2. SzRM gyakorlat (szept 26, kedd 2-4, 4-6; szept 28, csütörtök 4-6)

A feladatok megoldásai R-ben

```
# r prog a lin.ill-re
rm(list=ls())

y<-c(23,20,14,18,19,21)
x<-1:6
lm(y~x)
summary(lm(y~x))
anova(lm(y~x))


# --------------
# Call:
# lm(formula = y ~ x)
#
# Coefficients:
# (Intercept)            x
#     20.0667      -0.2571
# --------------

# --------------
# Call:
# lm(formula = y ~ x)
#
# Residuals:
#     1       2       3       4       5       6
#  3.1905  0.4476 -5.2952 -1.0381  0.2190  2.4762
#
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  20.0667     3.1459   6.379   0.0031 **
# x            -0.2571     0.8078  -0.318   0.7662
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 3.379 on 4 degrees of freedom
# Multiple R-squared: 0.02471,    Adjusted R-squared: -0.2191
# F-statistic: 0.1013 on 1 and 4 DF,  p-value: 0.7662
# ----------------------------

# ----------------------------
# Analysis of Variance Table
#
# Response: y
#           Df Sum Sq Mean Sq F value Pr(>F)
# x          1  1.157   1.157  0.1013 0.7662
# Residuals  4 45.676  11.419
# ----------------------------

# ============================
# b1       b0       -0,2571 20,0667
# se.b1    se.b0     0,8078  3,1459
# r2       se        0,02471 3,379
# F        errorDf   0,1013  4
# SSmodell SSerror   1,1571 45,676
# ============================

b1<-sum((x-mean(x))*(y-mean(y)))/sum((x-mean(x))^2)
b1 # -0,2571
b0<-mean(y)-b1*mean(x)
b0 # 20,0667

errorDf<-6-2
errorDf # 4
modellDf<-2-1

ysapka<-b0+b1*x

se<-sqrt(sum((ysapka-y)^2)/errorDf)
se # 3,379

se.b1<-se*sqrt(1/sum((x-mean(x))^2))
se.b1 # 0,8078

se.b0<-se*sqrt(sum(x^2)/sum(6*(x-mean(x))^2))
se.b0 # 3,1459

r<-sum((x-mean(x))*(y-mean(y)))/sqrt(sum((x-mean(x))^2)*sum((y-mean(y))^2))
(r2<-r^2) # 0,02471

SSmodell<-sum((ysapka-mean(ysapka))^2)
SSmodell # 1,1571

SSerror<-sum((ysapka-y)^2)
SSerror # 45,676

F<-(SSmodell/modellDf)/(SSerror/errorDf)
F # 0,1013

# ~~~~~~~~~~~~~~~~~
n<-length(x)
st<-sum((y-mean(y))^2)/n
sm<-sum((y-ysapka)^2)/n
sr<-sum((mean(ysapka)-ysapka)^2)/n
(st-sm)/st # 0,02471
sr/st # 0,02471   modell 'szoras'/ mfi 'szoras'
# -----
# fine
```

```
# ========================================
# elokeszites elemi fuggvenyekkel

rm(list=ls())
SS<-function(u,v=NULL) if(is.null(v)) sum((u-mean(u))^2) else sum((u-
mean(u))*(v-mean(v)))
y<-c(23,20,14,18,19,21)
x<-1:6
n<-length(x)
p<-2
b1<-SS(x,y)/SS(x)
b0<-mean(y)-mean(x)*b1

y.modell<-b0+b1*x #  y modell

SS.modell<-SS(y.modell)    # modell SS
SS.hiba<-SS(y-y.modell)   # hiba SS
SS.total<-SS(y)          # total SS

MS.modell<-SS(y.modell)/(p-1)    # modell MS
MS.hiba<-SS(y-y.modell)/(n-p)   # hiba MS
MS.total<-SS(y)/(n-1)          # total MS

df.modell<-p-1
df.hiba<-n-p
df.total<-n-1

F.stat<-(SS.modell/df.modell)/(SS.hiba/df.hiba)
F.stat.p.erteke<-1-pf(F.stat,df.modell,df.hiba)

r<-SS(x,y)/sqrt(SS(x)*SS(y))
r2<-r^2
r2corr<-1-(1-r2)*(n-1)/(n-2)
se.modell<-sqrt(SS.hiba/(n-p))

se.b0<-sqrt(sum(x^2)/(n*SS(x)))*se.modell
se.b1<-1/sqrt(SS(x))*se.modell

t.b0<-b0/se.b0
t.b1<-b1/se.b1

p.b0<-2*pt(-abs(t.b0),n-p)
p.b1<-2*pt(-abs(t.b1),n-p)

alpha.05<-.05
kv.05<-qt(1-alpha.05/2,n-p)
alpha.03<-.03
kv.03<-qt(1-alpha.03/2,n-p)

a5.b0<-b0-kv.05*se.b0
f5.b0<-b0+kv.05*se.b0
a3.b0<-b0-kv.03*se.b0
f3.b0<-b0+kv.03*se.b0

a5.b1<-b1-kv.05*se.b1
f5.b1<-b1+kv.05*se.b1
a3.b1<-b1-kv.03*se.b1
f3.b1<-b1+kv.03*se.b1

# ---
# ugyanez matrixosan
X<-cbind(1,x)
Y<-matrix(y)
