), kde z_j sú hodnoty triedneho znaku vypočítané podľa vzťahu (4) a uvedené v Tabuľke 9.

Tabuľka 9

I_j	$\langle 0;1)$	$\langle 1;2)$	$\langle 2;3)$	$\langle 3;4)$	$\langle 4;5\rangle$
z_j	$0,5$	$1,5$	$2,5$	$3,5$	$4,5$
n_j	60	20	11	5	4

Potom pre parameter λ platí

$$\lambda \approx \overline{x} = \frac{1}{100}\left(0,5\cdot 60 + 1,5\cdot 20 + 2,5\cdot 11 + 3,5\cdot 5 + 4,5\cdot 4\right) = 1,23.$$

Budeme testovať

$$H_0 : F(x) = G(x),\ \text{kde } G(x) \sim Ex(\lambda),$$
$$H_1 : F(x) \neq G(x).$$

Testujeme, či sa distribučná funkcia $F(x)$ rozdelenia, z ktorého pochádza náhodný výber, rovná distribučnej funkcii $G(x)$ exponenciálneho rozdelenia. Hodnotu testovacieho kritéria χ^2 určíme podľa vzťahu (115) a k výpočtu jeho hodnoty použijeme Tabuľku 10.

Tabuľka 10

j	t_j	$G(t_j)$	$G(t_{j+1})$	p_j	$n\,p_j$	n_j
1	0	0	$0,5565$	$0,5565$	$55,65$	60
2	1	$0,5565$	$0,8033$	$0,2468$	$24,68$	20
3	2	$0,8033$	$0,9128$	$0,1095$	$10,95$	11
4	3	$0,9128$	$0,9613$	$0,0485$	$4,85$	5
5	4	$0,9613$	1	$0,0387$	**3,87**	4
6	∞	1	–	–	–	–

V 2. stĺpci Tabuľky 10 sú hranice t_j triednych intervalov, pričom dolná hranica prvého intervalu, t. j. 0, zostane v tomto prípade nezmenená a hornú hranicu posledného intervalu nahradíme ∞ (viď definičný obor distribučnej funkcie exponenciálneho rozdelenia).

V 3. stĺpci Tabuľky 10 sú hodnoty distribučnej funkcie exponenciálneho rozdelenia $G(t_j)$, ktoré vypočítame podľa vzťahu pre distribučnú funkciu exponenciálneho rozdelenia (60)

$$G(t_j) = 1 - e^{-\frac{t_j}{\lambda}} = 1 - e^{-\frac{t_j}{1,23}}.$$

V 5. stĺpci Tabuľky 10 sú teoretické pravdepodobnosti p_j, ktoré vypočítame na základe rozdielu hodnôt distribučnej funkcie, t. j. $p_j = G(t_{j+1}) - G(t_j)$, kde hodnoty $G(t_{j+1})$ sú uvedené v 4. stĺpci Tabuľky 10.

Pri používaní testu sa požaduje splnenie Cochranovho pravidla. V našom prípade je počet tried $k = 5$ a počet odhadovaných parametrov $r = 1$, teda $k - r - 1 = 5 - 1 - 1 = 3 \geqq 3$. Stačí požadovať podmienku $np_j \geqq 4$.

V 6. stĺpci v piatom riadku Tabuľky 10 podmienka nie je splnená. Zlúčime štvrtý a piaty riadok v 6. stĺpci Tabuľky 10, čím dostaneme počet tried $k = 4$ (Tabuľka 11). Počet odhadovaných parametrov $r = 1$, teda $k - r - 1 = 4 - 1 - 1 = 2$. Teraz musí platiť $np_j \geqq 5$, čo máme splnené ako vidíme v Tabuľke 11.

Tabuľka 11

j	t_j	$G(t_j)$	p_j	$n\,p_j$	n_j	$\frac{(n_j-np_j)^2}{np_j}$
1	0	0	0,5565	55,65	60	0,3400
2	1	0,5565	0,2468	24,68	20	0,8875
3	2	0,8033	0,1095	10,95	11	0,0002
4	3	0,9128	0,0485	8,72	9	0,0090
5	4	0,9613	0,0387			
6	∞	1	–	–	–	$\sum_{j=1}^{5} \frac{(n_j - np_j)^2}{np_j} = 1,2367$

Hodnota testovacieho kritéria je $\chi^2 = 1,2367$ a kvantil χ^2–rozdelenia je $\chi^2_{1-\alpha}(k - r - 1) = \chi^2_{0,95}(2) = 5,99$ (Tabuľka 30). Kritická oblasť je

$$K_{0,05} = (5,99; \infty) \Rightarrow \chi^2 \notin K_{0,05}.$$

Hodnota testovacieho kritéria $\chi^2 = 1,2367$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ náhodný výber pochádza z exponenciálneho rozdelenia, pričom ale nepoznáme jeho parameter λ.

5.5.2 Kolmogorovov–Smirnovov test

Kolmogorovov–Smirnovov test je neparametrický test, ktorý sa používa pri testovaní rozdelení náhodných premenných spojitého typu. Navyše rozdelenie musí byť dané jednoznačne, t. j. bez neznámych parametrov. Na rozdiel od Pearsonovho testu dobrej zhody, test sa môže použiť aj pre náhodný výber malého rozsahu (teda nepočíta s triedením do intervalov).

1. Hypotézy:

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. Testovacie kritérium je
$$D = \sup_{x \in R} |F_n(x) - G(x)|, \tag{116}$$
pričom $G(x)$ je uvažovaná teoretická distribučná funkcia spojitého rozdelenia, $F_n(x)$ je empirická distribučná funkcia náhodného výberu, pre ktorú platí
$$F_n(x) = \frac{1}{n} \sum_{x_j \leqq x} n_j, \tag{117}$$
kde n_j početnosť hodnoty x_j v náhodnom výbere a n je rozsah náhodného výberu. Vzhľadom na takto definovanú empirickú distribučnú funkciu $F_n(x)$, je možné testovacie kritérium D vypočítať zo vzťahu
$$D = \max\{\max(d_j^-); \max(d_j^+)\}, \tag{118}$$
kde $d_j^- = |F_n(x_{j-1}) - G(x_j)|$, $d_j^+ = |F_n(x_j) - G(x_j)|$ pre $j = 1, 2, \ldots, n$.

3. Kritická oblasť K_α je
$$K_\alpha = (D_\alpha(n); \infty)\,,$$
kde $D_\alpha(n)$ je tabuľková kritická hodnota pre Kolmogorovov–Smirnovov test (Tabuľka 47). Pre veľké rozsahy náhodného výberu ($n > 100$) sa kritické hodnoty $D_\alpha(n)$ aproximujú vzťahom
$$D_\alpha(n) \doteq \sqrt{\frac{1}{2n} \ln \frac{2}{\alpha}}. \tag{119}$$

Poznámka: V prípade, že všetky namerané hodnoty $x_1, x_2, \ldots, x_n$ sú navzájom rôzne, môžeme použiť špeciálny tvar empirickej distribučnej funkcie
$$F_n(x) = \begin{cases} 0 & \text{pre } -\infty < x < x_{(1)}, \\ \frac{i}{n} & \text{pre } x_{(i)} \leqq x < x_{(i+1)},\ i = 1, 2, \ldots, n-1, \\ 1 & \text{pre } x_{(n)} < x < \infty, \end{cases}$$
kde hodnoty náhodného výberu sú usporiadané do neklesajúcej postupnosti podľa veľkosti (variačného radu), t. j.
$$x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)}.$$

Príklad 5.13 *Na hladine významnosti α=0,05 overme, či náhodná premenná daná tabuľkou pochádza zo základného súboru s exponenciálnym rozdelením s parametrom λ=1,23. Použime pritom Kolmogorovov–Smirnovov test.*

x_j	*0,5*	*1,5*	*2,5*	*3,5*	*4,5*
n_j	*60*	*20*	*11*	*5*	*4*

Riešenie: Rozsah súboru je $n = 100$, parameter exponenciálneho rozdelenia je $\lambda = 1,23$. Budeme testovať

$$H_0 : F(x) = G(x), \text{ kde } G(x) \sim Ex(1,23),$$
$$H_1 : F(x) \neq G(x).$$

Testujeme, či sa distribučná funkcia $F(x)$ rozdelenia, z ktorého pochádza náhodný výber, rovná distribučnej funkcii $G(x)$ exponenciálneho rozdelenia s parametrom $\lambda = 1,23$. Hodnotu testovacieho kritéria D určíme podľa vzťahu (118). K výpočtu jeho hodnoty použijeme Tabuľku 12.

Tabuľka 12

j	x_j	n_j	$G(x_j)$	$F_n(x_{j-1})$	$F_n(x_j)$	d_j^-	d_j^+
1	0,5	60	0,3340	0,00	0,60	**0,3340**	**0,2660**
2	1,5	20	0,7046	0,60	0,80	0,1046	0,0954
3	2,5	11	0,8600	0,80	0,91	0,0690	0,0410
4	3,5	5	0,9419	0,91	0,96	0,0319	0,0181
5	4,5	4	0,9742	0,96	1,00	0,0142	0,0258

Vo 4. stĺpci Tabuľky 12 sú hodnoty $G(x_j)$ distribučnej funkcie exponenciálneho rozdelenia, ktoré vypočítame podľa vzťahu pre distribučnú funkciu exponenciálneho rozdelenia (60)

$$G(x_j) = 1 - e^{-\frac{x_j}{\lambda}} = 1 - e^{-\frac{x_j}{1,23}}.$$

V 5. a 6. stĺpci Tabuľky 12 sú hodnoty empirickej distribučnej funkcie náhodného výberu $F_n(x_{j-1})$ a $F_n(x_j)$ vypočítané podľa vzťahu (117).

V 7. stĺpci Tabuľky 12 sú hodnoty d_j^-, kde $d_j^- = |F_n(x_{j-1}) - G(x_j)|$ a v 8. stĺpci Tabuľky 12 sú hodnoty d_j^+, kde $d_j^+ = |F_n(x_j) - G(x_j)|$.

Hodnota testovacieho kritéria je najväčšie číslo z posledných dvoch stĺpcov, t. j. $D = \max\{0,3340; 0,2660\} = 0,3340$.

Kritickú hodnotu $D_\alpha(n)$ vypočítame podľa vzťahu (119). Platí $D_\alpha(n) = D_{0,05}(100) = 0,1358$ (Tabuľka 47). Kritická oblasť je

$$K_{0,05} = (D_\alpha(n); \infty) = (0,1358; \infty) \Rightarrow D \in K_{0,05}.$$

Hodnota testovacieho kritéria $D = 0,334$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ náhodný výber nepochádza z exponenciálneho rozdelenia s parametrom $\lambda = 1,23$. To ale neznamená, že nepochádza z exponenciálneho rozdelenia s iným parametrom.

5.5.3 Andersonov–Darlingov test

Andersonov–Darlingov test umožňuje testovať zhodu empirickej distribučnej funkcie s teoretickou distribučnou funkciou normálneho, exponenciálneho, lognormálneho alebo Weibullovho rozdelenia. Na rozdiel od predchádzajúceho testu, kritické hodnoty sú závislé od teoretického rozdelenia.

1. Hypotézy:

$$H_0 : F(x) = G(x),$$
$$H_1 : F(x) \neq G(x).$$

2. Testovacie kritérium je

$$AD = \frac{1}{n}\sum_{i=1}^{n}(1-2i)\left[\ln G(x_{(i)}) + \ln\left(1 - G(x_{(n+1-i)})\right)\right] - n, \tag{120}$$

kde n je rozsah náhodného výberu. Podmienkou je, hodnoty náhodného výberu usporiadať do neklesajúcej postupnosti podľa veľkosti (variačného radu), t. j.

$$x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)}.$$

3. Kritická oblasť K_α je

$$K_\alpha = (AD_\alpha; \infty)\,,$$

kde AD_α je tabuľková kritická hodnota testu.

V prípade, že teoretická distribučná funkcia $G(x)$ pochádza z normálneho rozdelenia pri malom rozsahu výberu ($n \leqq 10$), hodnotu AD_α určíme podľa vzťahu

$$AD_\alpha \approx \frac{k(\alpha)}{1 + \frac{0{,}75}{n} + \frac{2{,}25}{n^2}}, \tag{121}$$

kde $k(\alpha)$ je konštanta pre hladinu významnosti α uvedená v Tabuľke 13.

Tabuľka 13: Hodnoty konštanty $k(\alpha)$ pre rôzne hladiny významnosti α

α	$0,005$	$0,01$	$0,0025$	$0,05$	$0,1$
$k(\alpha)$	$1,159$	$1,035$	$0,873$	$0,752$	$0,631$

V prípade, že teoretická distribučná funkcia $G(x)$ pochádza z exponenciálneho rozdelenia, hodnota AD_α je pre hladinu významnosti α je uvedená v Tabuľke 14.

Tabuľka 14: Hodnoty AD_α pre rôzne hladiny významnosti α

α	$0,005$	$0,01$	$0,0025$	$0,05$	$0,1$
AD_α	$2,244$	$1,959$	$1,591$	$1,321$	$1,162$

Poznámka: V prípade, že teoretická distribučná funkcia $G(x)$ pochádza z iného ako z normálneho alebo z exponenciálneho rozdelenia, hodnoty AD_α pre danú hladinu významnosti α sú dostupné na internete alebo v inej odbornej literatúre. My ich uvádzať nebudeme, nakoľko pri Andersonov–Darlingovom teste sa týmito rozdeleniami nebudeme zaoberať .

Príklad 5.14 *Pomocou Andersonov–Darlingovho testu na hladine významnosti α=0,05 overme, či hodnoty 10, 20, 35, 4, 30, 1 náhodného výberu pochádzajú z exponenciálneho rozdelenia.*

Riešenie: Rozsah súboru je $n = 6$. Na odhad parametra λ použijeme výberový aritmetický priemer $\overline{x}$, ktorý vypočítame podľa vzťahu (71). Teda

$$\lambda = \overline{x} = \frac{1}{6}(10 + 20 + 35 + 4 + 30 + 1) = 16,67.$$

Budeme testovať

$$H_0 : F(x) = G(x), \text{ kde } G(x) \sim Ex(\lambda),$$
$$H_1 : F(x) \neq G(x).$$

Testujeme, či sa distribučná funkcia $F(x)$ rozdelenia, z ktorého pochádza náhodný výber, rovná distribučnej funkcii $G(x)$ exponenciálneho rozdelenia. Hodnotu testovacieho kritéria AD určíme podľa vzťahu (120). K výpočtu jeho hodnoty použijeme Tabuľku 15, v ktorej hodnoty náhodného výberu usporiadame do neklesajúcej postupnosti (variačného radu).

Tabuľka 15

(i)	$x_{(i)}$	$G(x_{(i)})$	$\ln G(x_{(i)})$	$\ln\left[1 - G(x_{(n+1-i)})\right]$	$\ln G(x_{(i)}) + \ln\left[1 - G(x_{(n+1-i)})\right]$
1	1	$0,0582$	$-2,8433$	$-2,0996$	$-4,9432$
2	4	$0,2134$	$-1,5447$	$-1,8000$	$-3,3447$
3	10	$0,4512$	$-0,7959$	$-1,2000$	$-1,9959$
4	20	$0,6988$	$-0,3584$	$-0,6000$	$-0,9584$
5	30	$0,8347$	$-0,1807$	$-0,2400$	$-0,4207$
6	35	$0,8775$	$-0,1306$	$-0,0600$	$-0,1906$

V 2. stĺpci Tabuľky 15 sú hodnoty náhodného výberu usporiadané do neklesajúcej postupnosti.

V 3. stĺpci Tabuľky 15 sú hodnoty $G(x_{(i)})$ distribučnej funkcie exponenciálneho rozdelenia, ktoré vypočítame podľa vzťahu pre distribučnú funkciu exponenciálneho rozdelenia (60)

$$G(x_{(i)}) = 1 - e^{-\frac{x_{(i)}}{\lambda}} = 1 - e^{-\frac{x_{(i)}}{16,67}}.$$

V 4. stĺpci Tabuľky 15 sú hodnoty $\ln G(x_{(i)})$ a v 5. stĺpci sú hodnoty $\ln\left[1 - G(x_{(n+1-i)})\right]$.

V 6. stĺpci Tabuľky 15 je súčet hodnôt 4. a 5. stĺpca, t. j. $\ln G(x_{(i)}) + \ln\left[1 - G(x_{(n+1-i)})\right]$.

Hodnota testovacieho kritéria AD je určená podľa vzťahu (120)

$$AD = \frac{1}{6}\Big[-1 \cdot (-4,9432) - 3 \cdot (-3,3447) - 5 \cdot (-1,9959) - 7 \cdot (-0,9584) - \\ -9 \cdot (-0,4207) - 11 \cdot (-0,1906)\Big] - 6 = 0,2580.$$

Kritická hodnota testu $AD_\alpha = AD_{0,05} = 1,321$ (Tabuľka 14).

Kritická oblasť je

$$K_{0,05} = (AD_\alpha; \infty) = (1,321; \infty) \Rightarrow AD \notin K_{0,05}.$$

Hodnota testovacieho kritéria $AD = 0,2580$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ náhodný výber pochádza z exponenciálneho rozdelenia.

5.5.4 Shapiro–Wilkov test normality

Shapiro–Wilkov test normality je najčastejšie používaným testom normality. Je založený na zisťovaní skutočnosti, či sa body zostrojeného kvantil-kvantilového grafu (Q–Q graf alebo ang. Q–Q plot) významne líšia od regresnej priamky preloženej týmito bodmi. Používa sa predovšetkým pre náhodné výbery rozsahov $7 \leqq n \leqq 2000$. Uvažujme náhodný výber V_n, ktorého hodnoty usporiadame podľa veľkosti do neklesajúcej postupnosti (variačného radu)

$$x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)}.$$

1. Hypotézy:

$$H_0 : F(x) = G(x), \text{ kde } G(x) \sim N(\mu, \sigma^2),$$
$$H_1 : F(x) \neq G(x).$$

2. Testovacie kritérium je

$$W = \frac{\left(\sum\limits_{i=1}^{m} a_{i,n}\left(x_{(n-i+1)} - x_{(i)}\right)\right)^2}{\sum\limits_{i=1}^{n}\left(x_{(i)} - \overline{x}\right)^2}, \tag{122}$$

kde $a_{i,n}$ sú tabuľkové váhy (Tabuľka 48), $\overline{x}$ je výberový aritmetický priemer a $m = \frac{n}{2}$ pre n párne, resp. $m = \frac{n-1}{2}$ pre n nepárne.

3. Kritická oblasť K_α je
$$K_\alpha = (-\infty; W_\alpha(n)),$$
kde $W_\alpha(n)$ je tabuľková kritická hodnota pre Shapiro–Wilkov test normality (Tabuľka 49).

Poznámka: Čím viac sa hodnota testovacieho kritéria W blíži k 1, tým je zhoda medzi teoretickým a empirickým rozdelením lepšia.

Príklad 5.15 *Pomocou Shapiro–Wilkovho testu normality na hladine významnosti α=0,05 overme, či hodnoty 4,05; 4,15; 4,2; 4,25; 4,3; 4,35; 4,45; 4,55 náhodného výberu pochádzajú z normálneho rozdelenia.*

Riešenie: Budeme testovať
$$H_0 : F(x) = G(x), \text{ kde } G(x) \sim N(\mu, \sigma^2),$$
$$H_1 : F(x) \neq G(x).$$
Testujeme, či sa distribučná funkcia $F(x)$ rozdelenia, z ktorého pochádza náhodný výber, rovná distribučnej funkcii $G(x)$ normálneho rozdelenia. Rozsah súboru je $n = 8$. Všetky hodnoty sú usporiadané do neklesajúcej postupnosti. Hodnotu testovacieho kritéria W určíme podľa vzťahu (122), kde $m = \frac{n}{2} = 4$, výberový aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (71)
$$\overline{x} = \frac{1}{8}\big(4,05 + 4,15 + 4,2 + 4,25 + 4,3 + 4,3 + 4,45 + 4,55\big) = 4,2875$$
a tabuľkové hodnoty $a_{i,n}$ sú $a_{1,8} = 0,6052$, $a_{2,8} = 0,3164$, $a_{3,8} = 0,1743$, $a_{4,8} = 0,0561$ (Tabuľka 48). Potom hodnota testovacieho kritéria W je
$$W = \frac{[0,6052(4,55-4,05) + 0,3164(4,45-4,15) + 0,1743(4,35-4,2) + 0,0561(4,3-4,25)]^2}{\sum\limits_{i=1}^{n}\left(x_{(i)} - \overline{x}\right)^2},$$
kde
$$\sum_{i=1}^{n}\left(x_{(i)} - \overline{x}\right)^2 = (4,05-4,2875)^2 + (4,15-4,2875)^2 + (4,2-4,2875)^2 + (4,25-4,2875)^2 +$$
$$+(4,3-4,2875)^2 + (4,35-4,2875)^2 + (4,45-4,2875)^2 + (4,55-4,2875)^2 = 0,1838.$$
Teda $W = 0,99$. Kritická hodnota pre Shapiro–Wilkov test normality je $W_{0,05}(8) = 0,818$ (Tabuľka 49). Kritická oblasť je
$$K_{0,05} = (-\infty; 0,818) \Rightarrow W \notin K_{0,05}.$$

Hodnota testovacieho kritéria $W = 0,99$ nepatrí do kritickej oblasti $K_{0,05}$, preto na hladine významnosti $\alpha = 0,05$ nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že hodnoty náhodného výberu pochádzajú z normálneho rozdelenia.

5.6 Testy extrémnych hodnôt

Niekedy sa medzi nameranými hodnotami vyskytujú hodnoty, ktoré sa značne odlišujú od ostatných. Takéto hodnoty nazývame *extrémne (odľahlé) hodnoty*. Extrémne hodnoty skresľujú vypočítané charakteristiky náhodného výberu, čo môže viesť ku chybnému záveru analýzy daného súboru. Preto je dobré ich detekovať a následne vylúčiť z analyzovaného súboru. Ku grafickému nájdeniu extrémnych hodnôt slúži krabicový diagram (boxplot) (kapitola 2.2). Aby bolo možné extrémnu hodnotu z analyzovaného súboru vylúčiť, je potrebné štatisticky otestovať jej odchýlky od ostatných hodnôt. Medzi najčastejšie používané testy extrémnych hodnôt patria Grubbsov a Dixonov test. Oba predpokladajú splnenie normality rozdelenia základného súboru a dajú sa použiť pri malom rozsahu náhodného výberu ($n \leqq 30$).

