

Technická univerzita v Košiciach  
Fakulta elektrotechniky a informatiky



## Aplikovaná štatistika

10. cvičenie - riešené príklady v R

Jana Petrillová, Mária Švecová

```

# PRÍKLAD 7.1

x=c(160,250,320,500,750,1000,1500,2000)
y=c(789,800,851,874,1193,1335,1704,2073)
alpha = 0.05
n=length(x)

# a) -----
# lineárny model (balík stats, zaleží na poradí y~x alebo x ~ y)

model = lm(y~x)
summary(model)
b=model$coefficients          # bodové odhady parametrov regresného modelu
fy=model$fitted.values       # odhadované hodnoty pomocou regresného modelu (alebo fy=predict.lm(model))

# b) -----
# scatter.smooth(x,y) zakreslí body x,y a preloží to zhruba ako by mohla vyzerat závislosť medzi nimi

plot(x,y, xlab = "", ylab = "", type="p", col="red", pch=19, cex.axis=1.3)
title(xlab = list("Počet stromov v sade", cex = 1.4), ylab = list("Množstvo úrody jabĺk", cex = 1.4))
abline(model,col="blue",lty=2,lwd=2)          # alebo abline(reg = model) alebo abline(coef = coef(model))

# c) -----
# dostanem z modelu v a) ... R^2=Multiple R-squared
# alebo

s1=sum((y-fy)^2)
s2=sum((y-mean(y))^2)
RR=1-(s1/s2)                                # výberový koeficient determinácie

# d) -----
yy=mean(y)
SSE=sum((y-fy)^2)
m=2
TS=((n-m)*sum((fy-yy)^2))/((m-1)*SSE)        # testovacie kritérium
KH1=qf(1-alpha,m-1,n-m)                    # kritická hodnota

cat("Kritická oblasť (",KH1,", Inf)")

# vyhodnotenie pomocou kritickej oblasti
if(TS > KH1) cat("TS patri do kritickej oblasti, H0 zamietame prijimame H1") else
  cat("TS nepatri do kritickej oblasti, H0 nezamietame")

# alebo dostanem z modelu v a) ... F-statistic, p-value
p=2.825e-07

# vyhodnotenie pomocou p-hodnoty
if(p > alpha)
  cat("p > alpha, H0 nezamietame na hladine alpha") else
  cat("p <= alpha, H0 zamietame prijimame H1")

# e) -----
# pre beta0
xx=mean(x)
sb0=(SSE/(n-m))*(sum(x^2)/(n*sum((x-xx)^2)))
c=0
TS=(b[1]-c)/sqrt(sb0)                       # testovacie kritérium
KH2=qt(1-alpha/2, n-2)                      # kritická hodnota

cat("Kritická oblasť (-Inf, ",-KH2,") zjednotenie (",KH2,", Inf)")

# vyhodnotenie pomocou kritickej oblasti
if(TS > KH1) cat("TS patri do kritickej oblasti, H0 zamietame prijimame H1") else
  if (TS < -KH1) cat("TS patri do kritickej oblasti, H0 zamietame prijimame H1") else
    cat("TS nepatri do kritickej oblasti, H0 nezamietame")

# pre beta1
sb1=(SSE/(n-m))*(1/sum((x-xx)^2))
c=0
TS=(b[2]-c)/sqrt(sb1)                       # testovacie kritérium
KH2=qt(1-alpha/2, n-2)                      # kritická hodnota

cat("Kritická oblasť (-Inf, ",-KH2,") zjednotenie (",KH2,", Inf)")

# vyhodnotenie pomocou kritickej oblasti
if(TS > KH1) cat("TS patri do kritickej oblasti, H0 zamietame prijimame H1") else
  if (TS < -KH1) cat("TS patri do kritickej oblasti, H0 zamietame prijimame H1") else
    cat("TS nepatri do kritickej oblasti, H0 nezamietame")

```

```

# alebo dostanem z modelu v a) ... t-value v tabuľke Coefficients, vedľa stĺpec Pr - p-value,
# prvý riadok je pre koeficient b0 a druhý pre b1
# pre b0
pp=8.13e-07

# vyhodnotenie pomocou p-hodnoty
if(pp > alpha)
  cat("p > alpha, H0 nezamietame na hladine alpha") else
  cat("p <= alpha, H0 zamietame prijimame H1")

# pre b1
ppp=2.83e-07

# vyhodnotenie pomocou p-hodnoty
if(ppp > alpha)
  cat("p > alpha, H0 nezamietame na hladine alpha") else
  cat("p <= alpha, H0 zamietame prijimame H1")

# f) -----
# pre beta0
q=qt(1-alpha/2, n-2)
D0 = b[1] - q*sqrt(sb0)
H0 = b[1] + q*sqrt(sb0)
cat("<",D0,"",H0,">")          # interval spolahlivosti

# pre beta1
D1 = b[2] - q*sqrt(sb1)
H1 = b[2] + q*sqrt(sb1)
cat("<",D1,"",H1,">")          # interval spolahlivosti

# alebo
cbind(bod.odhad=coef(model),confint(model))

# v prípade iného ako 95% IS - napr. pre 99% IS použijeme cbind(bod.odhad=coef(model),confint(model, level=0.99))

# -----
# MSE, SSE, RMSE
RMSE=50.61      # Residual standard error z modelu
MSE=RMSE^2
SSE=MSE*(n-2)

```

# PRÍKLAD 7.2

```
x=c(160,250,320,500,750,1000,1500,2000)
y=c(789,800,851,874,1193,1335,1704,2073)
alpha = 0.05
n=length(x)
```

```
# a) -----
# exponencialny regresny model (balík stats, záleží na poradí y~x alebo x ~ y)

model = glm(log(y) ~ x)
summary(model)
b=model$coefficients # bodové odhady parametrov regresného modelu
summary(model)
fz=model$fitted.values # odhadované hodnoty pomocou regresného modelu
b0=exp(b[1]) # bodový odhad parametra beta0
b1=b[2] # bodový odhad parametra beta1
fy=exp(fz) # odhadované hodnoty upravené na exponenciálny model

# b) -----
# scatter.smooth(x,y) zakreslí body x,y a preloží to zhruba ako by mohla vyzerat závislosť medzi nimi

plot(x,y, xlab = "", ylab = "", type="p", col="red", pch=19, cex.axis=1.3)
title(xlab = list("Počet stromov v sade", cex = 1.4), ylab = list("Množstvo úrody jabák", cex = 1.4))
m=nls(y~a*exp(b*x),start=list(a=b0,b=b1) # aby pekne vykreslilo funkciu hodnoty x musia byť zoradené
# do variačného radu

lines(x,predict(m),col="blue",lty=2,lwd=2) # výsledný graf pre exponenciálny regresny model

# c) -----

s1=sum((y-fy)^2)
s2=sum((y-mean(y))^2)
RR=1-(s1/s2) # vyberový koeficient determinácie
```