Úlohy:

10.1. Vedenie podniku zamýšľa investovať do nákupu najmodernejších atrakcií s cieľom zvýšiť svoje zisky. K rozhodnutiu však pristúpi až po vypracovaní analýzy, ktorej cieľom je zistiť či zavedenie nových atrakcií priláka postačujúci počet návštevníkov. Z 12 ročnej histórie zábavného parku sú známe údaje o počte atrakcií (X) a zodpovedajúcom priemernom týždennom počte návštevníkov (Y) (v tisícoch). Výsledky sú v nasledujúcej tabuľke:

x_i	13	15	17	17	19	21	23	25	28	32	39	40
y_i	8	10	11	13	13	16	18	19	19	20	21	22

Predpokladajme, že veličiny X a Y spĺňajú predpoklady lineárneho regresného modelu. Riešte nasledujúce úlohy:

a) Odhadnite závislosť priemerného týždenného počtu návštevníkov od počtu atrakcií regresnou parabolou $f(x) = a_0 + a_1 x + a_2 x^2$.
b) Regresnú parabolu spolu s bodmi (x_i, y_i) znázornite.
c) Určte SSE, $RMSE$ a výberové smerodajné odchýlky odhadnutých regresných koeficientov.
d) Určte 95 %-né intervaly spoľahlivosti pre regresné koeficienty.
e) Na hladine významnosti $\alpha = 0{,}05$ testujte významnosť regresných koeficientov.
f) Určte bodový odhad funkčnej hodnoty $y(41)$ a 95 %-ný interval spoľahlivosti pre túto funkčnú hodnotu.
g) Charakterizujte vhodnosť modelu pomocou výberového koeficienta determinácie.
h) Na hladine významnosti $\alpha = 0{,}05$ testujte významnosť regresného modelu ako celku.
i) Overte predpoklady uvedeného lineárneho regresného modelu. Postupujte ako v riešenom príklade 10.3. Najprv vykonajte grafickú analýzu.

<u>Výsledok</u>: a) $\hat{y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2 = -10{,}9194 + 1{,}7603x - 0{,}0239x^2$;

c) $SSE = 6{,}9434$; $RMSE = 0{,}8783$; $s(\hat{a}_0) = 2{,}5798$; $s(\hat{a}_1) = 0{,}2098$; $s(\hat{a}_2) = 0{,}0039$;

d) $a_0 \in \langle -16{,}7552\,; -5{,}0836 \rangle$, $a_1 \in \langle 1{,}2857; 2{,}2349 \rangle$, $a_2 \in \langle -0{,}0326\,; -0{,}0151 \rangle$;

e) `regstats(y,x,'quadratic')`

t = [-4.2327 ; 8.3902 ; -6.1701] , pval = [0.0022 ; 1.5100e-05;1.6467e-04], pretože p-hodnota je vo všetkých prípadoch menšia ako $\alpha = 0{,}05$, tak regresné koeficienty považujeme za štatisticky významné. Ten istý výsledok vyplýva aj z IS, keďže žiaden z nich neobsahuje 0.

f) $\hat{y}(41) = 21{,}0959$; $y(41) \in \langle 19{,}4629\,; 22{,}7289 \rangle$; g) $R^2 = 0{,}9713$; $R^{*2} = 0{,}9649$;

h) `fitlm(x,y,'quadratic')`

F-statistic: 152, p-value = 1.16e-07; $p < 0{,}05$, prijímame hypotézu H_1 o štatistickej významnosti výberového koeficienta determinácie R^2.

i) Klasické a štandardizované rezíduá:

e = 0.0727 -0.1101 -1.1018 0.8982 -0.9025 0.4880
1.0696 0.8424 -0.6402 -0.9481 -0.3975 0.7294

eS = 0.1117 -0.1438 -1.3553 1.1047 -1.0957 0.5989
1.3395 1.0758 -0.8294 -1.2169 -0.5921 1.2080

V grafe $(\hat{y}_i, e_{Si})$ sú všetky rezíduá náhodne rozmiestnené okolo nuly v intervale $(-2\,, 2)$.