Uvažujme náhodný výber V_n, ktorý pochádza zo základného súboru s normálnym rozdelením so strednou hodnotou μ a smerodajnou odchýlkou σ. Hodnoty náhodného výberu usporiadame podľa veľkosti do neklesajúcej postupnosti (variačného radu)

$$x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n-1)} \leqq x_{(n)}.$$

5.6.1 Grubbsov test

Grubbsov test má dva tvary podľa toho, či testujeme extrémnosť minimálnej alebo maximálnej hodnoty.

Alternatíva pre testovanie extrémnosti minimálnej hodnoty:

1. Hypotézy:

$$H_0 : \text{hodnota } x_{(1)} \text{ nie je extrémna,}$$
$$H_1 : \text{hodnota } x_{(1)} \text{ je extrémna.}$$

2. Testovacie kritérium je

$$T(1) = \frac{\overline{x} - x_{(1)}}{s}\sqrt{\frac{n}{n-1}}, \tag{123}$$

kde $\overline{x}$ je výberový aritmetický priemer, s je výberová smerodajná odchýlka a n je rozsah náhodného výberu.

3. Kritická oblasť K_α je

$$K_\alpha = (T_\alpha(n); \infty)\,,$$

kde $T_\alpha(n)$ je kritická hodnota Grubbsovho testu tabelovaná pre $n \leqq 30$ pre $\alpha = 0,05$ a $\alpha = 0,01$ (Tabuľka 50).

Alternatíva pre testovanie extrémnosti maximálnej hodnoty:

1. Hypotézy:

$$H_0 : \text{hodnota } x_{(n)} \text{ nie je extrémna,}$$
$$H_1 : \text{hodnota } x_{(n)} \text{ je extrémna.}$$

2. Testovacie kritérium je

$$T(n) = \frac{x_{(n)} - \overline{x}}{s}\sqrt{\frac{n}{n-1}}, \tag{124}$$

kde $\overline{x}$ je výberový aritmetický priemer, s je výberová smerodajná odchýlka a n je rozsah náhodného výberu.

3. Kritická oblasť K_α je

$$K_\alpha = (T_\alpha(n); \infty),$$

kde $T_\alpha(n)$ je kritická hodnota Grubbsovho testu tabelovaná pre $n \leqq 30$ pre $\alpha = 0,05$ a $\alpha = 0,01$ (Tabuľka 50).

5.6.2 Dixonov test

Podobne ako Grubbsov test, aj *Dixonov test (Q-test)* má dva tvary podľa toho, či testujeme extrémnosť minimálnej alebo maximálnej hodnoty. V tomto prípade nie je potrebné počítať výberový aritmetický priemer a výberovú smerodajnú odchýlku náhodného výberu.

Alternatíva pre testovanie extrémnosti minimálnej hodnoty:

1. Hypotézy:

$$H_0 : \text{hodnota } x_{(1)} \text{ nie je extrémna,}$$
$$H_1 : \text{hodnota } x_{(1)} \text{ je extrémna.}$$

2. Testovacie kritérium je

$$Q(1) = \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}. \tag{125}$$

3. Kritická oblasť K_α je

$$K_\alpha = (Q_\alpha(n); \infty),$$

kde $Q_\alpha(n)$ je kritická hodnota Dixonovho testu tabelovaná pre $n \leqq 30$ pre $\alpha = 0,05$ a $\alpha = 0,01$ (Tabuľka 51).

Alternatíva pre testovanie extrémnosti maximálnej hodnoty:

1. Hypotézy:

$$H_0: \text{hodnota } x_{(n)} \text{ nie je extrémna,}$$
$$H_1: \text{hodnota } x_{(n)} \text{ je extrémna.}$$

2. Testovacie kritérium je

$$Q(n) = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}. \tag{126}$$

3. Kritická oblasť K_α je

$$K_\alpha = (Q_\alpha(n); \infty),$$

kde $Q_\alpha(n)$ je kritická hodnota Dixonovho testu tabelovaná pre $n \leqq 30$ pre $\alpha = 0,05$ a $\alpha = 0,01$ (Tabuľka 51).

Príklad 5.16 *Medzi hodnotami získanými z určitého merania 5,83; 5,80; 5,85; 5,88; 5,84; 5,83; 5,96; 5,78; 5,82; 5,81; 5,86; 5,82 preverme na hladine významnosti α=0,05, či hodnota 5,96 nie je extrémna. Predpokladáme, že hodnoty sa riadia normálnym rozdelením. Použime pritom*
a) Grubbsov test,
b) Dixonov test.

Riešenie: Všetky hodnoty usporiadame do neklesajúcej postupnosti: 5, 78; 5, 80; 5, 81; 5, 82; 5, 82; 5, 83; 5, 83; 5, 84; 5, 85; 5, 86; 5, 88; 5, 96. Pre rýchle overenie, či ide o extrémnu hodnotu, môžeme použiť aj boxplot (Obr. 24).

V tomto prípade ideme testovať extrémnosť maximálnej hodnoty. Rozsah súboru je $n = 12$. Budeme testovať

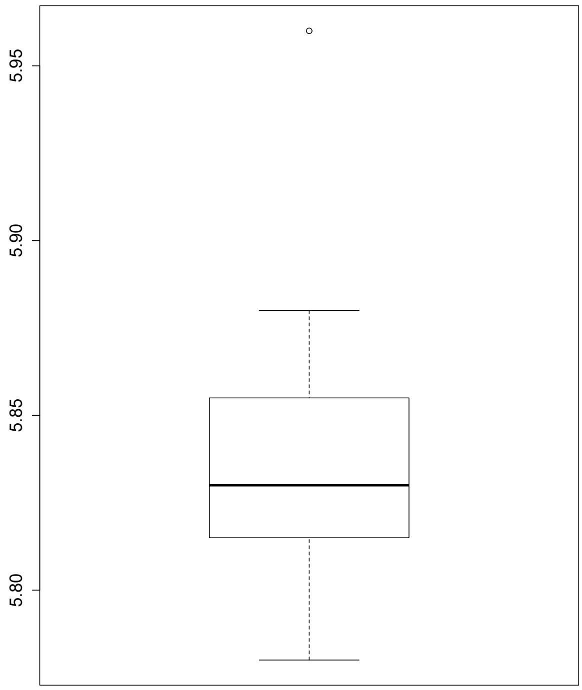
$$H_0: \text{hodnota } x_{(12)} = 5,96 \text{ nie je extrémna,}$$
$$H_1: \text{hodnota } x_{(12)} = 5,96 \text{ je extrémna.}$$

a) Hodnotu testovacieho kritéria $T(12)$ pre Grubbsov test určíme podľa vzťahu (124)

$$T(12) = \frac{x_{(12)} - \overline{x}}{s}\sqrt{\frac{12}{12-1}},$$

kde výberový aritmetický priemer $\overline{x}$ vypočítame podľa vzťahu (71)

$$\overline{x} = \frac{1}{12}(5,78 + 5,80 + 5,81 + 5,82 + 5,82 + 5,83 + 5,83+$$
$$+5,84 + 5,85 + 5,86 + 5,88 + 5,96) = 5,84$$



Obr. 24: Boxplot pre Príklad 5.16

a výberovú smerodajnú odchýlku s vypočítame podľa vzťahu (77). Pre výberový rozptyl s^2 platí (73)

$$s^2 = \frac{1}{11}\big[(5,78-5,84)^2 + (5,80-5,84)^2 + (5,81-5,84)^2 + 2.(5,82-5,84)^2 +$$
$$+2.(5,83-5,84)^2 + (5,84-5,84)^2 + (5,85-5,84)^2 + (5,86-5,84)^2 +$$
$$+(5,88-5,84)^2 + (5,96-5,84)^2\big] = 0,00214 \Rightarrow s = \sqrt{s^2} = 0,0463.$$

Hodnota testovacieho kritéria pre Grubbsov test je $T(12) = 2,7059$. Kritická hodnota je $T_{0,05}(12) = 2,387$ (Tabuľka 50). Kritická oblasť je

$$K_{0,05} = (2,387;\infty) \Rightarrow T(12) \in K_{0,05}.$$

Hodnota testovacieho kritéria $T(12) = 2,7059$ patrí do kritickej oblasti $K_{0,05}$, preto na hladine významnosti $\alpha = 0,05$ nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Na základe Grubbsovho testu môžeme predpokladať, že hodnota $x_{(12)} = 5,96$ je extrémna.

b) Hodnotu testovacieho kritéria $Q(12)$ pre Dixonov test určíme podľa vzťahu (126)

$$Q(12) = \frac{x_{(12)} - x_{(11)}}{x_{(12)} - x_{(1)}} = \frac{5,96-5,88}{5,96-5,78} = 0,4444.$$

Kritická hodnota Dixonovho testu je $Q_{0,05}(12) = 0,376$ (Tabuľka 51). Kritická oblasť je

$$K_{0,05} = (0,376; \infty) \Rightarrow Q(12) \in K_{0,05}.$$

Hodnota testovacieho kritéria $Q(12) = 0,4444$ patrí do kritickej oblasti $K_{0,05}$, preto na hladine významnosti $\alpha = 0,05$ nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Na základe Dixonovho testu môžeme predpokladať, že hodnota $x_{(12)} = 5,96$ je extrémna.

6 Korelačná analýza

Korelačná analýza sa zaoberá vzájomnými závislosťami medzi náhodnými premennými, pričom sa overuje predovšetkým kvalita (sila, tesnosť) ich vzájomného vzťahu. Základným prípadom štatistickej závislosti je tzv. *jednoduchá závislosť*, t. j. závislosť len medzi dvoma náhodnými premennými X a Y.

Ak uvažujeme závislosť náhodnej premennej Y na premenných X_1, X_2, ..., X_n, $n \geqq 2$, hovoríme o *viacnásobnej (mnohonásobnej) závislosti*. My sa budeme zaoberať len jednoduchou závislosťou náhodných premenných z hľadiska tesnosti (kvality) väzby medzi nimi, t. j. z hľadiska *korelácie*.

6.1 Pearsonov výberový korelačný koeficient

Uvažujme dve náhodné premenné X a Y, ktorých strednú hodnotu označíme $E(X)$, resp. $E(Y)$ a disperziu $D(X)$, resp. $D(Y)$. *Kovariancia* $\operatorname{cov}(X,Y)$ náhodných premenných X a Y je reálne číslo, pre ktoré platí nasledujúci vzťah

$$\operatorname{cov}(X,Y) = E(X \cdot Y) - E(X)E(Y). \tag{127}$$

Korelačný koeficient ρ ($\rho(X,Y)$) spojitých náhodných premenných X a Y je špeciálnym prípadom kovariancie dvoch náhodných premenných a je definovaný vzťahom

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\,\sqrt{D(Y)}}. \tag{128}$$

Korelačný koeficient ρ charakterizuje len lineárny vzťah dvoch náhodných premenných X, Y a nadobúda hodnoty z intervalu $\langle -1; 1\rangle$. Platí

- Ak $\rho = 0$, tak náhodné premenné X a Y sú nekorelované (lineárne nezávislé).
- Ak $\rho < 0$, tak náhodné premenné X a Y sú negatívne korelované (nepriama lineárna závislosť), zväčšovanie (zmenšovanie) jednej náhodnej premennej vedie v priemere k zmenšovaniu (zväčšovaniu) druhej náhodnej premennej.
- Ak $\rho > 0$, tak náhodné premenné X a Y sú pozitívne korelované (priama lineárna závislosť), zväčšovanie (zmenšovanie) jednej náhodnej premennej vedie v priemere k zväčšovaniu (zmenšovaniu) druhej náhodnej premennej.

Poznámka: V praktickej štatistike sa veľmi často nerobí rozdiel medzi pojmami nezávislé a nekorelované náhodné premenné, avšak korektné je to len pre náhodné premenné z normálneho rozdelenia.

Keďže sa v praxi intervalový odhad pre korelačný koeficient ρ často nepoužíva, budeme sa zaoberať len jeho bodovým odhadom – Pearsonovým výberovým korelačným koeficientom r_{xy}.

Uvažujme náhodný výber V_n s nameranými hodnotami $x_1, x_2, \ldots, x_n$ náhodnej premennej X a im zodpovedajúce hodnoty $y_1, y_2, \ldots, y_n$ náhodnej premennej Y. Potom je náhodný výber V_n z usporiadanej dvojice náhodných premenných (X, Y) reprezentovaný usporiadanými dvojicami $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Pearsonov výberový korelačný koeficient r_{xy} náhodných premenných X a Y je definovaný vzťahom

$$r_{xy} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2} - (\overline{x})^2}\,\sqrt{\overline{y^2} - (\overline{y})^2}}, \tag{129}$$

kde $\overline{x}$, $\overline{y}$ sú výberové aritmetické priemery náhodných premenných X a Y, pričom

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \qquad \overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2, \qquad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i\, y_i. \tag{130}$$

Poznámka: Priamo zo vzťahu (129) vyplýva jeho symetrickosť, t. j. $r_{xy} = r_{yx}$.

Pearsonov výberový korelačný koeficient r_{xy} nadobúda hodnoty z intervalu $\langle -1; 1\rangle$. Čím je hodnota $|r_{xy}|$ bližšia k 1, tým indikuje silnejšiu lineárnu korelačnú závislosť. Naopak čím je hodnota $|r_{xy}|$ bližšia k 0, tým indikuje slabšiu lineárnu korelačnú závislosť. Ak je $r_{xy} = 0$, tak hovoríme, že lineárna závislosť medzi náhodnými premennými X a Y neexistuje (môže však existovať nelineárna).

Keďže však ide len o výberovú charakteristiku, nevieme jednoznačne povedať, či korelačný koeficient ρ základného súboru je nulový. Preto to otestujeme pomocou *testu nekorelovanosti.*

Test nekorelovanosti

1. Hypotézy:

$$H_0 : \rho = 0,$$
$$H_1 : \rho \neq 0 \quad (\text{resp. } \rho < 0, \text{ resp. } \rho > 0).$$

resp.

H_0 : korelačný koeficient ρ nie je štatisticky významný,

H_1 : korelačný koeficient ρ je štatisticky významný.

2. Testovacie kritérium je

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1 - r_{xy}^2}} \sim t(n-2), \tag{131}$$

kde r_{xy} je Pearsonov výberový korelačný koeficient a n je rozsah náhodného výberu.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\rho = 0$	$\rho \neq 0$	$\left(-\infty; -t_{1-\frac{\alpha}{2}}(n-2)\right) \cup \left(t_{1-\frac{\alpha}{2}}(n-2); \infty\right)$
$\rho = 0$	$\rho > 0$	$(t_{1-\alpha}(n-2); \infty)$
$\rho = 0$	$\rho < 0$	$(-\infty; -t_{1-\alpha}(n-2))$

Príklad 6.1 *V nasledujúcej tabuľke sú údaje o množstve úrody jabĺk (Y) [v t] a počte stromov v sade (X) [v ks].*

x_i	*160*	*250*	*320*	*500*	*750*	*1000*	*1500*	*2000*
y_i	*789*	*800*	*851*	*874*	*1193*	*1335*	*1704*	*2073*

Predpokladáme, že medzi sledovanými premennými je lineárny vzťah a že ide o náhodný výber z normálneho rozdelenia.
a) Vypočítajme Pearsonov výberový korelačný koeficient.
b) Na hladine významnosti α=0,05 posúďme, či korelačný koeficient je štatisticky významný.
c) Na hladine významnosti α=0,05 posúďme, či množstvo stromov a množstvo úrody sú pozitívne korelované.

Riešenie: a) Pearsonov výberový korelačný koeficient r_{xy} vypočítame podľa vzťahu (129), kde výberové aritmetické priemery $\overline{x}$, $\overline{y}$ náhodných premenných X a Y vypočítame podľa vzťahu (71). Platí

$$\overline{x} = \frac{1}{8}(160 + 250 + 320 + 500 + 750 + 1000 + 1500 + 2000) = 810,$$

$$\overline{y} = \frac{1}{8}(789 + 800 + 851 + 874 + 1193 + 1335 + 1704 + 2073) = 1202,375.$$

Hodnoty $\overline{x^2}$, $\overline{y^2}$, $\overline{xy}$ vypočítame podľa vzťahu (130). Platí

$$\overline{x^2} = \frac{1}{8}\left(160^2 + 250^2 + 320^2 + 500^2 + 750^2 + 1000^2 + 1500^2 + 2000^2\right) = 1031625,$$

$$\overline{y^2} = \frac{1}{8}\left(789^2 + 800^2 + 851^2 + 874^2 + 1193^2 + 1335^2 + 1704^2 + 2073^2\right) = 1644627,$$

$$\overline{xy} = \frac{1}{8}\big(160 \cdot 789 + 250 \cdot 800 + 320 \cdot 851 + 500 \cdot 874 + 750 \cdot 1193 + \\ +1000 \cdot 1335 + 1500 \cdot 1704 + 2000 \cdot 2073\big) = 1245914.$$

Pearsonov výberový korelačný koeficient je $r_{xy} = 0,995$, teda medzi premennými X a Y je silná priama lineárna korelačná závislosť.

b) Budeme testovať

$$H_0 : \rho = 0,$$
$$H_1 : \rho \neq 0.$$

Nulová hypotéza hovorí, že korelačný koeficient ρ nie je štatistický významný. Proti obojstrannej alternatívnej hypotéze, ktorá hovorí, že korelačný koeficient ρ je štatistický významný. Hodnotu testovacieho kritéria t určíme podľa vzťahu (131)

$$t = \frac{0,995 \cdot \sqrt{8-2}}{\sqrt{1-0,995^2}} = 24,8046.$$

Kvantil Studentovho t–rozdelenia je $t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(6) = 2,4469$ (Tabuľka 29). Kritická oblasť je

$$K_{0,05} = (-\infty; -2,4469) \cup (2,4469; \infty) \Rightarrow t \in K_{0,05}.$$

Hodnota testovacieho kritéria $t = 24,8046$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ je korelačný koeficient ρ štatisticky významný.

c) Budeme testovať

$$H_0 : \rho = 0,$$
$$H_1 : \rho > 0,$$

Nulová hypotéza hovorí, že množstvo stromov a množstvo úrody nie sú korelované. Proti obojstrannej alternatívnej hypotéze, ktorá hovorí, že množstvo stromov a množstvo úrody sú pozitívne korelované. Hodnotu testovacieho kritéria t vypočítame rovnako ako v prípade b) a teda $t = 24,8046$. Kvantil Studentovho t–rozdelenia $t_{1-\alpha}(n-2) = t_{0,95}(6) = 1,9432$ (Tabuľka 29). Kritická oblasť je

$$K_{0,05} = (1,9432; \infty) \Rightarrow t \in K_{0,05}.$$

Hodnota testovacieho kritéria $t = 24,8046$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ množstvo stromov a množstvo úrody sú pozitívne korelované.

6.2 Spearmanov koeficient poradovej korelácie

Spearmanov koeficient poradovej korelácie r_S patrí medzi neparametrické miery korelácie. Na rozdiel od Pearsonovho výberového korelačného koeficientu r_{xy}, nie je koeficient r_S obmedzený na lineárnu závislosť a je odolný voči extrémnym (odľahlým) hodnotám. Avšak jeho nevýhodou je, že zavedením poradia hodnôt náhodných premenných strácame informácie o pôvodných pozorovaniach. Na druhej strane ho vďaka tomu môžeme použiť aj na zisťovanie závislosti dvoch poradových znakov (nadobúdajú diskrétne hodnoty a vieme ich jednoznačne usporiadať, napr. stupeň spokojnosti zákazníka).

Pozorované hodnoty x_i, resp. y_i zoradíme do samostatných variačných radov. Potom nahradíme hodnoty x_i premennej X poradovými číslami p_i a hodnoty y_i premennej Y poradovými číslami q_i. *Spearmanov koeficient poradovej korelácie* r_S náhodných premenných X a Y je definovaný vzťahom

$$r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (p_i - q_i)^2, \tag{132}$$

kde n je rozsah náhodného výberu.

Spearmanov koeficient poradovej korelácie r_S nadobúda hodnoty z intervalu $\langle -1; 1\rangle$. Podobne ako v prípade Pearsonovho výberového korelačného koeficientu, hodnoty $|r_S|$ bližšie k 1 ukazujú silnejšiu závislosť a naopak hodnoty $|r_S|$ bližšie k 0 indikujú slabšiu závislosť.

V náhodnom výbere sa môžu niektoré hodnoty aj opakovať. V takom prípade týmto hodnotám priradíme priemerné poradové číslo a použijeme *korigovaný Spearmanov koeficient poradovej korelácie* r_S^* definovaný vzťahom

$$r_S^* = 1 - \frac{6}{n(n^2-1) - c} \sum_{i=1}^{n} (p_i - q_i)^2, \text{ pričom } c = \sum \frac{(c_x^3 - c_x)}{2} + \sum \frac{(c_y^3 - c_y)}{2}, \tag{133}$$

kde c je opravný člen, c_x počet rovnakých hodnôt premennej X a c_y počet rovnakých hodnôt premennej Y.

Štatistickú významnosť Spearmanovho koeficientu poradovej korelácie r_S otestujeme *Spearmanovým testom nezávislosti.*

Spearmanov test nezávislosti

1. Hypotézy:

$$H_0: r_S = 0,$$
$$H_1: r_S \neq 0,$$

   resp.

   $H_0:$ Spearmanov koeficient poradovej korelácie r_S nie je štatisticky významný,

   $H_1:$ Spearmanov koeficient poradovej korelácie r_S je štatisticky významný.

2. Testovacie kritérium je

$$t = \frac{r_S\sqrt{n-2}}{\sqrt{1-r_S^2}} \sim t(n-2), \tag{134}$$

   kde r_S je Spearmanov koeficient poradovej korelácie a n je rozsah náhodného výberu.

3. Kritická oblasť K_α je

$$K_\alpha = \left(-\infty; -t_{1-\frac{\alpha}{2}}(n-2)\right) \cup \left(t_{1-\frac{\alpha}{2}}(n-2); \infty\right).$$

V prípade malého rozsahu náhodného výberu ($n \leqq 30$), môžeme použiť aj nasledujúci test.

1. Hypotézy:

$$H_0 : r_S = 0,$$
$$H_1 : r_S \neq 0,$$

   resp.

   H_0 : Spearmanov koeficient poradovej korelácie r_S nie je štatisticky významný,

   H_1 : Spearmanov koeficient poradovej korelácie r_S je štatisticky významný.

2. Testovacie kritérium je hodnota Spearmanovho koeficientu poradovej korelácie r_S.

3. Kritická oblasť K_α je

$$K_\alpha = (-\infty; -r_\alpha(n)) \cup (r_\alpha(n); \infty),$$

   kde $r_\alpha(n)$ je kritická hodnota tabelovaná pre $n \leqq 30$ a pre zvolenú hladinu významnosti α (Tabuľka 52).

Príklad 6.2 *Dvaja učitelia bodovo hodnotili testy u 6 študentov. Výsledky sú uvedené v tabuľke.*

Študent	*Š1*	*Š2*	*Š3*	*Š4*	*Š5*	*Š6*
Učiteľ A	*12*	*5*	*8*	*47*	*23*	*10*
Učiteľ B	*15*	*5*	*18*	*39*	*28*	*11*

a) Vypočítajme Spearmanov koeficient poradovej korelácie.
b) Na hladine významnosti α=0,05 posúďme, či Spearmanov koeficient poradovej korelácie je štatisticky významný.

Riešenie: a) K hodnoteniu učiteľov pridáme poradie od najnižšieho počtu bodov po najvyšší. Pri učiteľovi A označíme poradie p_i, pri učiteľovi B označíme poradie q_i (Tabuľka 16).

Hodnoty v jednotlivých náhodných výberoch sa neopakujú, preto Spearmanov koeficient poradovej korelácie r_S vypočítame podľa vzťahu (132). Platí

$$r_S = 1 - \frac{6}{6\cdot(36-1)}\left[(4-3)^2+(1-1)^2+(2-4)^2+(6-6)^2+(5-5)^2+(3-1)^2\right] = 0,8286.$$

Tabuľka 16

Študent	Š1	Š2	Š3	Š4	Š5	Š6
Učiteľ A	12	5	8	47	23	10
p_i	4	1	2	6	5	3
Učiteľ B	15	5	18	39	28	11
q_i	3	1	4	6	5	2

Z hodnoty $r_S = 0,8286$ vyplýva, že medzi premennými X a Y je silná korelačná závislosť.

b) Budeme testovať

$$H_0 : r_S = 0,$$

$$H_1 : r_S \neq 0.$$

Nulová hypotéza hovorí, že Spearmanov koeficient poradovej korelácie r_S nie je štatisticky významný. Proti obojstrannej alternatívnej hypotéze, ktorá hovorí, že Spearmanov koeficient poradovej korelácie r_S je štatisticky významný.
Keďže počet hodnôt je menší ako 30, hodnota testovacieho kritéria je hodnota Spearmanovho koeficientu poradovej korelácie $r_S = 0,8286$. Kritická hodnota testu $r_\alpha(n) = r_{0,05}(6) = 0,8286$ (Tabuľka 52). Kritická oblasť je

$$K_{0,05} = (-\infty; -0,8286) \cup (0,8286; \infty) \Rightarrow r_S \notin K_{0,05}.$$

Hodnota testovacieho kritéria $r_S = 0,8286$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ nie je Spearmanov koeficient poradovej korelácie štatisticky významný.

7 Regresná analýza

Zatiaľ čo korelačná analýza sa zaoberá existenciou závislosti (korelácie) medzi náhodnými premennými, *regresná analýza* zisťuje tvar tejto závislosti. Jej úlohou je teda nájsť matematickú funkciu, tzv. *regresnú funkciu (model)*, ktorá najlepšie popisuje tvar (priebeh) závislosti kvantitatívnych náhodných premenných. Kvalitu regresnej funkcie (modelu) vieme posúdiť pomocou korelačnej analýzy alebo iných štatistických metód.

Náhodná premenná Y, ktorej závislosť od iných premenných zisťujeme, sa nazýva *závislá (vysvetľovaná) premenná*. Náhodnú premennú X, resp. náhodné premenné $X_1, X_2, \ldots, X_n$, ktorej, resp. ktorých zmena pôsobí na zmenu závislej (vysvetľovanej) premennej Y, nazývame *nezávislá (vysvetľujúca) premenná*, resp. *nezávislé (vysvetľujúce) premenné.*

Vo všeobecnosti môžeme regresný model zapísať v tvare

$$Y = f(X_1, X_2, \ldots, X_n, \beta_0, \beta_1, \ldots, \beta_n) + \varepsilon, \tag{135}$$

kde f je regresná funkcia s n neznámymi premennými $X_1, X_2, \ldots, X_n$; $\beta_0, \beta_1, \ldots, \beta_n$ sú parametre modelu a ε je náhodná zložka, ktorá predstavuje pôsobenie náhodných vplyvov a ďalších faktorov, ktoré nie sú zaradené do modelu. Regresná funkcia f môže byť ľubovoľná matematická funkcia, my sa budeme zaoberať predovšetkým lineárnymi alebo linearizovateľnými nelineárnymi regresnými modelmi a obmedzíme sa len na jednoduché regresné modely.

7.1 Lineárny regresný model

O *lineárnom regresnom modele* hovoríme, ak sa vzťah medzi závislou premennou Y a nezávislou premennou X dá vyjadriť pomocou lineárnej funkcie, t. j. regresná funkcia má tvar

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{136}$$

resp.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ pre } i = 1, 2, \ldots, n,$$

kde y_i, resp. x_i predstavujú i–te hodnoty premennej Y, resp. X, ε_i je náhodná chyba i–teho pozorovania a β_0, β_1 sú parametre modelu.

Bodovým odhadom lineárneho regresného modelu je rovnica

$$\widehat{Y} = b_0 + b_1 X, \tag{137}$$

resp.

$$\widehat{y_i} = b_0 + b_1 x_i, \text{ pre } i = 1, 2, \ldots, n,$$

kde $\widehat{y}_i$ je i–ta teoretická hodnota závislej premennej Y a b_0, b_1 sú bodové odhady parametrov β_0, β_1 regresného modelu.

Parametre b_0, b_1 odhadneme metódou najmenších štvorcov, podľa ktorej potrebujeme minimalizovať súčet štvorcov odchýlok empirických hodnôt y_i od teoretických $\widehat{y}_i$, tzv. *reziduálny súčet štvorcov* SSE (ang. Sum of square errors). Platí

$$SSE = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \to \min, \tag{138}$$

kde rozdiel medzi empirickou hodnotou y_i a teoretickou hodnotou $\widehat{y}_i$ nazývame *rezíduum* e_i, t. j. $e_i = y_i - \widehat{y}_i$.

Po dosadení rovnice modelu $\widehat{y}_i = b_0 + b_1 x_i$ do vzťahu (138) dostaneme

$$SSE = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 \to \min. \tag{139}$$

Pre nájdenie minima musíme určiť parciálne derivácie podľa oboch parametrov b_0, b_1 a položiť ich rovné 0. Dostávame

$$\begin{aligned} \frac{\partial SSE}{\partial b_0} &= 2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)(-1) = 0, \\ \frac{\partial SSE}{\partial b_1} &= 2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)(-x_i) = 0. \end{aligned} \tag{140}$$

Úpravou rovníc (140) dostávame sústavu lineárnych rovníc v tvare

$$\begin{aligned} n\, b_0 + b_1 \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i, \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i\, y_i. \end{aligned} \tag{141}$$

Vyriešením sústavy rovníc (141) získame bodové odhady b_0, b_1 parametrov β_0, β_1 lineárneho regresného modelu.

Stanovením bodových odhadov b_0, b_1 parametrov β_0, β_1 lineárneho regresného modelu vieme pomocou reziduálneho súčtu štvorcov SSE vypočítať *výberový reziduálny rozptyl (priemerná kvadratická chyba)* MSE (ang. Mean square of error). Vo všeobecnosti má tvar

$$MSE = \frac{SSE}{n-m} = \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{n-m}, \tag{142}$$

kde m je počet odhadovaných parametrov regresného modelu. V prípade jednoduchého lineárneho modelu je $m = 2$. Výberový reziduálny rozptyl je bodovým odhadom rozptylu náhodných chýb $\widehat{\sigma}^2$.

Bodovým odhadom smerodajnej odchýlky náhodných chýb $\widehat{\sigma}$ je *výberová reziduálna smerodajná odchýlka* (ang. Root mean square of error), pre ktorú platí $RMSE$

$$RMSE = \sqrt{MSE}. \tag{143}$$

Bodovým odhadom rozptylu parametra b_0, resp. b_1 je *výberový rozptyl* $s^2(b_0)$, resp. $s^2(b_1)$, pre ktorý platí

$$s^2(b_0) = \widehat{\sigma}^2 \frac{\sum\limits_{i=1}^{n} x_i^2}{n \sum\limits_{i=1}^{n} (x_i - \overline{x})^2}, \quad \text{resp.} \quad s^2(b_1) = \frac{\widehat{\sigma}^2}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}, \tag{144}$$

kde $\overline{x}$ je výberový aritmetický priemer náhodnej premennej X a $\widehat{\sigma}^2$ je výberový reziduálny rozptyl MSE.

Obojstranný $100\,(1-\alpha)\%$*–ný interval spoľahlivosti pre parameter* β_i*,* $i = 0, 1$ má tvar

$$\left\langle b_i - t_{1-\frac{\alpha}{2}}(n-2)\, s(b_i); b_i + t_{1-\frac{\alpha}{2}}(n-2)\, s(b_i) \right\rangle, \tag{145}$$

kde $s(b_i)$ je výberová smerodajná odchýlka odhadnutého parametra b_i.

7.2 Nelineárne regresné modely

O *nelineárnom regresnom modele* hovoríme, ak sa vzťah medzi závislou premennou Y a nezávislou premennou X nedá vyjadriť pomocou lineárnej funkcie. Takéto regresné modely môžeme rozdeliť na *linearizovateľné* (je možné ich vhodnou transformáciou upraviť na lineárny regresný model) a *nelinearizovateľné* (nie je možné ich upraviť na lineárny regresný model). Keďže linearizovateľné nelineárne regresné modely vieme pomocou vhodnej transformácie upraviť na lineárny regresný model, vieme rovnako ako v predchádzajúcej kapitole, použiť na odhad parametrov regresného modelu metódu najmenších štvorcov. Následne po odhadnutí, sa parametre opäť retransformujú na parametre nelineárneho modelu.

7.2.1 Kvadratický regresný model

O *kvadratickom regresnom modele* hovoríme, ak sa vzťah medzi závislou premennou Y a nezávislou premennou X dá vyjadriť pomocou kvadratickej funkcie, t. j. regresná funkcia má tvar

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon, \tag{146}$$

resp.

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \text{ pre } i = 1, 2, \ldots, n,$$

kde y_i, resp. x_i predstavujú i–te hodnoty premennej Y, resp. X, ε_i je náhodná chyba i–teho pozorovania a β_0, β_1, β_2 sú parametre modelu.

Bodovým odhadom kvadratického regresného modelu je rovnica

$$\widehat{Y} = b_0 + b_1 X + b_2 X^2, \tag{147}$$

resp.

$$\widehat{y_i} = b_0 + b_1 x_i + b_2 x_i^2, \text{ pre } i = 1, 2, \ldots, n,$$

kde $\widehat{y_i}$ je i–ta teoretická hodnota závislej premennej Y a b_0, b_1, b_2 sú bodové odhady parametrov β_0, β_1, β_2 regresného modelu. Parametre b_0, b_1, b_2 odhadneme metódou najmenších štvorcov, podľa ktorej požadujeme, aby

$$SSE = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 \to \min. \tag{148}$$

Po dosadení rovnice modelu $\widehat{y_i} = b_0 + b_1 x_i + b_2 x_i^2$ do vzťahu (148) dostaneme

$$SSE = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i - b_2 x_i^2)^2 \to \min. \tag{149}$$

Pre nájdenie minima musíme určiť parciálne derivácie podľa parametrov b_0, b_1, b_2 a položiť ich rovné 0. Dostávame

$$\begin{aligned}
\frac{\partial SSE}{\partial b_0} &= 2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i - b_2 x_i^2)(-1) = 0, \\
\frac{\partial SSE}{\partial b_1} &= 2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i - b_2 x_i^2)(-x_i) = 0. \\
\frac{\partial SSE}{\partial b_2} &= 2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i - b_2 x_i^2)(-x_i^2) = 0.
\end{aligned} \tag{150}$$

Úpravou rovníc 150 dostávame sústavu rovníc v tvare

$$\begin{aligned}
n\, b_0 + b_1 \sum_{i=1}^{n} x_i + b_2 \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} y_i, \\
b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 + b_2 \sum_{i=1}^{n} x_i^3 &= \sum_{i=1}^{n} x_i\, y_i, \\
b_0 \sum_{i=1}^{n} x_i^2 + b_1 \sum_{i=1}^{n} x_i^3 + b_2 \sum_{i=1}^{n} x_i^4 &= \sum_{i=1}^{n} x_i^2\, y_i.
\end{aligned} \tag{151}$$

Vyriešením sústavy rovníc (151) získame bodové odhady b_0, b_1,b_2 parametrov β_0, β_1, β_2 kvadratického regresného modelu.

7.2.2 Hyperbolický regresný model

O *hyperbolickom regresnom modele* hovoríme, ak sa vzťah medzi závislou premennou Y a nezávislou premennou X dá vyjadriť pomocou hyperbolickej funkcie, t. j. regresná funkcia má tvar

$$Y = \beta_0 + \frac{\beta_1}{X} + \varepsilon, \tag{152}$$

resp.

$$y_i = \beta_0 + \frac{\beta_1}{x_i} + \varepsilon_i, \text{ pre } i = 1, 2, \dots, n,$$

kde y_i, resp. x_i predstavujú i–te hodnoty premennej Y, resp. X, ε_i je náhodná chyba i–teho pozorovania a β_0, β_1 sú parametre modelu.

Bodovým odhadom hyperbolického regresného modelu je rovnica

$$\widehat{Y} = b_0 + \frac{b_1}{X}, \tag{153}$$

resp.

$$\widehat{y_i} = b_0 + \frac{b_1}{x_i}, \text{ pre } i = 1, 2, \dots, n,$$

kde $\widehat{y_i}$ je i–ta teoretická hodnota závislej premennej Y a b_0, b_1 sú bodové odhady parametrov β_0, β_1 regresného modelu. Na odhad parametrov b_0, b_1 môžeme použiť transformáciu

$$T = \frac{1}{X},$$

ktorou prevedieme funkciu $\widehat{Y} = b_0 + \frac{b_1}{X}$ na lineárnu funkciu v tvare $\widehat{Y} = b_0 + b_1 T$. Potom pokračujeme ako v prípade lineárneho regresného modelu, čím dostaneme nasledujúcu sústavu rovníc

$$\begin{aligned} n\, b_0 + b_1 \sum_{i=1}^{n} t_i &= \sum_{i=1}^{n} y_i, \\ b_0 \sum_{i=1}^{n} t_i + b_1 \sum_{i=1}^{n} t_i^2 &= \sum_{i=1}^{n} t_i\, y_i, \end{aligned} \tag{154}$$

kde $t_i = \frac{1}{x_i}$ pre $i = 1, 2, \dots, n$. Vyriešením sústavy rovníc (154) získame bodové odhady b_0, b_1 parametrov β_0, β_1 hyperbolického regresného modelu.

7.2.3 Logaritmický regresný model

O *logaritmickom regresnom modele* hovoríme, ak sa vzťah medzi závislou premennou Y a nezávislou premennou X dá vyjadriť pomocou logaritmickej funkcie, t. j. regresná funkcia má tvar

$$Y = \beta_0 + \beta_1 \ln X + \varepsilon, \tag{155}$$

resp.

$$y_i = \beta_0 + \beta_1 \ln x_i + \varepsilon_i, \text{ pre } i = 1, 2, \ldots, n,$$

kde y_i, resp. x_i predstavujú i–te hodnoty premennej Y, resp. X, ε_i je náhodná chyba i–teho pozorovania a β_0, β_1 sú parametre modelu.

Bodovým odhadom logaritmického regresného modelu je rovnica

$$\widehat{Y} = b_0 + b_1 \ln X, \tag{156}$$

resp.

$$\widehat{y_i} = b_0 + b_1 \ln x_i, \text{ pre } i = 1, 2, \ldots, n,$$

kde $\widehat{y_i}$ je i–ta teoretická hodnota závislej premennej Y a b_0, b_1 sú bodové odhady parametrov β_0, β_1 regresného modelu. Na odhad parametrov b_0, b_1 môžeme použiť transformáciu

$$T = \ln X,$$

ktorou prevedieme funkciu $\widehat{Y} = b_0 + b_1 \ln X$ na lineárnu funkciu v tvare $\widehat{Y} = b_0 + b_1 T$. Potom pokračujeme ako v prípade lineárneho regresného modelu, čím dostaneme nasledujúcu sústavu rovníc

$$\begin{aligned} n\, b_0 + b_1 \sum_{i=1}^{n} t_i &= \sum_{i=1}^{n} y_i, \\ b_0 \sum_{i=1}^{n} t_i + b_1 \sum_{i=1}^{n} t_i^2 &= \sum_{i=1}^{n} t_i\, y_i. \end{aligned} \tag{157}$$

kde $t_i = \ln x_i$ pre $i = 1, 2, \ldots, n$. Vyriešením sústavy rovníc (157) získame bodové odhady b_0, b_1 parametrov β_0, β_1 logaritmického regresného modelu.

7.2.4 Exponenciálny regresný model

O *exponenciálnom regresnom modele* hovoríme, ak sa vzťah medzi závislou premennou Y a nezávislou premennou X dá vyjadriť pomocou exponenciálnej funkcie. V tomto prípade môže mať regresná funkcia dva tvary.

1. prípad: Regresná funkcia má tvar

$$Y = \beta_0\, e^{\beta_1 X} + \varepsilon, \tag{158}$$

resp.

$$y_i = \beta_0\, e^{\beta_1 x_i} + \varepsilon_i, \text{ pre } i = 1, 2, \ldots, n,$$

kde y_i, resp. x_i predstavujú i–te hodnoty premennej Y, resp. X, ε_i je náhodná chyba i–teho pozorovania a β_0, β_1 sú parametre modelu.

Bodovým odhadom exponenciálneho regresného modelu je rovnica

$$\widehat{Y} = b_0\, e^{b_1 X}, \tag{159}$$

resp.

$$\widehat{y_i} = b_0\, e^{b_1 x_i}, \text{ pre } i = 1, 2, \ldots, n,$$

kde $\widehat{y_i}$ je i–ta teoretická hodnota závislej premennej Y a b_0, b_1 sú bodové odhady parametrov β_0, β_1 regresného modelu. Na získanie lineárnej regresnej funkcie môžeme použiť logaritmickú transformáciu

$$\begin{aligned}
\ln \widehat{Y} &= \ln\left(b_0\, e^{b_1 X}\right),\\
\ln \widehat{Y} &= \ln b_0 + \ln e^{b_1 X},\\
\ln \widehat{Y} &= \ln b_0 + b_1 X.
\end{aligned}$$

Nahradením $\ln \widehat{Y} = \widehat{Z}$, $\ln b_0 = a_0$, $b_1 = a_1$ dostávame lineárnu funkciu v tvare $\widehat{Z} = a_0 + a_1 X$. Potom pokračujeme ako v prípade lineárneho regresného modelu. Dostaneme nasledujúcu sústavu rovníc

$$\begin{aligned}
n\, a_0 + a_1 \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} z_i,\\
a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i\, z_i,
\end{aligned} \tag{160}$$

kde $z_i = \ln y_i$ pre $i = 1, 2, \ldots, n$. Vyriešením sústavy rovníc (160) dostaneme a_0, a_1, pomocou ktorých získame bodové odhady b_0, b_1 parametrov β_0, β_1 exponenciálneho regresného modelu nasledovne: $b_0 = e^{a_0}$ a $b_1 = a_1$.

2. prípad: Regresná funkcia má tvar

$$Y = \beta_0\, \beta_1^X + \varepsilon, \tag{161}$$

resp.

$$y_i = \beta_0\, \beta_1^{x_i} + \varepsilon_i, \text{ pre } i = 1, 2, \ldots, n,$$

kde y_i, resp. x_i predstavujú i–te hodnoty premennej Y, resp. X, ε_i je náhodná chyba i–teho pozorovania a β_0, β_1 sú parametre modelu.

Bodovým odhadom exponenciálneho regresného modelu je rovnica

$$\widehat{Y} = b_0\, b_1^X, \tag{162}$$

resp.

$$\widehat{y_i} = b_0\, b_1^{x_i},\ \text{ pre } i = 1, 2, \ldots, n,$$

kde $\widehat{y_i}$ je i–ta teoretická hodnota závislej premennej Y a b_0, b_1 sú bodové odhady parametrov β_0, β_1 regresného modelu. Na získanie lineárnej regresnej funkcie môžeme použiť logaritmickú transformáciu

$$\begin{aligned}
\ln \widehat{Y} &= \ln\left(b_0\, b_1^X\right), \\
\ln \widehat{Y} &= \ln b_0 + \ln b_1^X, \\
\ln \widehat{Y} &= \ln b_0 + X \ln b_1.
\end{aligned}$$

Nahradením $\ln \widehat{Y} = \widehat{Z}$, $\ln b_0 = a_0$, $\ln b_1 = a_1$ dostávame lineárnu funkciu v tvare $\widehat{Z} = a_0 + a_1 X$. Potom pokračujeme ako v prípade lineárneho regresného modelu, čím dostaneme nasledujúcu sústavu rovníc

$$\begin{aligned}
n\, a_0 + a_1 \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} z_i, \\
a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i\, z_i,
\end{aligned} \tag{163}$$

kde $z_i = \ln y_i$ pre $i = 1, 2, \ldots, n$. Vyriešením sústavy rovníc (163) dostaneme a_0, a_1, pomocou ktorých získame bodové odhady b_0, b_1 parametrov β_0, β_1 exponenciálneho regresného modelu nasledovne: $b_0 = e^{a_0}$ a $b_1 = e^{a_1}$.