V grafe (i, e_i) spolu s 95 %-nými intervalmi spoľahlivosti pre náhodné chyby vidíme, že IS všetkých náhodných chýb obsahujú 0, čo svedčí v prospech hypotézy, že náhodné chyby sú normálne rozdelené s nulovou strednou hodnotou.

Aplikácia A-D testu na overenie normality náhodných chýb na hladine významnosti $\alpha = 0{,}05$: $p = 0{,}3305 > 0{,}05$, prijímame hypotézu H_0 o normálnom rozdelení náhodných chýb.

Test nulovej strednej hodnoty náhodných chýb pomocou *t*-testu na hladine významnosti $\alpha = 0{,}05$: $p = 1 > 0{,}05$, prijímame hypotézu H_0 o nulovej strednej hodnote náhodných chýb.

Test homoskedasticity na hladine významnosti $\alpha = 0{,}05$: $SSE_d = 0{,}5512$; $SSE_h = 2{,}708$;

$H_0: \sigma_d^2 = \sigma_h^2$, $H_1: \sigma_h^2 > \sigma_d^2$; F = 2,708/0,5512 = 4,9129; $F < F_{0,95;\,4;\,4} = 6{,}3882$; prijímame hypotézu o homoskedasticite náhodných chýb.

Test nekorelovanosti náhodných chýb pomocou *DW* štatistiky na hladine významnosti $\alpha = 0{,}05$: dw = 2.0809, pval = 0.3985; $p > \alpha$, prijímame hypotézu H_0 o nekorelovanosti náhodných chýb.

Test náhodnosti chýb na základe bodov zvratu na hladine významnosti $\alpha = 0{,}05$:
BodZ = -1.1018 0.8982 -0.9025 1.0696 -0.9481; Z = 5, q = 1.96, U = -1.2384, prijímame hypotézu H_0 o náhodnosti chýb.

Poznámka: Riešte všetky zadané úlohy aj pre prípad regresnej hyperboly $f(x) = a_0 + \frac{a_1}{x}$.

Niektoré výsledky:
General model: f(x) = a0+a1./x
Coefficients (with 95% confidence bounds):
 a0 = 28.98 (27.14, 30.81)
 a1 = -280.3 (-317.4, -243.3)
Goodness of fit: SSE: 8.203, R-square: 0.9661, Adjusted R-square: 0.9627, RMSE: 0.9057

10.2. Tabuľka uvádza vek strojov (*X*) v rokoch a náklady na ich údržbu (*Y*) v eurách:

x_i	14	0,8	3	7,5	8,4	14,8	4,5	15,6	17,3	11,5	13,2	1,5
y_i	47,5	8	10	17	22	76,4	12,5	76	94,5	25	30,6	12

Riešte nasledujúce úlohy:

a) Odhadnite závislosť nákladov od veku stroja regresnou exponenciálnou funkciou $f(x) = a_0 \cdot a_1^x$.
b) Regresnú exponenciálnu krivku spolu s bodmi (x_i, y_i) znázornite.
c) Vypočítajte výberový reziduálny rozptyl a výberovú reziduálnu smerodajnú odchýlku.
d) Posúďte súlad použitého modelu so skutočnosťou na základe indexu determinácie, resp. modifikovaného indexu determinácie.

Výsledok: Riešenie pomocou logaritmickej transformácie: a) $\hat{y} = \hat{a}_0 \cdot \hat{a}_1^x = 6{,}9473 \cdot 1{,}1511^x$; c) $MSE = 116{,}609$; $RMSE = 10{,}7986$; d) $R^2 = 0{,}8843$; $R^{*2} = 0{,}8728$; modelom exponenciálnej regresie je vysvetlených približne 87 % variability znaku *Y*.