7.2.5 Mocninový regresný model

O *mocninovom regresnom modele* hovoríme, ak sa vzťah medzi závislou premennou Y a nezávislou premennou X dá vyjadriť rovnicou

$$Y = \beta_0\, X^{\beta_1} + \varepsilon, \tag{164}$$

resp.

$$y_i = \beta_0\, x_i^{\beta_1} + \varepsilon_i,\ \text{ pre } i = 1, 2, \ldots, n,$$

kde y_i, resp. x_i predstavujú i–te hodnoty premennej Y, resp. X, ε_i je náhodná chyba i–teho pozorovania a β_0, β_1 sú parametre modelu.

Bodovým odhadom mocninového regresného modelu je rovnica

$$\widehat{Y} = b_0\, X^{b_1}, \tag{165}$$

resp.

$$\widehat{y_i} = b_0\, x_i^{b_1},\ \text{ pre } i = 1, 2, \ldots, n,$$

kde $\widehat{y_i}$ je i–ta teoretická hodnota závislej premennej Y a b_0, b_1 sú bodové odhady parametrov β_0, β_1 regresného modelu. Na získanie lineárnej regresnej funkcie môžeme použiť logaritmickú transformáciu

$$\ln \widehat{Y} = \ln\left(b_0\, X^{b_1}\right),$$
$$\ln \widehat{Y} = \ln b_0 + \ln X^{b_1},$$
$$\ln \widehat{Y} = \ln b_0 + b_1 \ln X.$$

Nahradením $\ln \widehat{Y} = \widehat{Z}$, $\ln b_0 = a_0$, $b_1 = a_1$ a $\ln X = T$ dostávame lineárnu funkciu v tvare $\widehat{Z} = a_0 + a_1 T$. Potom pokračujeme ako v prípade lineárneho regresného modelu, čím dostaneme nasledujúcu sústavu rovníc

$$\begin{aligned} n\, a_0 + a_1 \sum_{i=1}^{n} t_i &= \sum_{i=1}^{n} z_i, \\ a_0 \sum_{i=1}^{n} t_i + a_1 \sum_{i=1}^{n} t_i^2 &= \sum_{i=1}^{n} t_i\, z_i, \end{aligned} \tag{166}$$

kde $z_i = \ln y_i$, $t_i = \ln x_i$ pre $i = 1, 2, \ldots, n$. Vyriešením sústavy rovníc (166) dostaneme a_0, a_1, pomocou ktorých získame bodové odhady b_0, b_1 parametrov β_0, β_1 mocninového regresného modelu nasledovne: $b_0 = e^{a_0}$ a $b_1 = a_1$.

7.2.6 Niektoré ďalšie nelineárne regresné modely

Prehľad vybraných nelineárnych regresných modelov, ktoré sa dajú podobne ako v predchádzajúcich prípadoch pretransformovať na lineárny regresný model, sú zhrnuté v Tabuľke 17.

Tabuľka 17: Ďalšie nelineárne regresné modely

Bodový odhad regresného modelu	Linearizujúca transformácia	Bodový odhad lineárneho regresného modelu
$\widehat{Y} = b_0 + b_1\, e^X$	$e^X = T$	$\widehat{Y} = b_0 + b_1 T$
$\widehat{Y} = b_0 + b_1\, a^X$	$a^X = T$	$\widehat{Y} = b_0 + b_1 T$
$\widehat{Y} = b_0 + b_1 \sin X$	$\sin X = T$	$\widehat{Y} = b_0 + b_1 T$
$\widehat{Y} = \frac{1}{b_0 + b_1 X}$	$\frac{1}{\widehat{Y}} = b_0 + b_1 X$, $\frac{1}{\widehat{Y}} = \widehat{Z}$	$\widehat{Z} = b_0 + b_1 X$
$\widehat{Y} = \frac{X}{b_0 + b_1 X}$	$\frac{1}{\widehat{Y}} = \frac{b_0}{X} + b_1$, $\frac{1}{\widehat{Y}} = \widehat{Z}$, $\frac{1}{X} = t$	$\widehat{Z} = b_0 T + b_1$

7.3 Posudzovanie vhodnosti a významnosti regresného modelu

V ďalšej časti nás zaujíma, či regresný model odhadnutý pomocou metódy najmenších štvorcov je vhodný na analýzu skúmaného náhodného výberu a či je štatisticky významný.

7.3.1 Výberový koeficient determinácie

Na posudzovanie vhodnosti zvoleného regresného modelu odhadnutého pomocou metódy najmenších štvorcov slúži *výberový koeficient (index) determinácie* R^2. Tento koeficient je popisnou mierou vhodnosti použitia regresnej funkcie na predikciu a je definovaný nasledovne

$$R^2 = 1 - \frac{SSE}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \tag{167}$$

kde $\overline{y}$ je výberový aritmetický priemer závislej premennej Y a n je rozsah náhodného výberu.

Výberový koeficient determinácie nadobúda len hodnoty z intervalu $\langle 0; 1\rangle$ a určuje tú časť celkovej variability pozorovaných hodnôt, ktorú je možné vysvetliť daným regresným modelom. Vyjadrený v percentách ($100\,R^2\%$) udáva, koľko percent celkovej variability je vysvetliteľných regresným modelom. Z toho vyplýva, že hodnoty blízke nule naznačujú, že zvolená regresná funkcia nie je vhodná. Naopak, hodnoty blízke 1 naznačujú, že regresná funkcia veľmi dobre vystihuje závislosť medzi jednotlivými náhodnými premennými.

Pre malé rozsahy výberu n je tento odhad vychýlený a nadhodnocuje priliehavosť k regresnému modelu, preto sa používa nevychýlený odhad *modifikovaný (korigovaný) výberový koeficient determinácie* definovaný vzťahom

$$R^{*2} = R^2 - (1 - R^2)\,\frac{m-1}{n-m}, \tag{168}$$

kde m je počet odhadovaných parametrov regresnej funkcie.

7.3.2 Test významnosti regresného modelu

Test významnosti regresného modelu súčasne testuje významnosť výberového koeficienta determinácie a všetkých parametrov modelu.

1. Hypotézy:

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \ldots = \beta_{m-1} = 0,$$
$$H_1 : \text{aspoň jeden z parametrov } \beta_0, \beta_1, \beta_2, \ldots, \beta_{m-1} \neq 0,$$

resp.

$$H_0 : \text{regresný model nie je štatisticky významný},$$
$$H_1 : \text{regresný model je štatisticky významný}.$$

2. Testovacie kritérium je

$$F = \frac{(n-m)\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{(m-1)\ SSE} = \frac{(n-m)\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{(m-1)\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \sim F(m-1, n-m), \tag{169}$$

   kde $\overline{y}$ je výberový aritmetický priemer náhodnej premennej Y, m je počet odhadovaných parametrov regresnej funkcie a n je rozsah náhodného výberu.

3. Kritická oblasť K_α je

$$K_\alpha = (F_{1-\alpha}(m-1, n-m); \infty)\,.$$

Niekedy je užitočné otestovať, či niektorý z parametrov regresnej funkcie sa rovná nejakej konkrétnej konštante b.

Test zhody parametrov regresnej funkcie so známou konštantou

1. Hypotézy:

$$H_0 : \beta_i = b,$$
$$H_1 : \beta_i \neq b.$$

2. Testovacie kritérium je

$$t = \frac{b_i - b}{s(b_i)} \sim t(n-2), \tag{170}$$

   kde $s(b_i)$ je výberová smerodajná odchýlka odhadnutého parametra b_i a n je rozsah náhodného výberu.

3. Kritická oblasť K_α pre rôzne alternatívy je uvedená v nasledujúcej tabuľke:

H_0	H_1	K_α
$\beta_i = b$	$\beta_i \neq b$	$\left(-\infty; -t_{1-\frac{\alpha}{2}}(n-2)\right) \cup \left(t_{1-\frac{\alpha}{2}}(n-2); \infty\right)$
$\beta_i = b$	$\beta_i > b$	$(t_{1-\alpha}(n-2); \infty)$
$\beta_i = b$	$\beta_i < b$	$(-\infty; -t_{1-\alpha}(n-2))$

Poznámka: V prípade, že v predchádzajúcom teste budeme testovať nulovú hypotézu $\beta_i = 0$, ide o *test významnosti parametra* β_i, $i = 0, 1, \ldots, m$.

7.3.3 Štatistická analýza rezíduí

Rezíduá e_i sú dôležité pri hodnotení kvality regresnej funkcie a údajov, taktiež pri posudzovaní oprávnenosti predpokladov u zvoleného lineárneho regresného modelu. Predstavujú bodové odhady náhodných chýb ε_i modelu. Metóda najmenších štvorcov predstavuje najlepšie neskreslené odhady lineárneho modelu, ak budeme predpokladať, že náhodné chyby ε_i modelu pre $i = 1, 2, \ldots, n$ majú nasledujúce vlastnosti:

1. Stredná hodnota náhodnej chyby ε_i je nula, t. j. $E(\varepsilon_i) = 0$.
2. Rozptyl náhodnej chyby ε_i je konštantný a nadobúda hodnoty σ_ε^2, $\sigma_\varepsilon > 0$.
3. Rozdelenie pravdepodobnosti náhodnej chyby ε_i je normálne, t. j. $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$.
4. Náhodné chyby sú medzi sebou vzájomne nezávislé, t. j. $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$, pre všetky $i \neq j$.

Splnenie týchto predpokladov sa dá overiť až po zvolení regresného modelu, pretože až vtedy vieme určiť rezíduá. Pokiaľ však model nespĺňa niektorú z uvedených podmienok, nie je možné daný model použiť pre skúmané údaje.

Grafickou analýzou, či sú dané predpoklady splnené, môže byť bodový graf závislosti rezíduí e_i od teoretických hodnôt, t. j. bodový graf $(\widehat{y}_i, e_i)$, $i = 0, 1, \ldots, n$. Ak sú splnené predpoklady modelu, rezíduá musia byť náhodne rozptýlené okolo nuly. Rezíduá je možné ohraničiť dvomi priamkami rovnobežnými s osou x, pričom v grafe by nemal byť žiaden náznak potenciálneho trendu.

Po grafickej analýze si jednotlivé vlastnosti náhodných chýb overíme aj pomocou štatistických testov. Prvá vlastnosť náhodných chýb ε_i, ktorú potrebujeme overiť je, či rozdelenie pravdepodobnosti náhodných chýb je normálne rozdelenie. K tomu pomôžu testy dobrej zhody (kapitola 5.5). Nulovú strednú hodnotu náhodných chýb môžeme overiť pomocou jednovýberového t–testu strednej hodnoty so známou konštantou (kapitola 5.1.2). Ďalej potrebujeme overiť je, či náhodné chyby majú konštantný rozptyl. K tomu môžeme použiť Goldfeldov-Quandtov test, ktorý je obdobou testu zhody rozptylov dvoch nezávislých normálnych súborov.

Goldfeldov-Quandtov test

Predpokladáme, že náhodné chyby ε_i majú normálne rozdelenie. Údaje usporiadame do neklesajúcej postupnosti podľa hodnôt premennej X a rozdelíme ich do dvoch skupín. Horná časť obsahuje menšie hodnoty premennej X (ich počet označíme n_h) a dolná časť obsahuje väčšie hodnoty premennej X (ich počet označíme n_d). Následne odhadneme nové parametre regresného modelu zvlášť pre hornú a dolnú časť pozorovaní pomocou metódy najmenších štvorcov. Súčty štvorcov rezíduí pre obidve časti označíme SSE_h a SSE_d.

1. Hypotézy:

$$H_0 : \sigma_d^2 = \sigma_h^2,$$
$$H_1 : \sigma_d^2 \neq \sigma_h^2.$$

2. Testovacie kritérium je

$$F = \frac{SSE_d}{SSE_h}\,\frac{n_h - m}{n_d - m} = \frac{MSE_d}{MSE_h} \sim F(n_d - m, n_h - m), \tag{171}$$

kde m je počet odhadovaných parametrov regresnej funkcie.

3. Kritická oblasť K_α je

$$K_\alpha = (F_{1-\alpha}(n_d - m, n_h - m); \infty)\,.$$

Poznámka: V špeciálnom prípade, ak rozdelíme dáta do dvoch rovnako početných častí (t. j. $n_d = n_h = \frac{n}{2}$), testovacie kritérium F má tvar

$$F = \frac{SSE_d}{SSE_h} \sim F\left(\frac{n}{2} - m, \frac{n}{2} - m\right), \tag{172}$$

kde m je počet odhadovaných parametrov regresnej funkcie a n je rozsah náhodného výberu.

Posledná vlastnosť náhodných chýb, ktorú potrebujeme overiť, je vzájomná nezávislosť náhodných chýb. Mierou závislosti (korelovanosti) rezíduí je *Durbin–Watsonova štatistika DW*, ktorá má tvar

$$DW = \frac{\sum\limits_{i=2}^{n} (e_i - e_{i-1})^2}{\sum\limits_{i=1}^{n} e_i^2}, \tag{173}$$

kde e_i sú rezíduá. Durbin–Watsonova štatistika nadobúda hodnoty z intervalu $\langle 0; 4\rangle$. Na jej interpretáciu sa v praxi používa nasledujúca stupnica:

- ak $DW \in \langle 1,4; 2,6\rangle$, tak rezíduá e_i nevykazujú autokoreláciu (t. j. majú náhodný charakter, sú nezávislé) a model je dobrý,
- ak $DW < 1,4$, tak rezíduá e_i sú kladne korelované a model je nevyhovujúci,
- ak $DW > 2,6$, tak rezíduá e_i sú záporne korelované a model je nevyhovujúci.

Analýzu rezíduí, resp. náhodných chýb, uzavrieme overením náhodnosti chýb. K tomu použijeme test náhodnosti chýb založený na tzv. bodoch zvratu. Číslo e_i, $2 \leqq i \leqq n-1$ budeme považovať za *bod zvratu* v postupnosti rôznych čísel $e_1, e_2, \ldots, e_n$, ak platí buď podmienka $e_{i-1} < e_i > e_{i+1}$ alebo podmienka $e_{i-1} > e_i < e_{i+1}$. Pokiaľ sú niektoré susedné hodnoty rovnaké, jednu z nich škrtneme. Označme Z celkový počet bodov zvratu v postupnosti $e_1, e_2, \ldots, e_n$.

Test náhodnosti chýb (test bodov zvratu)

1. Hypotézy:

$$H_0 : \text{chyby sú náhodné,}$$
$$H_1 : \text{chyby nie sú náhodné.}$$

2. Testovacie kritérium je

$$U = \frac{Z - \frac{2n-4}{3}}{\sqrt{\frac{16n-29}{90}}} \sim N(0,1), \tag{174}$$

kde Z je celkový počet bodov zvratu v postupnosti $e_1, e_2, \ldots, e_n$.

3. Kritická oblasť K_α je

$$K_\alpha = \left(-\infty; -u_{1-\frac{\alpha}{2}}\right) \cup \left(u_{1-\frac{\alpha}{2}}; \infty\right).$$

Príklad 7.1 *V nasledujúcej tabuľke sú údaje o množstve úrody jabĺk (Y) [v t] a počte stromov v sade (X) [v ks].*

x_i	*160*	*250*	*320*	*500*	*750*	*1000*	*1500*	*2000*
y_i	*789*	*800*	*851*	*874*	*1193*	*1335*	*1704*	*2073*

Predpokladáme, že medzi sledovanými premennými je lineárny vzťah a že ide o náhodný výber z normálneho rozdelenia.

a) Odhadnime závislosť množstva úrody od počtu stromov regresnou priamkou.

b) Zostrojme korelačný graf náhodných premenných X a Y.

c) Posúďme vhodnosť regresného modelu.

d) Otestujme na hladine významnosti α=0,05 významnosť regresného modelu.

e) Otestujme na hladine významnosti α=0,05 významnosť regresných parametrov.

f) Určme 95%–né intervaly spoľahlivosti pre regresné parametre.

Riešenie: a) Predpokladáme, že medzi sledovanými premennými je lineárny vzťah. Lineárny regresný model je v tvare (136)

$$Y = \beta_0 + \beta_1 X + \varepsilon.$$

Bodovým odhadom lineárneho regresného modelu je rovnica (137)

$$\widehat{Y} = b_0 + b_1 X,$$

kde parametre b_0, b_1 odhadneme metódou najmenších štvorcov. Potrebujeme vyriešiť sústavu (141). Pri jej riešení je vhodné vytvoriť Tabuľku 18.

Tabuľka 18

i	x_i	y_i	x_i^2	$x_i\, y_i$
1	160	789	25600	126240
2	250	800	62500	200000
3	320	851	102400	272320
4	500	874	250000	437000
5	750	1193	562500	894750
6	1000	1335	1000000	1335000
7	1500	1704	2250000	2556000
8	2000	2073	4000000	4146000
$\sum$	6480	9619	8253000	9967310

Po dosadení hodnôt z Tabuľky 18 dostávame sústavu lineárnych rovníc v tvare

$$8b_0 + 6480b_1 = 9619,$$
$$6480b_0 + 8253000b_1 = 9967310.$$

Vyriešením sústavy lineárnych rovníc získame bodové odhady parametrov $b_0 = 615,6979$, $b_1 = 0,7243$. Regresná priamka má tvar

$$\widehat{Y} = 615,6979 + 0,7243X.$$

Teoretické hodnoty $\widehat{y}_i$ sú uvedené v Tabuľke 19.

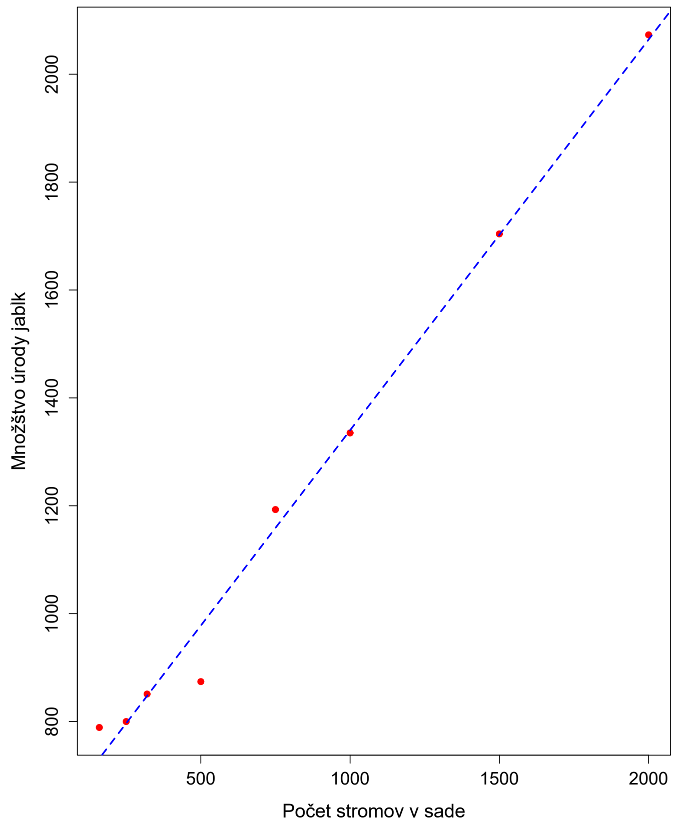
b) Korelačný graf je znázornený na Obr. 25.

c) Vhodnosť regresného modelu posúdime pomocou výberového koeficientu determinácie R^2, ktorý vypočítame pomocou vzťahu (167). Platí

$$\overline{y} = \frac{1}{8}\,(789 + 800 + 851 + 874 + 1193 + 1335 + 1704 + 2073) = 1202,375;$$

Tabuľka 19

x_i	160	250	320	500	750	1000	1500	2000
y_i	789	800	851	874	1193	1335	1704	2073
$\widehat{y}_i$	$731,59$	$796,77$	$847,47$	$977,84$	$1158,92$	$1339,99$	$1702,14$	$2064,28$



Obr. 25: Korelačný graf z Príkladu 7.1

$$SSE=\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2=(789-731,5848)^2+(800-796,7711)^2+(851-847,4716)^2+$$
$$+(874-977,8443)^2+(1193-1158,9174)^2+(1335-1339,9906)^2+$$
$$+(1704-1702,1369)^2+(2073-2064,2833)^2=15368,9969;$$
$$\sum_{i=1}^{n}(y_i-\overline{y})^2=(789-1202,375)^2+(800-1202,375)^2+(851-1202,375)^2+$$
$$+(874-1202,375)^2+(1193-1202,375)^2+(1335-1202,375)^2+$$
$$+(1704-1202,375)^2+(2073-1202,375)^2=1591371,875.$$

Potom výberový koeficient determinácie R^2 je

$$R^2 = 1 - \frac{15368,9969}{1591371,875} = 0,9903.$$

Čo znamená, že až 99,03% hodnôt je možné vysvetliť týmto regresným modelom.

d) Pri teste významnosti regresného modelu budeme testovať

$$H_0 : \text{regresný model nie je štatisticky významný,}$$

$$H_1 : \text{regresný model je štatisticky významný.}$$

Hodnotu testovacieho kritéria F určíme podľa vzťahu (169)

$$F = \frac{(8-2)\sum_{i=1}^{n}(\widehat{y_i} - \overline{y})^2}{(2-1)\cdot 15368,9969},$$

kde

$$\sum_{i=1}^{n}(\widehat{y_i} - \overline{y})^2 = (731,5848 - 1202,375)^2 + (796,7711 - 1202,375)^2 + (847,4716 - 1202,375)^2 +$$

$$+(977,8443 - 1202,375)^2 + (1158,9174 - 1202,375)^2 + (1339,9906 - 1202,375)^2 +$$

$$+(1702,1369 - 1202,375)^2 + (2064,2833 - 1202,375)^2 = 1576002,8781.$$

Potom hodnota testovacieho kritéria je $F = 615,3$.
Kvantil Fisherovho F–rozdelenia je $F_{1-\alpha}(m-1, n-m) = F_{0,95}(1,6) = 5,9874$ (Tabuľka 33) a kritická oblasť je

$$K_{0,05} = (5,9874; \infty) \Rightarrow F \in K_{0,05}.$$

Hodnota testovacieho kritéria $F = 615,3$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ je regresný model štatisticky významný.

e) Pri teste významnosti parametra β_0 budeme testovať

$$H_0 : \beta_0 = 0,$$

$$H_1 : \beta_0 \neq 0.$$

Hodnotu testovacieho kritéria t určíme podľa vzťahu (170)

$$t = \frac{615,6979 - 0}{s(b_0)},$$

kde $s(b_0)$ je výberová smerodajná odchýlka odhadnutého parametra b_0, ktorú vypočítame pomocou vzťahu (144). Platí

$$\sum_{i=1}^{n} x_i^2 = 160^2 + 250^2 + 320^2 + 500^2 + 750^2 + 1000^2 + 1500^2 + 2000^2 = 8253000,$$

$$\overline{x} = \frac{1}{8}(160 + 250 + 320 + 500 + 750 + 1000 + 1500 + 2000) = 810,$$

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = (160 - 810)^2 + (250 - 810)^2 + (320 - 810)^2 + (500 - 810)^2 + (750 - 810)^2 +$$

$$+(1000 - 810)^2 + (1500 - 810)^2 + (2000 - 810)^2 = 3004200.$$

Potom výberová smerodajná odchýlka $s(b_0)$ odhadnutého parametra b_0 je

$$s(b_0) = \sqrt{s^2(b_0)} = \sqrt{\frac{15368,9969}{8-2} \cdot \frac{8253000}{8 \cdot 3004200}} = \sqrt{879,6042} = 29,6581.$$

Hodnota testovacieho kritéria je $t = 20,7598$ a kvantil Studentovho t–rozdelenia je $t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(6) = 2,4469$ (Tabuľka 29). Kritická oblasť je

$$K_{0,05} = (-\infty; -2,4469) \cup (2,4469; \infty) \Rightarrow t \in K_{0,05}.$$

Hodnota testovacieho kritéria $t = 20,7598$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ je parameter β_0 štatisticky významný.

Pri teste významnosti parametra β_1, budeme testovať

$$H_0 : \beta_1 = 0,$$

$$H_1 : \beta_1 \neq 0.$$

Hodnotu testovacieho kritéria t určíme podľa vzťahu (170)

$$t = \frac{b_1 - b}{s(b_1)} = \frac{0,7243 - 0}{s(b_1)},$$

kde $s(b_1)$ je výberová smerodajná odchýlka odhadnutého parametra b_1, ktorú vypočítame pomocou vzťahu (144). Platí

$$s(b_1) = \sqrt{s^2(b_1)} = \sqrt{\frac{15368,9969}{8-2} \cdot \frac{1}{3004200}} = \sqrt{0,00085} = 0,02915.$$

Hodnota testovacieho kritéria je $t = 24,80$ a kvantil Studentovho t–rozdelenia je $t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(6) = 2,4469$ (Tabuľka 29). Kritická oblasť je

$$K_{0,05} = (-\infty; -2,4469) \cup (2,4469; \infty) \Rightarrow t \in K_{0,05}.$$

Hodnota testovacieho kritéria $t = 24,80$ patrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 zamietame a prijímame alternatívnu hypotézu H_1. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ je parameter β_1 štatisticky významný.

f) Chceme určiť intervaly spoľahlivosti pre parametre β_0 a β_1. Kvantil Studentovho t–rozdelenia je $t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(6) = 2,4469$ (Tabuľka 29) a výberové smerodajné odchýlky pre oba parametre sme už vypočítali v časti e). Potom 95%–ný interval spoľahlivosti pre parameter β_0 je podľa vzťahu (145)

$$\langle 615,6979 - 2,4469 \cdot 29,6581; 615,6979 + 2,4469 \cdot 29,6581 \rangle = \langle 543,1271; 688,2688 \rangle$$

a 95%–ný interval spoľahlivosti pre parameter β_1 je

$$\langle 0,7243 - 2,4469 \cdot 0,02915; 0,7243 + 2,4469 \cdot 0,02915 \rangle = \langle 0,6528; 0,7957 \rangle .$$

Príklad 7.2 *V nasledujúcej tabuľke sú údaje o množstve úrody jabĺk (Y) [v t] a počte stromov v sade (X) [v ks].*

x_i	160	250	320	500	750	1000	1500	2000
y_i	789	800	851	874	1193	1335	1704	2073

a) Odhadnime závislosť množstva úrody od počtu stromov exponenciálnym regresným modelom v tvare $Y = \beta_0\, e^{\beta_1 X} + \varepsilon$.
b) Zostrojme korelačný graf náhodných premenných X a Y.
c) Posúďme vhodnosť regresného modelu.

Riešenie: Exponenciálny regresný model je v tvare 158

$$Y = \beta_0\, e^{\beta_1 X} + \varepsilon.$$

Bodovým odhadom exponenciálneho regresného modelu je rovnica 159

$$\widehat{Y} = b_0\, e^{b_1 X},$$

kde parametre b_0, b_1 sú bodové odhady parametrov β_0, β_1. Na získanie parametrov modelu použijeme logaritmickú transformáciu uvedenú v kapitole 7.2.4. Označíme $\ln \widehat{Y} = \widehat{Z}$, $\ln b_0 = a_0$, $b_1 = a_1$, čím dostávame lineárnu funkciu v tvare $\widehat{Z} = a_0 + a_1 X$.

Potom pokračujeme ako v prípade lineárneho regresného modelu. Potrebujeme vyriešiť sústavu (160). Pri jej riešení je vhodné vytvoriť Tabuľku 20. Po dosadení hodnôt z Tabuľky 20 dostávame sústavu lineárnych rovníc v tvare

$$8a_0 + 6480a_1 = 56,2333;$$
$$6480a_0 + 8253000a_1 = 47228,3388.$$

Tabuľka 20

i	x_i	$z_i = \ln y_i$	x_i^2	$x_i\, z_i$
1	160	6,6708	25600	1067,3226
2	250	6,6846	62500	1671,1529
3	320	6,7464	102400	2158,8519
4	500	6,7731	250000	3386,5402
5	750	7,0842	562500	5313,1698
6	1000	7,1967	1000000	7196,6866
7	1500	7,4407	2250000	11161,1006
8	2000	7,6368	4000000	15273,5042
$\sum$	6480	56,2333	8253000	47228,3288

Vyriešením sústavy rovníc získame parametre $a_0 = 6,5764$, $a_1 = 0,0006$. Potom bodové odhady parametrov β_0, β_1 exponenciálneho regresného modelu sú $b_0 = e^{a_0} = e^{6,5764} = 717,9213$ a $b_1 = a_1 = 0,0006$. Hľadaná regresná funkcia má tvar

$$\widehat{Y} = 717,9213\, e^{0,0006X}.$$

Teoretické hodnoty $\widehat{y}_i = e^{z_i}$ sú uvedené v Tabuľke 21.

Tabuľka 21

x_i	160	250	320	500	750	1000	1500	2000
y_i	789	800	851	874	1193	1335	1704	2073
$\widehat{y}_i$	785, 09	825, 60	858, 55	949, 43	1091, 84	1255, 60	1660, 50	2195, 98

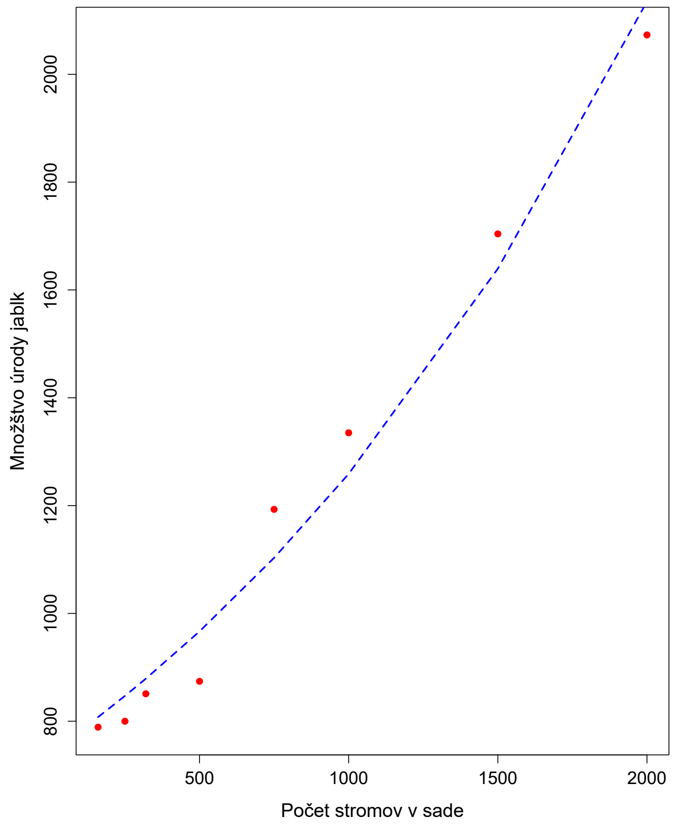
b) Korelačný graf je znázornený na Obr. 26.

c) Vhodnosť regresného modelu posúdime pomocou výberového koeficientu determinácie R^2, ktorý vypočítame pomocou vzťahu (167). Platí

$$\overline{y} = \frac{1}{8}(789 + 800 + 851 + 874 + 1193 + 1335 + 1704 + 2073) = 1202,375;$$

$$SSE = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = (789 - 785,0925)^2 + (800 - 825,6018)^2 + (851 - 858,548)^2 +$$
$$+(874 - 949,4333)^2 + (1193 - 1091,8382)^2 + (1335 - 1255,6024)^2 +$$
$$+(1704 - 1660,5036)^2 + (2073 - 2195,9755)^2 = 39970,4838;$$

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = (789 - 1202,375)^2 + (800 - 1202,375)^2 + (851 - 1202,375)^2 + (874 - 1202,375)^2 +$$



Obr. 26: Korelačný graf z Príkladu 7.2

$+(1193-1202,375)^2+(1335-1202,375)^2+(1704-1202,375)^2+(2073-1202,375)^2 = 1591371,875.$

Hodnota výberového koeficienta determinácie R^2 je

$$R^2 = 1 - \frac{39970,4838}{1591371,875} = 0,9745.$$

Čo znamená, že až 97,45% hodnôt je možné vysvetliť týmto regresným modelom.

Príklad 7.3 *V nasledujúcej tabuľke sú údaje o množstve nakúpeného dreva (Y) [v kg] a počte vyrobených stoličiek (X) [v ks].*

x_i	160	250	320	500	750	1000	1500	2000
y_i	789	800	851	990	1193	1335	1704	2073

Vytvorme lineárny regresný model a posúďme mieru kvality získaného modelu.

Riešenie: Lineárny regresný model je v tvare (136)

$$Y = \beta_0 + \beta_1 X + \varepsilon.$$

Bodovým odhadom lineárneho regresného modelu je rovnica (137)

$$\widehat{Y} = b_0 + b_1 X,$$

kde parametre b_0, b_1 odhadneme metódou najmenších štvorcov. Potrebujeme vyriešiť sústavu (141). Pri jej riešení je vhodné vytvoriť Tabuľku 22.

Tabuľka 22

i	x_i	y_i	x_i^2	$x_i\,y_i$
1	160	789	25600	126240
2	250	800	62500	200000
3	320	851	102400	272320
4	500	990	250000	495000
5	750	1193	562500	894750
6	1000	1335	1000000	1335000
7	1500	1704	2250000	2556000
8	2000	2073	4000000	4146000
$\sum$	6480	9619	8253000	10025310

Po dosadení hodnôt z Tabuľky 22 dostávame sústavu lineárnych rovníc v tvare

$$\begin{aligned} 8b_0 + 6480b_1 &= 9735, \\ 6480b_0 + 8253000b_1 &= 10025310. \end{aligned}$$

Vyriešením danej sústavy rovníc získame bodové odhady parametrov $b_0 = 639,8936$, $b_1 = 0,7123$. Regresná priamka má tvar

$$\widehat{Y} = 639,8936 + 0,7123X.$$

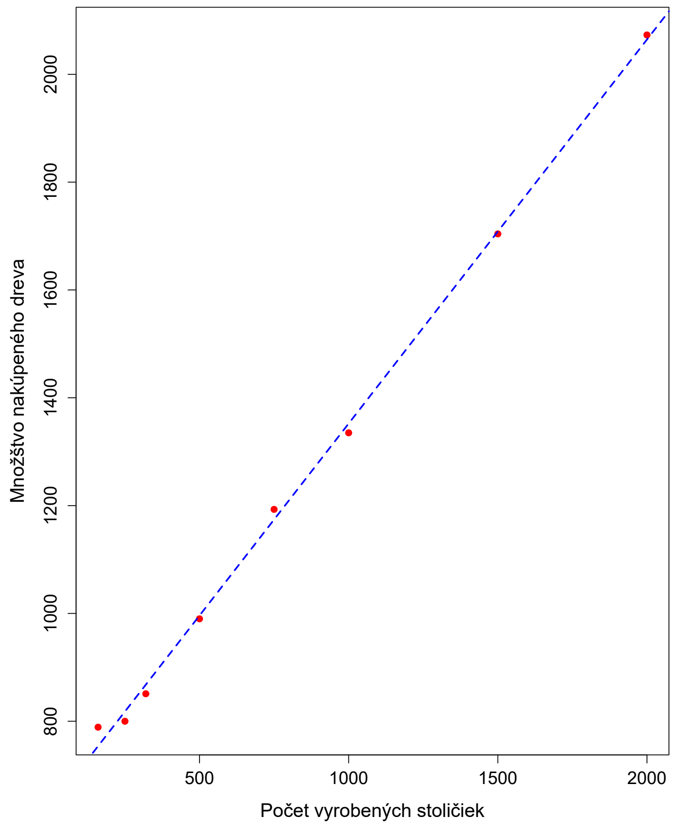
Teoretické hodnoty $\widehat{y}_i$ sú uvedené v Tabuľke 23.

Tabuľka 23

x_i	160	250	320	500	750	1000	1500	2000
y_i	789	800	851	990	1193	1335	1704	2073
$\widehat{y}_i$	753,87	817,97	867,84	996,06	1174,14	1352,22	1708,38	2064,54

Kvôli názornosti je korelačný graf znázornený na Obr. 27.

Aby zvolená metóda najmenších štvorcov odhadla najlepšie a neskreslene parametre regresného modelu, náhodné chyby ε_i musia spĺňať určité vlastnosti. Keďže rezíduá $e_i = y_i - \widehat{y}_i$



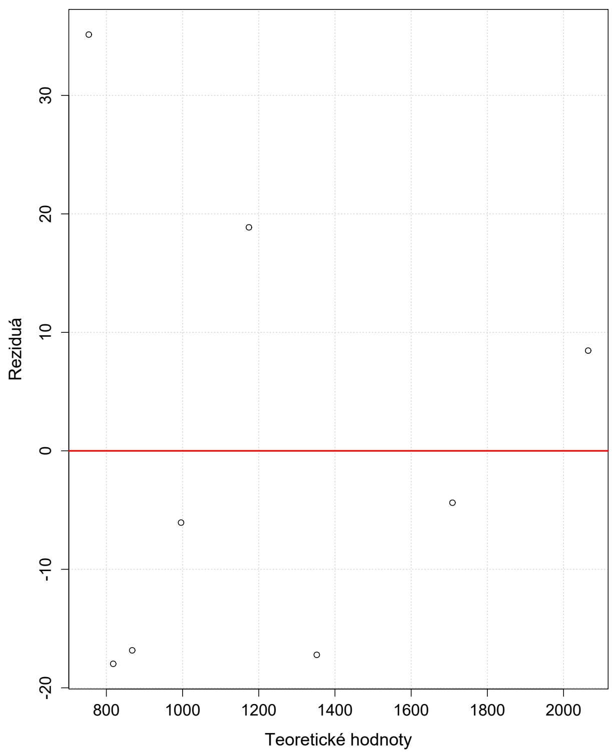
Obr. 27: Korelačný graf z Príkladu 7.3

predstavujú bodové odhady náhodných chýb ε_i, budeme ich potrebovať k posúdeniu miery kvality modelu (Tabuľka 24). Použijeme štatistickú analýzu rezíduí, kde za hladinu významnosti zvolíme $\alpha = 0,05$.

Tabuľka 24

x_i	y_i	$\widehat{y}_i$	e_i
160	789	$753,8652$	$35,1348$
250	800	$817,9743$	$-17,9743$
320	851	$867,8369$	$-16,8369$
500	990	$996,0549$	$-6,0549$
750	1193	$1174,1356$	$18,8644$
1000	1335	$1352,2163$	$-17,2163$
1500	1704	$1708,3777$	$-4,3777$
2000	2073	$2064,5391$	$8,4609$

Najprv urobíme grafickú analýzu, teda vytvoríme bodový graf závislosti rezíduí od teoretických hodnôt (Obr. 28). Pretože rezíduá sú náhodne rozptýlené okolo nuly, predpoklady modelu sú splnené.



Obr. 28: Závislosť rezíduí od teoretických hodnôt z Príkladu 7.3

Prvá vlastnosť náhodných chýb, ktorú potrebujeme overiť, je, či rozdelenie pravdepodobnosti náhodných chýb je normálne rozdelenie. K tomu použijeme napríklad Shapiro-Wilkov test normality. Budeme testovať

$$H_0 : F(x) = G(x), \text{ kde } G(x) \sim N(\mu, \sigma^2),$$
$$H_1 : F(x) \neq G(x).$$

Rozsah súboru je $n = 8$. Usporiadame hodnoty rezíduí e_i do neklesajúcej postupnosti (Tabuľka 25).

Tabuľka 25: Hodnoty rezíduí e_i usporiadané do neklesajúcej postupnosti

i	1	2	3	4	5	6	7	8
$e_{(i)}$	$-17,9743$	$-17,2163$	$-16,8369$	$-6,0549$	$-4,3777$	$8,4609$	$18,8644$	$35,1348$

Hodnotu testovacieho kritéria W určíme podľa vzťahu (122), kde $m = \frac{n}{2} = 4$. Výberový aritmetický priemer vypočítaný podľa vzťahu (71) je rovný nule (t. j. $\overline{e} = 0$). Tabuľkové hodnoty $a_{i,n}$ sú $a_{1,8} = 0,6052$, $a_{2,8} = 0,3164$, $a_{3,8} = 0,1743$, $a_{4,8} = 0,0561$ (Tabuľka 48). Potom

$$\sum_{i=1}^{4} a_{i,n}\left(e_{(n-i+1)} - e_{(i)}\right) = 0,6052 \cdot (35,1348 + 17,9743) + 0,3164 \cdot (18,8644 + 17,2163) +$$

$$+0,1743 \cdot (8,4609 + 16,8369) + 0,0561 \cdot (-4,3777 + 6,0549) = 48,0611,$$

$$\sum_{i=1}^{8} \left(e_{(i)} - \overline{e}\right)^2 = (-17,9743)^2 + (-17,2163)^2 + (-16,8369)^2 + (-6,0549)^2 +$$

$$+(-4,3777)^2 + (8,4609)^2 + (18,8644)^2 + (35,1348)^2 = 2620,687.$$

Hodnota testovacieho kritéria je $W = 0,8814$. Kritická hodnota pre Shapiro-Wilkov test normality je $W_{0,05}(8) = 0,818$ (Tabuľka 49). Kritická oblasť je

$$K_{0,05} = (-\infty; 0,818) \Rightarrow W \notin K_{0,05}.$$

Hodnota testovacieho kritéria $W = 0,8814$ nepatrí do kritickej oblasti $K_{0,05}$, preto na hladine významnosti $\alpha = 0,05$ nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že rozdelenie pravdepodobnosti náhodných chýb je normálne rozdelenie.

Na overenie nulovej strednej hodnoty náhodných chýb použijeme jednovýberový test strednej hodnoty, pričom nepoznáme rozptyl (jednovýberový t–test). Budeme testovať

$$H_0 : \overline{e} = 0,$$

$$H_1 : \overline{e} \neq 0.$$

Hodnotu testovacieho kritéria t určíme podľa vzťahu (97)

$$t = \frac{\overline{e} - \mu_0}{s}\sqrt{n} = \frac{0-0}{s} \cdot \sqrt{8} = 0.$$

Kvantil Studentovho t–rozdelenia je $t_{1-\frac{\alpha}{2}}(n-1) = t_{0,975}(7) = 2,365$ (Tabuľka 29). Kritická oblasť je

$$K_{0,05} = (-\infty; -2,365) \cup (2,365; \infty) \Rightarrow t \notin K_{0,05}.$$

Hodnota testovacieho kritéria $t = 0$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ sa stredná hodnota náhodných chýb významne nelíši od nuly.

Na overenie konštantného rozptylu náhodných chýb použijeme Goldfeldov-Quandtov test. Hodnoty náhodnej premennej X, ktoré tvoria neklesajúcu postupnosť, rozdelíme do dvoch skupín (Tabuľka 26).

Tabuľka 26

x_i	y_i	$\widehat{y_i}$	e_i^2
160	789	765,5183	551,3918
250	800	821,6427	468,4071
320	851	865,2951	204,3488
500	990	977,5440	155,1530
750	1193	1175,7120	298,8796
1000	1335	1353,7290	350,7685
1500	1704	1709,7630	33,2089
2000	2073	2065,7970	51,8888

Pre obe skupiny vytvoríme nové regresné modely, ktoré majú nasledujúce tvary

$$\widehat{Y_h} = 665,7415 + 0,6236X_h; \qquad \widehat{Y_d} = 641,6610 + 0,7121X_d.$$

Budeme testovať

$$H_0 : \sigma_d^2 = \sigma_h^2,$$
$$H_1 : \sigma_d^2 \neq \sigma_h^2.$$

Nulová hypotéza hovorí, že je splnený predpoklad rovnosti rozptylov oboch skupín. Proti alternatívnej hypotéze, ktorá hovorí, že nie je splnený predpoklad rovnosti rozptylov oboch skupín. Hodnotu testovacieho kritéria F určíme podľa vzťahu (171), kde $m = 2$ je počet odhadovaných parametrov lineárnej regresnej funkcie a $n_h = n_d = 4$. Hodnoty súčtov štvorcov rezíduí pre dolnú a hornú časť vypočítame nasledovne

$$SSE_h = \sum_{i=1}^{m} e_i^2 = 1379,3007; \qquad SSE_d = \sum_{i=1}^{m} e_i^2 = 734,7458;$$

kde za hodnoty rezíduí e_i dosadíme len tie, ktoré patria do jednotlivých častí (Tabuľka 26). Hodnota testovacieho kritéria je $F = 0,5327$. Kvantil Fisherovho F–rozdelenia je $F_{1-\alpha}(n_d - m, n_h - m) = F_{0,95}(2,2) = 19$ (Tabuľka 33). Kritická oblasť K_α je

$$K_{0,05} = (19; \infty) \Rightarrow F \notin K_{0,05}.$$

Hodnota testovacieho kritéria $F = 0,5327$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ je rozptyl náhodných chýb oboch skupín konštantný.

Posledná vlastnosť náhodných chýb, ktorú potrebujeme zistiť, je overenie vzájomnej nezávislosti náhodných chýb. K tomu použijeme Durbin–Watsonovu štatistiku DW (173). Platí

$$\sum_{i=2}^{8}(e_i - e_{i-1})^2 = (-17,9743 - 35,1348)^2 + (-16,8369 + 17,9743)^2 +$$
$$+(-6,0549 + 16,8369)^2 + (18,8644 + 6,0549)^2 + (-17,2163 - 18,8644)^2 +$$
$$+(-4,3777 + 17,2163)^2 + (8,4609 + 4,3777)^2 = 5190,895,$$
$$\sum_{i=1}^{8} e_i^2 = (-17,9743)^2 + (-17,2163)^2 + (-16,8369)^2 + (-6,0549)^2 +$$
$$+(-4,3777)^2 + (8,4609)^2 + (18,8644)^2 + (35,1348)^2 = 2620,6867.$$

Hodnotu Durbin–Watsonovej štatistiky určíme podľa vzťahu (173)

$$DW = \frac{5190,895}{2620,6867} = 1,98.$$

Hodnota Durbin–Watsonovej štatistiky $DW = 1,98$ patrí do intervalu $\langle 1,4; 2,6\rangle$, teda rezíduá nevykazujú autokoreláciu a preto je model dobrý.

Nakoniec urobíme overenie náhodnosti chýb pomocou testu náhodnosti chýb (testu bodov zvratu). Budeme testovať

$$H_0 : \text{chyby sú náhodné},$$
$$H_1 : \text{chyby nie sú náhodné}.$$

V Tabuľke 24 sú vypočítané rezíduá e_i. Potom body zvratu nájdeme nasledovne

$$35,1348 > \mathbf{-17{,}9743} < -16,8369 < -6,0549 < \mathbf{18{,}8644} > \mathbf{-17{,}2163} < -4,3777 < 8,4609.$$

Teda počet bodov zvratu je $Z = 3$. Hodnotu testovacieho kritéria U určíme podľa vzťahu (174)

$$U = \frac{3 - \frac{16-4}{3}}{\sqrt{\frac{128-29}{90}}} = -0,9535.$$

Kvantil normovaného normálneho rozdelenia je $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$ (Tabuľka 28). Kritická oblasť K_α je

$$K_{0,05} = (-\infty; -1,96) \cup (1,96; \infty) \Rightarrow U \notin K_{0,05}.$$

Hodnota testovacieho kritéria $U = -0,9535$ nepatrí do kritickej oblasti $K_{0,05}$, preto nulovú hypotézu H_0 nezamietame. Môžeme predpokladať, že na hladine významnosti $\alpha = 0,05$ sú chyby náhodné.

Náhodné chyby ε_i majú všetky požadované vlastnosti a teda môžeme daný lineárny regresný model považovať za kvalitný.

Literatúra

[1] Anděl, J., *Základy matematické statistiky*, Matfyzpress, Praha (2005).

[2] Andrejiová M., *Štatistické metódy v praxi*, Technická univerzita v Košiciach, Košice (2016).

[3] Buša J., Pirč V., Schrötter Š., *Numerické metódy, pravdepodobnosť a matematická štatistika*, Technická univerzita v Košiciach, Košice (2006).

[4] Chajdiak J., Rublíková E., Gudába M., *Štatistické metódy v praxi*, Statis, Bratislava, (1994).

[5] Knežo D., Andrejiová M., Ižaríková G., *Základné štatistické metódy*, Technická univerzita v Košiciach, Košice (2011).

[6] Ostertágová E., *Aplikovaná štatistika v počítačovom prostredí MATLABu*, Technická univerzita v Košiciach, Košice (2015).

[7] Riečan B., Lamoš F., Lenárt C., *Pravdepodobnosť a matematická štatistika*, Alfa, Bratislava, (1984).

[8] Skřivánková, V., Hančová, M., *Štatistika v príkladoch*, Univerzita Pavla Jozefa Šafárika v Košiciach, Košice (2005).

[9] Varga Š., *Matematická štatistika*, STU Bratislava, (2012).

[10] Volauf P., *Pravdepodobnosť a matematická štatistika*, STU Bratislava, (2014).

Štatistické tabuľky

Tabuľka 27: Distribučná funkcia normovaného normálneho rozdelenia

u	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5754
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6470	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7258	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7996	0,8023	0,8051	0,8079	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8598	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9430	0,9441
1,6	0,9452	0,9463	0,9474	0,9485	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9762	0,9767
2,0	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9865	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	1,0000	1,0000	1,0000
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabuľka 28: Kvantily normovaného normálneho rozdelenia u_α

α	u_α	α	u_α	α	u_α	α	u_α
0,500	0,00000	0,850	1,03643	0,930	1,47579	0,965	1,81191
0,510	0,02507	0,860	1,08032	0,931	1,48328	0,966	1,82501
0,520	0,05015	0,870	1,12639	0,932	1,49085	0,967	1,83842
0,530	0,07527	0,880	1,17499	0,933	1,49851	0,968	1,85218
0,540	0,10043	0,890	1,22653	0,934	1,50626	0,969	1,86630
0,550	0,12566	0,900	1,28155	0,935	1,51410	0,970	1,88079
0,560	0,15097	0,901	1,28727	0,936	1,52204	0,971	1,89570
0,570	0,17637	0,902	1,29303	0,937	1,53007	0,972	1,91104
0,580	0,20189	0,903	1,29884	0,938	1,53820	0,973	1,92684
0,590	0,22754	0,904	1,30469	0,939	1,54643	0,974	1,94313
0,600	0,25335	0,905	1,31058	0,940	1,55477	0,975	1,95996
0,610	0,27932	0,906	1,31652	0,941	1,56322	0,976	1,97737
0,620	0,30548	0,907	1,32251	0,942	1,57179	0,977	1,99539
0,630	0,33185	0,908	1,32854	0,943	1,58047	0,978	2,01409
0,640	0,35846	0,909	1,33462	0,944	1,58927	0,979	2,03352
0,650	0,38532	0,910	1,34076	0,945	1,59819	0,980	2,05375
0,660	0,41246	0,911	1,34694	0,946	1,60725	0,981	2,07485
0,670	0,43991	0,912	1,35317	0,947	1,61644	0,982	2,09693
0,680	0,46770	0,913	1,35946	0,948	1,62576	0,983	2,12007
0,690	0,49585	0,914	1,36581	0,949	1,63523	0,984	2,14441
0,700	0,52440	0,915	1,37220	0,950	1,64485	0,985	2,17009
0,710	0,55338	0,916	1,37866	0,951	1,65463	0,986	2,19729
0,720	0,58284	0,917	1,38517	0,952	1,66456	0,987	2,22621
0,730	0,61281	0,918	1,39174	0,953	1,67466	0,988	2,25713
0,740	0,64335	0,919	1,39838	0,954	1,68494	0,989	2,29037
0,750	0,67449	0,920	1,40507	0,955	1,69540	0,990	2,32635
0,760	0,70630	0,921	1,41183	0,956	1,70604	0,991	2,36562
0,770	0,73885	0,922	1,41865	0,957	1,71689	0,992	2,40892
0,780	0,77219	0,923	1,42554	0,958	1,72793	0,993	2,45726
0,790	0,80642	0,924	1,43250	0,959	1,73920	0,994	2,51214
0,800	0,84162	0,925	1,43953	0,960	1,75069	0,995	2,57583
0,810	0,87790	0,926	1,44663	0,961	1,76241	0,996	2,65207
0,820	0,91537	0,927	1,45381	0,962	1,77438	0,997	2,74778
0,830	0,95417	0,928	1,46106	0,963	1,78661	0,998	2,87816
0,840	0,99446	0,929	1,46838	0,964	1,79912	0,999	3,09023

Tabuľka 29: Kvantily Studentovho t–rozdelenia $t_\alpha(n)$

n	$\alpha = 0,900$	$\alpha = 0,950$	$\alpha = 0,975$	$\alpha = 0,990$	$\alpha = 0,995$	$\alpha = 0,999$
1	3,0777	6,3138	12,7062	31,8205	63,6567	318,3088
2	1,8856	2,9200	4,3027	6,9646	9,9248	22,3271
3	1,6377	2,3534	3,1824	4,5407	5,8409	10,2145
4	1,5332	2,1318	2,7764	3,7469	4,6041	7,1732
5	1,4759	2,0150	2,5706	3,3649	4,0321	5,8934
6	1,4398	1,9432	2,4469	3,1427	3,7074	5,2076
7	1,4149	1,8946	2,3646	2,9980	3,4995	4,7853
8	1,3968	1,8595	2,3060	2,8965	3,3554	4,5008
9	1,3830	1,8331	2,2622	2,8214	3,2498	4,2968
10	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437
11	1,3634	1,7959	2,2010	2,7181	3,1058	4,0247
12	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296
13	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520
14	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874
15	1,3406	1,7531	2,1314	2,6025	2,9467	3,7328
16	1,3368	1,7459	2,1199	2,5835	2,9208	3,6862
17	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458
18	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105
19	1,3277	1,7291	2,0930	2,5395	2,8609	3,5794
20	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518
21	1,3232	1,7207	2,0796	2,5176	2,8314	3,5272
22	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050
23	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850
24	1,3178	1,7109	2,0639	2,4922	2,7969	3,4668
25	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502
26	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350
27	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210
28	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082
29	1,3114	1,6991	2,0452	2,4620	2,7564	3,3962
30	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852
$>$30	1,2816	1,6449	1,9600	2,3263	2,5758	3,0000

Tabuľka 30: Kvantily χ^2–rozdelenia $\chi^2(n)$

α / n	$0,005$	$0,010$	$0,025$	$0,050$	$0,100$	$0,900$	$0,950$	$0,975$	$0,990$	$0,995$
1	0,00	0,00	0,00	0,00	0,00	2,71	3,84	5,02	6,64	7,88
2	0,01	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	6,25	7,82	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	21,95
9	1,74	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	4,58	5,58	17,28	19,68	21,92	24,72	26,76
12	3,07	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,83	7,04	19,81	22,36	24,74	27,69	29,82
14	4,08	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41	35,72
18	6,27	7,02	8,23	9,39	10,86	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,82	12,44	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	49,65
28	12,46	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	53,67
31	14,46	15,66	17,54	19,28	21,43	41,42	44,99	48,23	52,19	55,00
32	15,13	16,36	18,29	20,07	22,27	42,59	46,19	49,48	53,49	56,33
33	15,82	17,07	19,05	20,87	23,11	43,75	47,40	50,73	54,78	57,65
34	16,50	17,79	19,81	21,66	23,95	44,90	48,60	51,97	56,06	58,96
35	17,19	18,51	20,57	22,47	24,80	46,06	49,80	53,20	57,34	60,28

Tabuľka 31: Kvantily χ^2–rozdelenia $\chi^2(n)$

α / n	$0,005$	$0,010$	$0,025$	$0,050$	$0,100$	$0,900$	$0,950$	$0,975$	$0,990$	$0,995$
36	17,89	19,23	21,34	23,27	25,64	47,21	51,00	54,44	58,62	61,58
37	18,59	19,96	22,11	24,08	26,49	48,36	52,19	55,67	59,89	62,88
38	19,29	20,69	22,88	24,88	27,34	49,51	53,38	56,90	61,16	64,18
39	19,99	21,43	23,65	25,70	28,20	50,66	54,57	58,12	62,43	65,48
40	20,70	22,16	24,43	26,51	29,05	51,81	55,76	59,34	63,69	66,77
41	21,42	22,91	25,22	27,33	29,91	52,95	56,94	60,56	64,95	68,05
42	22,14	23,65	26,00	28,14	30,77	54,09	58,12	61,78	66,21	69,34
43	22,86	24,40	26,79	28,97	31,63	55,23	59,30	63,00	67,46	70,62
44	23,58	25,15	27,58	29,79	32,49	56,37	60,48	64,20	68,71	71,89
45	24,31	25,90	28,37	30,61	33,35	57,51	61,66	65,41	69,96	73,17
46	25,04	26,66	29,16	31,44	34,26	58,64	62,83	66,62	71,20	74,44
47	25,78	27,42	29,96	32,27	35,08	59,77	64,00	67,82	72,44	75,70
48	26,51	28,18	30,76	33,10	35,95	60,91	65,17	69,02	73,68	76,97
49	27,25	28,94	31,56	33,93	36,82	62,04	66,34	70,22	74,92	78,23
50	27,99	29,71	32,36	34,76	37,69	63,17	67,51	71,42	76,15	79,49
55	31,74	33,57	36,40	38,96	42,06	68,80	73,31	77,38	82,29	85,75
60	35,53	37,49	40,48	43,19	46,46	74,40	79,08	83,30	88,38	91,95
65	39,38	41,44	44,60	47,45	50,88	79,97	84,82	89,18	94,42	98,11
70	43,28	45,44	48,76	51,74	55,33	85,53	90,53	95,02	100,43	104,21
75	47,21	49,48	52,94	56,05	59,80	91,06	96,22	100,84	106,39	110,29
80	51,17	53,54	57,15	60,39	64,28	96,58	101,88	106,63	112,33	116,32
85	55,17	57,63	61,39	64,75	68,78	102,08	107,52	112,39	118,24	122,32
90	59,20	61,75	65,65	69,13	73,29	107,57	113,15	118,14	124,12	128,30
95	63,25	65,90	69,93	73,52	77,82	113,04	118,75	123,86	129,97	134,25
100	67,33	70,07	74,22	77,93	82,36	118,50	124,34	129,56	135,81	140,17

Tabuľka 32: Kvantily χ^2–rozdelenia $\chi^2(n)$

α / n	$0,005$	$0,010$	$0,025$	$0,050$	$0,100$	$0,900$	$0,950$	$0,975$	$0,990$	$0,995$
105	71,43	74,25	78,54	82,35	86,91	123,95	129,92	135,25	141,62	146,07
110	75,55	78,46	82,87	86,79	91,47	129,39	135,48	140,92	147,41	151,95
115	79,69	82,68	87,21	91,24	96,04	134,81	141,03	146,57	153,19	157,81
120	83,85	86,92	91,57	95,71	100,62	140,23	146,57	152,21	158,96	163,65
125	88,03	91,18	95,95	100,18	105,21	145,64	152,09	157,84	164,69	169,47
130	92,22	95,45	100,33	104,66	109,81	151,05	157,61	163,45	170,42	175,28
135	96,43	99,74	104,73	109,16	114,42	156,44	163,12	169,06	176,14	181,07
140	100,65	104,03	109,14	113,66	119,03	161,83	168,61	174,65	181,84	186,85
145	104,89	108,35	113,56	118,17	123,65	167,21	174,10	180,23	187,53	192,61
150	109,14	112,67	117,98	122,69	128,28	172,58	179,58	185,80	193,21	198,36
160	117,68	121,35	126,87	131,76	137,55	183,31	190,52	196,92	204,53	209,82
170	126,26	130,06	135,79	140,86	146,84	194,02	201,42	208,00	215,81	221,24
180	134,88	138,82	144,74	149,97	156,15	204,70	212,30	219,04	227,06	232,62
190	143,55	147,61	153,72	159,11	165,49	215,37	223,16	230,06	238,27	243,96
200	152,24	156,43	162,73	168,28	174,84	226,02	233,99	241,06	249,45	255,26
300	240,66	245,97	253,91	260,88	269,07	331,79	341,40	349,87	359,91	366,84
400	330,90	337,16	346,48	354,64	364,21	436,65	447,63	457,31	468,72	476,61
500	422,30	429,39	439,94	449,15	459,93	540,93	553,13	563,85	576,49	585,21
600	514,53	522,37	534,02	544,18	556,06	644,80	658,09	669,77	683,52	692,98
700	607,38	615,91	628,58	639,61	652,50	748,36	762,66	775,21	789,97	800,13
800	700,72	709,90	723,51	735,36	749,19	851,67	866,91	880,28	895,98	906,79
900	794,47	804,25	818,76	831,37	846,07	954,78	970,90	985,03	1001,6	1013,0
1000	888,56	898,91	914,26	927,59	943,13	1057,7	1074,7	1089,5	1107,0	1118,9

Tabuľka 33: Kvantily Fisherovho F–rozdelenia $F_{0,95}(k_1, k_2)$

k_1 / k_2	1	2	3	4	5	6	7
1	161,450	199,50	215,7074	224,5832	230,1619	233,986	236,7684
2	18,5128	19,000	19,1643	19,2468	19,2964	19,3295	19,3532
3	10,1280	9,5521	9,2766	9,1172	9,0135	8,9406	8,8867
4	7,7086	6,9443	6,5914	6,3882	6,2561	6,1631	6,0942
5	6,6079	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759
6	5,9874	5,1433	4,7571	4,5337	4,3874	4,2839	4,2067
7	5,5914	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870
8	5,3177	4,4590	4,0662	3,8379	3,6875	3,5806	3,5005
9	5,1174	4,2565	3,8625	3,6331	3,4817	3,3738	3,2927
10	4,9646	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355
11	4,8443	3,9823	3,5874	3,3567	3,2039	3,0946	3,0123
12	4,7472	3,8853	3,4903	3,2592	3,1059	2,9961	2,9134
13	4,6672	3,8056	3,4105	3,1791	3,0254	2,9153	2,8321
14	4,6001	3,7389	3,3439	3,1122	2,9582	2,8477	2,7642
15	4,5431	3,6823	3,2874	3,0556	2,9013	2,7905	2,7066
16	4,4940	3,6337	3,2389	3,0069	2,8524	2,7413	2,6572
17	4,4513	3,5915	3,1968	2,9647	2,8100	2,6987	2,6143
18	4,4139	3,5546	3,1599	2,9277	2,7729	2,6613	2,5767
19	4,3807	3,5219	3,1274	2,8951	2,7401	2,6283	2,5435
20	4,3512	3,4928	3,0984	2,8661	2,7109	2,5990	2,5140
21	4,3248	3,4668	3,0725	2,8401	2,6848	2,5727	2,4876
22	4,3009	3,4434	3,0491	2,8167	2,6613	2,5491	2,4638
23	4,2793	3,4221	3,0280	2,7955	2,6400	2,5277	2,4422
24	4,2597	3,4028	3,0088	2,7763	2,6207	2,5082	2,4226
25	4,2417	3,3852	2,9912	2,7587	2,6030	2,4904	2,4047
26	4,2252	3,3690	2,9752	2,7426	2,5868	2,4741	2,3883
27	4,2100	3,3541	2,9604	2,7278	2,5719	2,4591	2,3732
28	4,1960	3,3404	2,9467	2,7141	2,5581	2,4453	2,3593
29	4,1830	3,3277	2,9340	2,7014	2,5454	2,4324	2,3463
30	4,1709	3,3158	2,9223	2,6896	2,5336	2,4205	2,3343
40	4,0847	3,2317	2,8387	2,6060	2,4495	2,3359	2,2490
60	4,0012	3,1504	2,7581	2,5252	2,3683	2,2541	2,1665
80	3,9604	3,1108	2,7188	2,4859	2,3287	2,2142	2,1263
120	3,9201	3,0718	2,6802	2,4472	2,2899	2,1750	2,0868
> 120	3,8415	2,9957	2,6049	2,3719	2,2141	2,0986	2,0096

Tabuľka 34: Kvantily Fisherovho F–rozdelenia $F_{0,95}(k_1, k_2)$

k_1 / k_2	8	9	10	11	12	13	14
1	238,8827	240,5433	241,8818	242,9835	243,906	244,6899	245,364
2	19,3710	19,3848	19,3959	19,4050	19,4125	19,4189	19,4244
3	8,8452	8,8123	8,7855	8,7633	8,7446	8,7287	8,7149
4	6,0410	5,9988	5,9644	5,9358	5,9117	5,8911	5,8733
5	4,8183	4,7725	4,7351	4,7040	4,6777	4,6552	4,6358
6	4,1468	4,0990	4,0600	4,0274	3,9999	3,9764	3,9559
7	3,7257	3,6767	3,6365	3,6030	3,5747	3,5503	3,5292
8	3,4381	3,3881	3,3472	3,3130	3,2839	3,2590	3,2374
9	3,2296	3,1789	3,1373	3,1025	3,0729	3,0475	3,0255
10	3,0717	3,0204	2,9782	2,9430	2,9130	2,8872	2,8647
11	2,9480	2,8962	2,8536	2,8179	2,7876	2,7614	2,7386
12	2,8486	2,7964	2,7534	2,7173	2,6866	2,6602	2,6371
13	2,7669	2,7144	2,6710	2,6347	2,6037	2,5769	2,5536
14	2,6987	2,6458	2,6022	2,5655	2,5342	2,5073	2,4837
15	2,6408	2,5876	2,5437	2,5068	2,4753	2,4481	2,4244
16	2,5911	2,5377	2,4935	2,4564	2,4247	2,3973	2,3733
17	2,5480	2,4943	2,4499	2,4126	2,3807	2,3531	2,3290
18	2,5102	2,4563	2,4117	2,3742	2,3421	2,3143	2,2900
19	2,4768	2,4227	2,3779	2,3402	2,3080	2,2800	2,2556
20	2,4471	2,3928	2,3479	2,3100	2,2776	2,2495	2,2250
21	2,4205	2,3660	2,3210	2,2829	2,2504	2,2222	2,1975
22	2,3965	2,3419	2,2967	2,2585	2,2258	2,1975	2,1727
23	2,3748	2,3201	2,2747	2,2364	2,2036	2,1752	2,1502
24	2,3551	2,3002	2,2547	2,2163	2,1834	2,1548	2,1298
25	2,3371	2,2821	2,2365	2,1979	2,1649	2,1362	2,1111
26	2,3205	2,2655	2,2197	2,1811	2,1479	2,1192	2,0939
27	2,3053	2,2501	2,2043	2,1655	2,1323	2,1035	2,0781
28	2,2913	2,2360	2,1900	2,1512	2,1179	2,0889	2,0635
29	2,2783	2,2229	2,1768	2,1379	2,1045	2,0755	2,0500
30	2,2662	2,2107	2,1646	2,1256	2,0921	2,0630	2,0374
40	2,1802	2,1240	2,0772	2,0376	2,0035	1,9738	1,9476
60	2,0970	2,0401	1,9926	1,9522	1,9174	1,8870	1,8602
80	2,0564	1,9991	1,9512	1,9105	1,8753	1,8445	1,8174
120	2,0164	1,9588	1,9105	1,8693	1,8337	1,8026	1,7750
>120	1,9384	1,8799	1,8307	1,7886	1,7522	1,7202	1,6918

Tabuľka 35: Kvantily Fisherovho F–rozdelenia $F_{0,95}(k_1, k_2)$

k_1 / k_2	15	16	17	18	19	20	25
1	245,9499	246,4639	246,9184	247,3232	247,6861	248,0131	249,2601
2	19,4291	19,4333	19,437	19,4402	19,4431	19,4458	19,4558
3	8,7029	8,6923	8,6829	8,6745	8,6670	8,6602	8,6341
4	5,8578	5,8441	5,8320	5,8211	5,8114	5,8025	5,7687
5	4,6188	4,6038	4,5904	4,5785	4,5678	4,5581	4,5209
6	3,9381	3,9223	3,9083	3,8957	3,8844	3,8742	3,8348
7	3,5107	3,4944	3,4799	3,4669	3,4551	3,4445	3,4036
8	3,2184	3,2016	3,1867	3,1733	3,1613	3,1503	3,1081
9	3,0061	2,9890	2,9737	2,9600	2,9477	2,9365	2,8932
10	2,8450	2,8276	2,8120	2,7980	2,7854	2,7740	2,7298
11	2,7186	2,7009	2,6851	2,6709	2,6581	2,6464	2,6014
12	2,6169	2,5989	2,5828	2,5684	2,5554	2,5436	2,4977
13	2,5331	2,5149	2,4987	2,4841	2,4709	2,4589	2,4123
14	2,4630	2,4446	2,4282	2,4134	2,4000	2,3879	2,3407
15	2,4034	2,3849	2,3683	2,3533	2,3398	2,3275	2,2797
16	2,3522	2,3335	2,3167	2,3016	2,2880	2,2756	2,2272
17	2,3077	2,2888	2,2719	2,2567	2,2429	2,2304	2,1815
18	2,2686	2,2496	2,2325	2,2172	2,2033	2,1906	2,1413
19	2,2341	2,2149	2,1977	2,1823	2,1683	2,1555	2,1057
20	2,2033	2,1840	2,1667	2,1511	2,1370	2,1242	2,0739
21	2,1757	2,1563	2,1389	2,1232	2,1090	2,0960	2,0454
22	2,1508	2,1313	2,1138	2,0980	2,0837	2,0707	2,0196
23	2,1282	2,1086	2,0910	2,0751	2,0608	2,0476	1,9963
24	2,1077	2,0880	2,0703	2,0543	2,0399	2,0267	1,9750
25	2,0889	2,0691	2,0513	2,0353	2,0207	2,0075	1,9554
26	2,0716	2,0518	2,0339	2,0178	2,0032	1,9898	1,9375
27	2,0558	2,0358	2,0179	2,0017	1,9870	1,9736	1,9210
28	2,0411	2,0210	2,0030	1,9868	1,9720	1,9586	1,9057
29	2,0275	2,0073	1,9893	1,9730	1,9581	1,9446	1,8915
30	2,0148	1,9946	1,9765	1,9601	1,9452	1,9317	1,8782
40	1,9245	1,9037	1,8851	1,8682	1,8529	1,8389	1,7835
60	1,8364	1,8151	1,7959	1,7784	1,7625	1,7480	1,6902
80	1,7932	1,7716	1,7520	1,7342	1,7180	1,7032	1,6440
120	1,7505	1,7285	1,7085	1,6904	1,6739	1,6587	1,5980
> 120	1,6640	1,6435	1,6228	1,6038	1,5865	1,5705	1,5061

Tabuľka 36: Kvantily Fisherovho F–rozdelenia $F_{0,95}(k_1, k_2)$

k_1 / k_2	30	40	60	80	120	>120
1	250,0952	251,1432	252,1957	252,7237	253,2529	254,31
2	19,4624	19,4707	19,4791	19,4832	19,4874	19,496
3	8,6166	8,5944	8,5720	8,5607	8,5494	8,5264
4	5,7459	5,7170	5,6877	5,6730	5,6581	5,6281
5	4,4957	4,4638	4,4314	4,4150	4,3985	4,3650
6	3,8082	3,7743	3,7398	3,7223	3,7047	3,6689
7	3,3758	3,3404	3,3043	3,2860	3,2674	3,2298
8	3,0794	3,0428	3,0053	2,9862	2,9669	2,9276
9	2,8637	2,8259	2,7872	2,7675	2,7475	2,7067
10	2,6996	2,6609	2,6211	2,6008	2,5801	2,5379
11	2,5705	2,5309	2,4901	2,4692	2,4480	2,4045
12	2,4663	2,4259	2,3842	2,3628	2,3410	2,2962
13	2,3803	2,3392	2,2966	2,2747	2,2524	2,2064
14	2,3082	2,2664	2,2229	2,2006	2,1778	2,1307
15	2,2468	2,2043	2,1601	2,1373	2,1141	2,0658
16	2,1938	2,1507	2,1058	2,0826	2,0589	2,0096
17	2,1477	2,1040	2,0584	2,0348	2,0107	1,9604
18	2,1071	2,0629	2,0166	1,9927	1,9681	1,9168
19	2,0712	2,0264	1,9795	1,9552	1,9302	1,8780
20	2,0391	1,9938	1,9464	1,9217	1,8963	1,8432
21	2,0102	1,9645	1,9165	1,8915	1,8657	1,8117
22	1,9842	1,9380	1,8894	1,8641	1,8380	1,7831
23	1,9605	1,9139	1,8648	1,8392	1,8128	1,7570
24	1,9390	1,8920	1,8424	1,8164	1,7896	1,7330
25	1,9192	1,8718	1,8217	1,7955	1,7684	1,7110
26	1,9010	1,8533	1,8027	1,7762	1,7488	1,6906
27	1,8842	1,8361	1,7851	1,7584	1,7306	1,6717
28	1,8687	1,8203	1,7689	1,7418	1,7138	1,6541
29	1,8543	1,8055	1,7537	1,7264	1,6981	1,6376
30	1,8409	1,7918	1,7396	1,7121	1,6835	1,6223
40	1,7444	1,6928	1,6373	1,6077	1,5766	1,5089
60	1,6491	1,5943	1,5343	1,5019	1,4673	1,3893
80	1,6017	1,5449	1,4821	1,4477	1,4107	1,3247
120	1,5543	1,4952	1,4290	1,3922	1,3519	1,2539
> 120	1,4591	1,3940	1,3180	1,2735	1,2214	1,0000

Tabuľka 37: Kvantily Fisherovho F–rozdelenia $F_{0,975}(k_1, k_2)$

k_1 / k_2	1	2	3	4	5	6	7
1	647,789	799,500	864,163	899,5833	921,8479	937,1111	948,2169
2	38,5063	39,0000	39,1655	39,2484	39,2982	39,3315	39,3552
3	17,4434	16,0441	15,4392	15,1010	14,8848	14,7347	14,6244
4	12,2179	10,6491	9,9792	9,6045	9,3645	9,1973	9,0741
5	10,0070	8,4336	7,7636	7,3879	7,1464	6,9777	6,8531
6	8,8131	7,2599	6,5988	6,2272	5,9876	5,8198	5,6955
7	8,0727	6,5415	5,8898	5,5226	5,2852	5,1186	4,9949
8	7,5709	6,0595	5,4160	5,0526	4,8173	4,6517	4,5286
9	7,2093	5,7147	5,0781	4,7181	4,4844	4,3197	4,1970
10	6,9367	5,4564	4,8256	4,4683	4,2361	4,0721	3,9498
11	6,7241	5,2559	4,6300	4,2751	4,0440	3,8807	3,7586
12	6,5538	5,0959	4,4742	4,1212	3,8911	3,7283	3,6065
13	6,4143	4,9653	4,3472	3,9959	3,7667	3,6043	3,4827
14	6,2979	4,8567	4,2417	3,8919	3,6634	3,5014	3,3799
15	6,1995	4,7650	4,1528	3,8043	3,5764	3,4147	3,2934
16	6,1151	4,6867	4,0768	3,7294	3,5021	3,3406	3,2194
17	6,0420	4,6189	4,0112	3,6648	3,4379	3,2767	3,1556
18	5,9781	4,5597	3,9539	3,6083	3,3820	3,2209	3,0999
19	5,9216	4,5075	3,9034	3,5587	3,3327	3,1718	3,0509
20	5,8715	4,4613	3,8587	3,5147	3,2891	3,1283	3,0074
21	5,8266	4,4199	3,8188	3,4754	3,2501	3,0895	2,9686
22	5,7863	4,3828	3,7829	3,4401	3,2151	3,0546	2,9338
23	5,7498	4,3492	3,7505	3,4083	3,1835	3,0232	2,9023
24	5,7166	4,3187	3,7211	3,3794	3,1548	2,9946	2,8738
25	5,6864	4,2909	3,6943	3,3530	3,1287	2,9685	2,8478
26	5,6586	4,2655	3,6697	3,3289	3,1048	2,9447	2,8240
27	5,6331	4,2421	3,6472	3,3067	3,0828	2,9228	2,8021
28	5,6096	4,2205	3,6264	3,2863	3,0626	2,9027	2,7820
29	5,5878	4,2006	3,6072	3,2674	3,0438	2,8840	2,7633
30	5,5675	4,1821	3,5894	3,2499	3,0265	2,8667	2,7460
40	5,4239	4,0510	3,4633	3,1261	2,9037	2,7444	2,6238
60	5,2856	3,9253	3,3425	3,0077	2,7863	2,6274	2,5068
80	5,2184	3,8643	3,2841	2,9504	2,7295	2,5708	2,4502
120	5,1523	3,8046	3,2269	2,8943	2,6740	2,5154	2,3948
> 120	5,0239	3,6889	3,1161	2,7858	2,5665	2,4082	2,2875

Tabuľka 38: Kvantily Fisherovho F–rozdelenia $F_{0,975}(k_1, k_2)$

k_1 / k_2	8	9	10	11	12	13	14
1	956,6562	963,2846	968,6274	973,0252	976,708	979,8368	982,5278
2	39,3730	39,3869	39,3980	39,4071	39,4146	39,4210	39,4265
3	14,5399	14,4731	14,4189	14,3742	14,3366	14,3045	14,2768
4	8,9796	8,9047	8,8439	8,7935	8,7512	8,7150	8,6838
5	6,7572	6,6811	6,6192	6,5678	6,5245	6,4876	6,4556
6	5,5996	5,5234	5,4613	5,4098	5,3662	5,3290	5,2968
7	4,8993	4,8232	4,7611	4,7095	4,6658	4,6285	4,5961
8	4,4333	4,3572	4,2951	4,2434	4,1997	4,1622	4,1297
9	4,1020	4,0260	3,9639	3,9121	3,8682	3,8306	3,7980
10	3,8549	3,7790	3,7168	3,6649	3,6209	3,5832	3,5504
11	3,6638	3,5879	3,5257	3,4737	3,4296	3,3917	3,3588
12	3,5118	3,4358	3,3736	3,3215	3,2773	3,2393	3,2062
13	3,3880	3,3120	3,2497	3,1975	3,1532	3,1150	3,0819
14	3,2853	3,2093	3,1469	3,0946	3,0502	3,0119	2,9786
15	3,1987	3,1227	3,0602	3,0078	2,9633	2,9249	2,8915
16	3,1248	3,0488	2,9862	2,9337	2,8890	2,8506	2,8170
17	3,0610	2,9849	2,9222	2,8696	2,8249	2,7863	2,7526
18	3,0053	2,9291	2,8664	2,8137	2,7689	2,7302	2,6964
19	2,9563	2,8801	2,8172	2,7645	2,7196	2,6808	2,6469
20	2,9128	2,8365	2,7737	2,7209	2,6758	2,6369	2,6030
21	2,8740	2,7977	2,7348	2,6819	2,6368	2,5978	2,5638
22	2,8392	2,7628	2,6998	2,6469	2,6017	2,5626	2,5285
23	2,8077	2,7313	2,6682	2,6152	2,5699	2,5308	2,4966
24	2,7791	2,7027	2,6396	2,5865	2,5411	2,5019	2,4677
25	2,7531	2,6766	2,6135	2,5603	2,5149	2,4756	2,4413
26	2,7293	2,6528	2,5896	2,5363	2,4908	2,4515	2,4171
27	2,7074	2,6309	2,5676	2,5143	2,4688	2,4293	2,3949
28	2,6872	2,6106	2,5473	2,4940	2,4484	2,4089	2,3743
29	2,6686	2,5919	2,5286	2,4752	2,4295	2,3900	2,3554
30	2,6513	2,5746	2,5112	2,4577	2,4120	2,3724	2,3378
40	2,5289	2,4519	2,3882	2,3343	2,2882	2,2481	2,2130
60	2,4117	2,3344	2,2702	2,2159	2,1692	2,1286	2,0929
80	2,3549	2,2775	2,2130	2,1584	2,1115	2,0706	2,0346
120	2,2994	2,2217	2,1570	2,1021	2,0548	2,0136	1,9773
> 120	2,1918	2,1136	2,0483	1,9927	1,9447	1,9027	1,8656

Tabuľka 39: Kvantily Fisherovho F–rozdelenia $F_{0,975}(k_1, k_2)$

k_1 / k_2	15	16	17	18	19	20	25
1	984,8668	986,9187	988,7331	990,3490	991,7973	993,1028	998,0808
2	39,4313	39,4354	39,4391	39,4424	39,4453	39,4479	39,4579
3	14,2527	14,2315	14,2127	14,1960	14,1810	14,1674	14,1155
4	8,6565	8,6326	8,6113	8,5924	8,5753	8,5599	8,5010
5	6,4277	6,4032	6,3814	6,3619	6,3444	6,3286	6,2679
6	5,2687	5,2439	5,2218	5,2021	5,1844	5,1684	5,1069
7	4,5678	4,5428	4,5206	4,5008	4,4829	4,4667	4,4045
8	4,1012	4,0761	4,0538	4,0338	4,0158	3,9995	3,9367
9	3,7694	3,7441	3,7216	3,7015	3,6833	3,6669	3,6035
10	3,5217	3,4963	3,4737	3,4534	3,4351	3,4185	3,3546
11	3,3299	3,3044	3,2816	3,2612	3,2428	3,2261	3,1616
12	3,1772	3,1515	3,1286	3,1081	3,0896	3,0728	3,0077
13	3,0527	3,0269	3,0039	2,9832	2,9646	2,9477	2,8821
14	2,9493	2,9234	2,9003	2,8795	2,8607	2,8437	2,7777
15	2,8621	2,8360	2,8128	2,7919	2,7730	2,7559	2,6894
16	2,7875	2,7614	2,7380	2,7170	2,6980	2,6808	2,6138
17	2,7230	2,6968	2,6733	2,6522	2,6331	2,6158	2,5484
18	2,6667	2,6404	2,6168	2,5956	2,5764	2,5590	2,4912
19	2,6171	2,5907	2,5670	2,5457	2,5265	2,5089	2,4408
20	2,5731	2,5465	2,5228	2,5014	2,4821	2,4645	2,3959
21	2,5338	2,5071	2,4833	2,4618	2,4424	2,4247	2,3558
22	2,4984	2,4717	2,4478	2,4262	2,4067	2,3890	2,3198
23	2,4665	2,4396	2,4157	2,3940	2,3745	2,3567	2,2871
24	2,4374	2,4105	2,3865	2,3648	2,3452	2,3273	2,2574
25	2,4110	2,3840	2,3599	2,3381	2,3184	2,3005	2,2303
26	2,3867	2,3597	2,3355	2,3137	2,2939	2,2759	2,2054
27	2,3644	2,3373	2,3131	2,2912	2,2713	2,2533	2,1826
28	2,3438	2,3167	2,2924	2,2704	2,2505	2,2324	2,1615
29	2,3248	2,2976	2,2732	2,2512	2,2313	2,2131	2,1419
30	2,3072	2,2799	2,2554	2,2334	2,2134	2,1952	2,1237
40	2,1819	2,1542	2,1293	2,1068	2,0864	2,0677	1,9943
60	2,0613	2,0330	2,0076	1,9846	1,9636	1,9445	1,8687
80	2,0026	1,9741	1,9483	1,9250	1,9037	1,8843	1,8071
120	1,9450	1,9161	1,8900	1,8663	1,8447	1,8249	1,7462
> 120	1,8326	1,8028	1,7759	1,7515	1,7291	1,7085	1,6259

Tabuľka 40: Kvantily Fisherovho F–rozdelenia $F_{0,975}(k_1, k_2)$

k_1 / k_2	30	40	60	80	120	>120
1	1001,414	1005,598	1009,800	1011,908	1014,020	1018,300
2	39,4646	39,4729	39,4812	39,4854	39,4896	39,4980
3	14,0805	14,0365	13,9921	13,9697	13,9473	13,9020
4	8,4613	8,4111	8,3604	8,3349	8,3092	8,2573
5	6,2269	6,1750	6,1225	6,0960	6,0693	6,0153
6	5,0652	5,0125	4,9589	4,9318	4,9044	4,8491
7	4,3624	4,3089	4,2544	4,2268	4,1989	4,1423
8	3,8940	3,8398	3,7844	3,7563	3,7279	3,6702
9	3,5604	3,5055	3,4493	3,4207	3,3918	3,3329
10	3,3110	3,2554	3,1984	3,1694	3,1399	3,0798
11	3,1176	3,0613	3,0035	2,9740	2,9441	2,8828
12	2,9633	2,9063	2,8478	2,8178	2,7874	2,7249
13	2,8372	2,7797	2,7204	2,6900	2,6590	2,5955
14	2,7324	2,6742	2,6142	2,5833	2,5519	2,4872
15	2,6437	2,5850	2,5242	2,4930	2,4611	2,3953
16	2,5678	2,5085	2,4471	2,4154	2,3831	2,3163
17	2,5020	2,4422	2,3801	2,3481	2,3153	2,2474
18	2,4445	2,3842	2,3214	2,2890	2,2558	2,1869
19	2,3937	2,3329	2,2696	2,2368	2,2032	2,1333
20	2,3486	2,2873	2,2234	2,1902	2,1562	2,0853
21	2,3082	2,2465	2,1819	2,1485	2,1141	2,0422
22	2,2718	2,2097	2,1446	2,1108	2,0760	2,0032
23	2,2389	2,1763	2,1107	2,0766	2,0415	1,9677
24	2,2090	2,1460	2,0799	2,0454	2,0099	1,9353
25	2,1816	2,1183	2,0516	2,0169	1,9811	1,9055
26	2,1565	2,0928	2,0257	1,9907	1,9545	1,8781
27	2,1334	2,0693	2,0018	1,9665	1,9299	1,8527
28	2,1121	2,0477	1,9797	1,9441	1,9072	1,8291
29	2,0923	2,0276	1,9591	1,9232	1,8861	1,8072
30	2,0739	2,0089	1,9400	1,9039	1,8664	1,7867
40	1,9429	1,8752	1,8028	1,7644	1,7242	1,6371
60	1,8152	1,7440	1,6668	1,6252	1,5810	1,4821
80	1,7523	1,6790	1,5987	1,5549	1,5079	1,3997
120	1,6899	1,6141	1,5299	1,4834	1,4327	1,3104
> 120	1,5660	1,4835	1,3883	1,3329	1,2684	1,0000

Tabuľka 41: Kvantily Fisherovho F–rozdelenia $F_{0,995}(k_1, k_2)$

k_1 / k_2	1	2	3	4	5	6	7
1	16211	20000	21615	22500	23056	23438	23715
2	198,50	199,00	199,17	199,25	199,30	199,33	199,36
3	55,5520	49,7993	47,4672	46,1946	45,3916	44,8385	44,4341
4	31,3328	26,2843	24,2591	23,1545	22,4564	21,9746	21,6217
5	22,7848	18,3138	16,5298	15,5561	14,9396	14,5133	14,2004
6	18,6350	14,5441	12,9166	12,0275	11,4637	11,0730	10,7859
7	16,2356	12,4040	10,8824	10,0505	9,5221	9,1553	8,8854
8	14,6882	11,0424	9,5965	8,8051	8,3018	7,9520	7,6941
9	13,6136	10,1067	8,7171	7,9559	7,4712	7,1339	6,8849
10	12,8265	9,4270	8,0807	7,3428	6,8724	6,5446	6,3025
11	12,2263	8,9122	7,6004	6,8809	6,4217	6,1016	5,8648
12	11,7542	8,5096	7,2258	6,5211	6,0711	5,7570	5,5245
13	11,3735	8,1865	6,9258	6,2335	5,7910	5,4819	5,2529
14	11,0603	7,9216	6,6804	5,9984	5,5623	5,2574	5,0313
15	10,7980	7,7008	6,4760	5,8029	5,3721	5,0708	4,8473
16	10,5755	7,5138	6,3034	5,6378	5,2117	4,9134	4,6920
17	10,3842	7,3536	6,1556	5,4967	5,0746	4,7789	4,5594
18	10,2181	7,2148	6,0278	5,3746	4,9560	4,6627	4,4448
19	10,0725	7,0935	5,9161	5,2681	4,8526	4,5614	4,3448
20	9,9439	6,9865	5,8177	5,1743	4,7616	4,4721	4,2569
21	9,8295	6,8914	5,7304	5,0911	4,6809	4,3931	4,1789
22	9,7271	6,8064	5,6524	5,0168	4,6088	4,3225	4,1094
23	9,6348	6,7300	5,5823	4,9500	4,5441	4,2591	4,0469
24	9,5513	6,6609	5,5190	4,8898	4,4857	4,2019	3,9905
25	9,4753	6,5982	5,4615	4,8351	4,4327	4,1500	3,9394
26	9,4059	6,5409	5,4091	4,7852	4,3844	4,1027	3,8928
27	9,3423	6,4885	5,3611	4,7396	4,3402	4,0594	3,8501
28	9,2838	6,4403	5,3170	4,6977	4,2996	4,0197	3,8110
29	9,2297	6,3958	5,2764	4,6591	4,2622	3,9831	3,7749
30	9,1797	6,3547	5,2388	4,6234	4,2276	3,9492	3,7416
40	8,8279	6,0664	4,9758	4,3738	3,9860	3,7129	3,5088
60	8,4946	5,7950	4,7290	4,1399	3,7599	3,4918	3,2911
80	8,3346	5,6652	4,6113	4,0285	3,6524	3,3867	3,1876
120	8,1788	5,5393	4,4972	3,9207	3,5482	3,2849	3,0874
$>$ 120	7,8795	5,2983	4,2794	3,7151	3,3499	3,0913	2,8968

Tabuľka 42: Kvantily Fisherovho F–rozdelenia $F_{0,995}(k_1, k_2)$

k_1 / k_2	8	9	10	11	12	13	14
1	23925	24091	24225	24334	24426	24505	24572
2	199,38	199,39	199,40	199,41	199,42	199,43	199,43
3	44,1256	43,8824	43,6858	43,5236	43,3874	43,2715	43,1716
4	21,3520	21,1391	20,9667	20,8243	20,7047	20,6027	20,5148
5	13,9610	13,7716	13,6182	13,4912	13,3845	13,2934	13,2148
6	10,5658	10,3915	10,2500	10,1329	10,0343	9,9501	9,8774
7	8,6781	8,5138	8,3803	8,2697	8,1764	8,0967	8,0279
8	7,4959	7,3386	7,2106	7,1045	7,0149	6,9384	6,8721
9	6,6933	6,5411	6,4172	6,3142	6,2274	6,1530	6,0887
10	6,1159	5,9676	5,8467	5,7462	5,6613	5,5887	5,5257
11	5,6821	5,5368	5,4183	5,3197	5,2363	5,1649	5,1031
12	5,3451	5,2021	5,0855	4,9884	4,9062	4,8358	4,7748
13	5,0761	4,9351	4,8199	4,7240	4,6429	4,5733	4,5129
14	4,8566	4,7173	4,6034	4,5085	4,4281	4,3591	4,2993
15	4,6744	4,5364	4,4235	4,3295	4,2497	4,1813	4,1219
16	4,5207	4,3838	4,2719	4,1785	4,0994	4,0314	3,9723
17	4,3894	4,2535	4,1424	4,0496	3,9709	3,9033	3,8445
18	4,2759	4,1410	4,0305	3,9382	3,8599	3,7926	3,7341
19	4,1770	4,0428	3,9329	3,8410	3,7631	3,6961	3,6378
20	4,0900	3,9564	3,8470	3,7555	3,6779	3,6111	3,5530
21	4,0128	3,8799	3,7709	3,6798	3,6024	3,5358	3,4779
22	3,9440	3,8116	3,7030	3,6122	3,5350	3,4686	3,4108
23	3,8822	3,7502	3,6420	3,5515	3,4745	3,4083	3,3506
24	3,8264	3,6949	3,5870	3,4967	3,4199	3,3538	3,2962
25	3,7758	3,6447	3,5370	3,4470	3,3704	3,3044	3,2469
26	3,7297	3,5989	3,4916	3,4017	3,3252	3,2594	3,2020
27	3,6875	3,5571	3,4499	3,3602	3,2839	3,2182	3,1608
28	3,6487	3,5186	3,4117	3,3222	3,2460	3,1803	3,1231
29	3,6131	3,4832	3,3765	3,2871	3,2110	3,1454	3,0882
30	3,5801	3,4505	3,3440	3,2547	3,1787	3,1132	3,0560
40	3,3498	3,2220	3,1167	3,0284	2,9531	2,8880	2,8312
60	3,1344	3,0083	2,9042	2,8166	2,7419	2,6771	2,6205
80	3,0320	2,9066	2,8031	2,7159	2,6413	2,5767	2,5201
120	2,9330	2,8083	2,7052	2,6183	2,5439	2,4794	2,4228
> 120	2,7444	2,6211	2,5188	2,4325	2,3583	2,2938	2,2371

Tabuľka 43: Kvantily Fisherovho F–rozdelenia $F_{0,995}(k_1, k_2)$

k_1 / k_2	15	16	17	18	19	20	25
1	24630	24682	24727	24767	24803	24836	24960
2	199,43	199,44	199,44	199,44	199,45	199,45	199,46
3	43,0847	43,0083	42,9407	42,8804	42,8263	42,7775	42,5910
4	20,4383	20,3710	20,3113	20,2581	20,2104	20,1673	20,0024
5	13,1463	13,0861	13,0327	12,9850	12,9422	12,9035	12,7554
6	9,8140	9,7582	9,7086	9,6644	9,6247	9,5888	9,4511
7	7,9678	7,9148	7,8678	7,8258	7,7881	7,7540	7,6230
8	6,8143	6,7633	6,7180	6,6775	6,6411	6,6082	6,4817
9	6,0325	5,9829	5,9388	5,8994	5,8639	5,8318	5,7084
10	5,4707	5,4221	5,3789	5,3403	5,3055	5,2740	5,1528
11	5,0489	5,0011	4,9586	4,9205	4,8863	4,8552	4,7356
12	4,7213	4,6741	4,6321	4,5945	4,5606	4,5299	4,4115
13	4,4600	4,4132	4,3716	4,3344	4,3008	4,2703	4,1528
14	4,2468	4,2005	4,1592	4,1221	4,0888	4,0585	3,9417
15	4,0698	4,0237	3,9827	3,9459	3,9127	3,8826	3,7662
16	3,9205	3,8747	3,8338	3,7972	3,7641	3,7342	3,6182
17	3,7929	3,7473	3,7066	3,6701	3,6372	3,6073	3,4916
18	3,6827	3,6373	3,5967	3,5603	3,5275	3,4977	3,3822
19	3,5866	3,5412	3,5008	3,4645	3,4318	3,4020	3,2867
20	3,5020	3,4568	3,4164	3,3802	3,3475	3,3178	3,2025
21	3,4270	3,3818	3,3416	3,3054	3,2728	3,2431	3,1279
22	3,3600	3,3150	3,2748	3,2387	3,2060	3,1764	3,0613
23	3,2999	3,2549	3,2148	3,1787	3,1461	3,1165	3,0014
24	3,2456	3,2007	3,1606	3,1246	3,0920	3,0624	2,9472
25	3,1963	3,1515	3,1114	3,0754	3,0429	3,0133	2,8981
26	3,1515	3,1067	3,0666	3,0306	2,9981	2,9685	2,8533
27	3,1104	3,0656	3,0256	2,9896	2,9571	2,9275	2,8123
28	3,0727	3,0279	2,9879	2,9520	2,9194	2,8899	2,7746
29	3,0379	2,9932	2,9532	2,9173	2,8847	2,8551	2,7398
30	3,0057	2,9611	2,9211	2,8852	2,8526	2,8230	2,7076
40	2,7811	2,7365	2,6966	2,6607	2,6281	2,5984	2,4823
60	2,5705	2,5259	2,4859	2,4498	2,4171	2,3872	2,2697
80	2,4700	2,4254	2,3854	2,3492	2,3163	2,2862	2,1678
120	2,3727	2,3280	2,2878	2,2514	2,2183	2,1881	2,0686
>120	2,1868	2,1417	2,1011	2,0642	2,0306	1,9998	1,8771

Tabuľka 44: Kvantily Fisherovho F–rozdelenia $F_{0,995}(k_1, k_2)$

k_1 / k_2	30	40	60	80	120	>120
1	25044	25148	25253	25306	25359	25464
2	199,47	199,47	199,50	199,50	199,50	199,50
3	42,4658	42,3082	42,1494	42,0696	41,9895	41,8283
4	19,8915	19,7518	19,6107	19,5397	19,4684	19,3247
5	12,6556	12,5297	12,4024	12,3383	12,2737	12,1435
6	9,3582	9,2408	9,1219	9,0619	9,0015	8,8793
7	7,5345	7,4224	7,3088	7,2513	7,1933	7,0760
8	6,3961	6,2875	6,1772	6,1213	6,0649	5,9506
9	5,6248	5,5186	5,4104	5,3555	5,3001	5,1875
10	5,0706	4,9659	4,8592	4,8050	4,7501	4,6385
11	4,6543	4,5508	4,4450	4,3912	4,3367	4,2255
12	4,3309	4,2282	4,1229	4,0693	4,0149	3,9039
13	4,0727	3,9704	3,8655	3,8120	3,7577	3,6465
14	3,8619	3,7600	3,6552	3,6017	3,5473	3,4359
15	3,6867	3,5850	3,4803	3,4267	3,3722	3,2602
16	3,5389	3,4372	3,3324	3,2787	3,2240	3,1115
17	3,4124	3,3108	3,2058	3,1520	3,0971	2,9839
18	3,3030	3,2014	3,0962	3,0422	2,9871	2,8732
19	3,2075	3,1058	3,0004	2,9462	2,8908	2,7762
20	3,1234	3,0215	2,9159	2,8614	2,8058	2,6904
21	3,0488	2,9467	2,8408	2,7861	2,7302	2,6140
22	2,9821	2,8799	2,7736	2,7187	2,6625	2,5455
23	2,9221	2,8197	2,7132	2,6581	2,6015	2,4837
24	2,8679	2,7654	2,6585	2,6031	2,5463	2,4276
25	2,8187	2,7160	2,6088	2,5532	2,4961	2,3765
26	2,7738	2,6709	2,5633	2,5075	2,4501	2,3297
27	2,7327	2,6296	2,5217	2,4656	2,4079	2,2867
28	2,6949	2,5916	2,4834	2,4270	2,3690	2,2470
29	2,6600	2,5565	2,4479	2,3914	2,3331	2,2102
30	2,6278	2,5241	2,4151	2,3584	2,2998	2,1760
40	2,4015	2,2958	2,1838	2,1249	2,0636	1,9318
60	2,1874	2,0789	1,9622	1,8998	1,8341	1,6885
80	2,0845	1,9739	1,8540	1,7892	1,7203	1,5634
120	1,9840	1,8709	1,7469	1,6789	1,6055	1,4311
> 120	1,7891	1,6691	1,5325	1,4540	1,3637	1,0036

Tabuľka 45: Kritické hodnoty k_1 znamienkového testu

n_0	$\alpha = 0,05$	$\alpha = 0,01$	n_0	$\alpha = 0,05$	$\alpha = 0,01$
6	0	-	31	9	7
7	0	-	32	9	8
8	0	0	33	10	8
9	1	0	34	10	9
10	1	0	35	11	9
11	1	0	36	11	9
12	2	1	37	12	10
13	2	1	38	12	10
14	2	1	39	12	11
15	3	2	40	13	11
16	3	2	41	13	11
17	4	2	42	14	12
18	4	3	43	14	12
19	4	3	44	15	13
20	5	3	45	15	13
21	5	4	46	15	13
22	5	4	47	16	14
23	6	4	48	16	14
24	6	5	49	17	15
25	7	5	50	17	15
26	7	6	51	18	15
27	7	6	52	18	16
28	8	6	53	18	16
29	8	7	54	19	17
30	9	7	55	19	17

Tabuľka 46: Kritické hodnoty $U_\alpha(n_1, n_2)$ Wilcoxonovho dvojvýberového testu pre $\alpha = 0,05$

n_1 / n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	0	1	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	1	2	3	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	1	3	5	6	8	-	-	-	-	-	-	-	-	-	-	-	-	-
8	0	2	4	6	8	10	13	-	-	-	-	-	-	-	-	-	-	-	-
9	0	2	4	7	10	12	15	17	-	-	-	-	-	-	-	-	-	-	-
10	0	3	5	8	11	14	17	20	23	-	-	-	-	-	-	-	-	-	-
11	0	3	6	9	13	16	19	23	26	30	-	-	-	-	-	-	-	-	-
12	1	4	7	11	14	18	22	26	29	33	37	-	-	-	-	-	-	-	-
13	1	4	8	12	16	20	24	28	33	37	41	45	-	-	-	-	-	-	-
14	1	5	9	13	17	22	26	31	36	40	45	50	55	-	-	-	-	-	-
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64	-	-	-	-	-
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	-	-	-	-
17	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87	-	-	-
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	-	-
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	-
20	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
21	2	8	15	22	29	36	43	50	58	65	73	80	88	96	103	111	119	126	134
22	3	9	16	23	30	38	45	53	61	69	77	85	93	101	109	117	125	133	141
23	3	9	17	24	32	40	48	56	64	73	81	89	98	106	115	123	132	140	149
24	3	10	17	25	33	42	50	59	67	76	85	94	102	111	120	129	138	147	156
25	3	10	18	27	35	44	53	62	71	80	89	98	107	117	126	135	145	154	161
26	4	11	19	28	37	46	55	64	74	83	93	102	112	122	132	141	151	161	171
27	4	11	20	29	38	48	57	67	77	87	97	107	117	127	137	147	158	168	178
28	4	12	21	30	40	50	60	70	80	90	101	111	122	132	143	154	164	175	186
29	4	13	22	32	42	52	62	73	83	94	105	116	127	138	149	160	171	182	193
30	5	13	23	33	43	54	65	76	87	98	109	120	131	143	154	166	177	189	200

Tabuľka 47: Kritické hodnoty $D_\alpha(n)$ pre Kolmogorovov–Smirnovov test

n	$\alpha = 0,05$	$\alpha = 0,01$	n	$\alpha = 0,05$	$\alpha = 0,01$
1	0,9750	0,9950	31	0,2379	0,2853
2	0,8419	0,9293	32	0,2342	0,2809
3	0,7076	0,8290	33	0,2308	0,2768
4	0,6239	0,7342	34	0,2274	0,2728
5	0,5633	0,6685	35	0,2243	0,2690
6	0,5193	0,6166	36	0,2212	0,2653
7	0,4834	0,5758	37	0,2183	0,2618
8	0,4543	0,5418	38	0,2154	0,2584
9	0,4300	0,5133	39	0,2127	0,2552
10	0,4093	0,4889	40	0,2101	0,2521
11	0,3912	0,4677	41	0,2076	0,2490
12	0,3754	0,4491	42	0,2052	0,2461
13	0,3614	0,4325	43	0,2028	0,2433
14	0,3489	0,4176	44	0,2006	0,2406
15	0,3376	0,4042	45	0,1984	0,2380
16	0,3273	0,3920	46	0,1963	0,2534
17	0,3180	0,3809	47	0,1942	0,2330
18	0,3094	0,3706	48	0,1922	0,2306
19	0,3014	0,3612	49	0,1903	0,2283
20	0,2941	0,3524	50	0,1884	0,2260
21	0,2873	0,3443	51	0,1866	0,2239
22	0,2809	0,3367	52	0,1848	0,2217
23	0,2749	0,3295	53	0,1831	0,2197
24	0,2693	0,3229	54	0,1814	0,2177
25	0,2640	0,3166	55	0,1798	0,2157
26	0,2591	0,3106	56	0,1782	0,2138
27	0,2544	0,3050	57	0,1767	0,2120
28	0,2499	0,2997	58	0,1752	0,2102
29	0,2457	0,2947	59	0,1737	0,2084
30	0,2417	0,2899	60	0,1723	0,2067

Tabuľka 48: Váhy $a_{i,n}$ pre Shapiro–Wilkov test normality

i	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$
1	0,7007	1,7007	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888
2	-	0,0000	0,1667	0,2413	0,2806	0,3031	0,3164	0,3244
3	-	-	-	0,0000	0,0875	0,1401	0,1743	0,1976
4	-	-	-	-	-	0,0000	0,0561	0,0947
5	-	-	-	-	-	-	-	0,0000
6	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-
i	$n=10$	$n=11$	$n=12$	$n=13$	$n=14$	$n=15$	$n=16$	$n=17$
1	0,5739	0,5601	0,5475	0,5359	0,5251	0,5150	0,5056	0,4968
2	0,3291	0,3315	0,3325	0,3325	0,3318	0,3306	0,3290	0,3273
3	0,2141	0,2260	0,2347	0,2412	0,2460	0,2495	0,2521	0,2540
4	0,1224	0,1429	0,1586	0,1707	0,1802	0,1878	0,1939	0,1988
5	0,0399	0,0695	0,0922	0,1099	0,1240	0,1353	0,1447	0,1524
6	-	0,0000	0,0303	0,0539	0,7270	0,0880	0,1005	0,1109
7	-	-	-	0,0000	0,0240	0,0433	0,0593	0,0725
8	-	-	-	-	-	0,0000	0,0196	0,0359
9	-	-	-	-	-	-	-	0,0000
10	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-
i	$n=18$	$n=19$	$n=20$	$n=21$	$n=22$	$n=23$	$n=24$	$n=25$
1	0,4886	0,4808	0,4734	0,4643	0,4590	0,4542	0,0449	0,4450
2	0,3253	0,3232	0,3211	0,3185	0,3156	0,3126	0,3098	0,3069
3	0,2553	0,2561	0,2565	0,2578	0,2571	0,2563	0,2554	0,2543
4	0,2027	0,2059	0,2085	0,2119	0,2131	0,2139	0,2145	0,2148
5	0,1587	0,1641	0,1686	0,1736	0,1764	0,1787	0,1807	0,1822
6	0,1197	0,1271	0,1334	0,1399	0,1443	0,1480	0,1512	0,1539
7	0,0837	0,0932	0,1013	0,1092	0,1150	0,1201	0,1245	0,1283
8	0,0496	0,0612	0,0711	0,0804	0,0878	0,0941	0,0997	0,1046
9	0,0163	0,0303	0,0422	0,0530	0,0618	0,0696	0,0764	0,0823
10	-	0,0000	0,0140	0,0263	0,0368	0,0459	0,0539	0,0610
11	-	-	-	0,0000	0,0122	0,0228	0,0321	0,0403
12	-	-	-	-	-	0,0000	0,0107	0,0200

Tabuľka 49: Kritické hodnoty $W_\alpha(n)$ pre Shapiro–Wilkov test normality

n	$\alpha = 0,05$	$\alpha = 0,01$
3	0,767	0,753
4	0,748	0,687
5	0,762	0,686
6	0,788	0,713
7	0,803	0,730
8	0,818	0,749
9	0,829	0,764
10	0,842	0,781
11	0,850	0,792
12	0,859	0,805
13	0,866	0,814
14	0,874	0,825
15	0,881	0,835
16	0,887	0,844
17	0,892	0,851
18	0,897	0,858
19	0,901	0,863
20	0,905	0,868
21	0,908	0,873
22	0,911	0,878
23	0,914	0,881
24	0,916	0,884
25	0,918	0,888
26	0,920	0,891
27	0,923	0,894
28	0,924	0,896
29	0,926	0,898
30	0,927	0,900

Tabuľka 50: Kritické hodnoty $T_\alpha(n)$ pre Grubbsov test

n	$\alpha = 0,05$	$\alpha = 0,01$
3	1,142	1,414
4	1,689	1,723
5	1,187	1,955
6	1,996	2,130
7	2,093	2,265
8	2,172	2,374
9	2,238	2,464
10	2,294	2,540
11	2,340	2,606
12	2,387	2,663
13	2,426	2,713
14	2,461	2,759
15	2,494	2,800
16	2,523	2,837
17	2,551	2,871
18	2,577	2,903
19	2,601	2,932
20	2,623	2,959
21	2,644	2,984
22	2,664	3,008
23	2,683	3,030
24	2,701	3,051
25	2,718	3,071
26	2,734	3,089
27	2,749	3,107
28	2,764	3,124
29	2,778	3,140
30	2,792	3,156

Tabuľka 51: Kritické hodnoty $Q_\alpha(n)$ pre Dixonov test

n	$\alpha = 0,05$	$\alpha = 0,01$
3	0,941	0,988
4	0,765	0,889
5	0,642	0,780
6	0,560	0,698
7	0,507	0,637
8	0,468	0,590
9	0,437	0,555
10	0,412	0,527
11	0,392	0,502
12	0,376	0,482
13	0,361	0,465
14	0,349	0,450
15	0,338	0,438
16	0,329	0,426
17	0,320	0,416
18	0,313	0,407
19	0,306	0,398
20	0,300	0,391
21	0,295	0,384
22	0,290	0,378
23	0,285	0,321
24	0,281	0,367
25	0,277	0,362
26	0,273	0,357
27	0,269	0,353
28	0,266	0,349
29	0,263	0,345
30	0,260	0,341

Tabuľka 52: Kritické hodnoty $r_\alpha(n)$ pre Spearmanov test nezávislosti

n	$\alpha = 0,05$	$\alpha = 0,01$
5	0,9000	-
6	0,8286	0,9429
7	0,7450	0,8929
8	0,6905	0,8571
9	0,6833	0,8167
10	0,6364	0,7818
11	0,6091	0,7545
12	0,5804	0,7273
13	0,5549	0,6978
14	0,5341	0,6747
15	0,5179	0,6536
16	0,5000	0,6324
17	0,4853	0,6152
18	0,4716	0,5975
19	0,4579	0,5825
20	0,4451	0,5684
21	0,4351	0,5545
22	0,4241	0,5426
23	0,4150	0,5306
24	0,4061	0,5200
25	0,3977	0,5100
26	0,3894	0,5002
27	0,3822	0,4915
28	0,3749	0,4828
29	0,3685	0,4744
30	0,3620	0,4665

Názov: Aplikovaná štatistika

Autori: Anna Grinčová, Jana Petrillová

Vydavateľ: Technická univerzita v Košiciach
Rok: 2019
Vydanie: prvé
Náklad: 50 ks
Rozsah: 170 strán

ISBN 978-80-553-3213-0

ISBN 978-80-553-3213-0