b<-solve(t(X)%*%X)%*%t(X)%*%Y
b-c(b0,b1)

y.sapka<-X%*%b
y.sapka-y.modell

se<-sqrt(SS(y-y.sapka)/(n-p))
se.modell-se

sigma.b<-sqrt(diag(solve(t(X)%*%X)))*se
c(se.b0,se.b1)-sigma.b

# ========================================
# EXCEL eredmenyek reprodukcioja

# ---
# Regresszios statisztika
RS<-data.frame(c(r,r2,r2corr,se.modell,n))
colnames(RS)<-" "
rownames(RS)<-c("r értéke","r-négyzet","Korrigált r-négyzet",
                "Standard hiba","Megfigyelések")
RS

# ---
# A variancia analízis tabla
VA<-data.frame(c(df.modell,df.hiba,df.total),
               c(SS.modell,SS.hiba,SS.total),
               c(MS.modell,MS.hiba,NA),
               c(F.stat,NA,NA),
               c(F.stat.p.erteke,NA,NA)     )
colnames(VA)<-c("df","SS","MS","F","F szignifikanciája")
rownames(VA)<-c("Regresszió","Maradék","Összesen")
VA

# ---
# Az együtthatók elemzése
EE<-matrix(c(b0,se.b0,t.b0,p.b0,a5.b0,f5.b0,a3.b0,f3.b0,
             b1,se.b1,t.b1,p.b1,a5.b1,f5.b1,a3.b1,f3.b1),
           2,byrow=TRUE )

colnames(EE)<-c("Koefficiensek","Standard hiba","t érték","p-érték",
                "Alsó 95%","Felső 95%","Alsó 97%","Felső 97%")
rownames(EE)<-c("Tengelymetszet","X változó 1")
EE

----
fine
```

```
#
# Y          X1        X2                  X3                  X4             X5              X6
# halálozás, csapadék, januári hőmérséklet, júliusi hőmérséklet, öregek aránya, háztartás méret, iskolázottság,
# X7                 X8          X9                X10                X11                X12 X13  X14  X15
# összkomfort aránya, népsűrűség, afroamerikai arány, nem fizikai munkás, szegények  aránya, CO2, NO, SO2, pára
#

D<-matrix(c(
 921.870, 36, 27, 71,  8.1, 3.34, 11.4, 81.5, 3243,  8.8, 42.6, 11.7, 21,   15,  59, 59,
 997.875, 35, 23, 72, 11.1, 3.14, 11.0, 78.8, 4281,  3.6, 50.7, 14.4,  8,   10,  39, 57,
 962.354, 44, 29, 74, 10.4, 3.21,  9.8, 81.6, 4260,  0.8, 39.4, 12.4,  6,    6,  33, 54,
 982.291, 47, 45, 79,  6.5, 3.41, 11.1, 77.5, 3125, 27.1, 50.2, 20.6, 18,    8,  24, 56,
1071.289, 43, 35, 77,  7.6, 3.44,  9.6, 84.6, 6441, 24.4, 43.7, 14.3, 43,   38, 206, 55,
1030.380, 53, 45, 80,  7.7, 3.45, 10.2, 66.8, 3325, 38.5, 43.1, 25.5, 30,   32,  72, 54,
 934.700, 43, 30, 74, 10.9, 3.23, 12.1, 83.9, 4679,  3.5, 49.2, 11.3, 21,   32,  62, 56,
 899.529, 45, 30, 73,  9.3, 3.29, 10.6, 86.0, 2140,  5.3, 40.4, 10.5,  6,    4,   4, 56,
1001.902, 36, 24, 70,  9.0, 3.31, 10.5, 83.2, 6582,  8.1, 42.5, 12.6, 18,   12,  37, 61,
 912.347, 36, 27, 72,  9.5, 3.36, 10.7, 79.3, 4213,  6.7, 41.0, 13.2, 12,    7,  20, 59,
1017.613, 52, 42, 79,  7.7, 3.39,  9.6, 69.2, 2302, 22.2, 41.3, 24.2, 18,    8,  27, 56,
1024.885, 33, 26, 76,  8.6, 3.20, 10.9, 83.4, 6122, 16.3, 44.9, 10.7, 88,   63, 278, 58,
 970.467, 40, 34, 77,  9.2, 3.21, 10.2, 77.0, 4101, 13.0, 45.7, 15.1, 26,   26, 146, 57,
 985.950, 35, 28, 71,  8.8, 3.29, 11.1, 86.3, 3042, 14.7, 44.6, 11.4, 31,   21,  64, 60,
 958.839, 37, 31, 75,  8.0, 3.26, 11.9, 78.4, 4259, 13.1, 49.6, 13.9, 23,    9,  15, 58,
 860.101, 35, 46, 85,  7.1, 3.22, 11.8, 79.9, 1441, 14.8, 51.2, 16.1,  1,    1,   1, 54,
 936.234, 36, 30, 75,  7.5, 3.35, 11.4, 81.9, 4029, 12.4, 44.0, 12.0,  6,    4,  16, 58,
 871.766, 15, 30, 73,  8.2, 3.15, 12.2, 84.2, 4824,  4.7, 53.1, 12.7, 17,    8,  28, 38,
 959.221, 31, 27, 74,  7.2, 3.44, 10.8, 87.0, 4834, 15.8, 43.5, 13.6, 52,   35, 124, 59,
 941.181, 30, 24, 72,  6.5, 3.53, 10.8, 79.5, 3694, 13.1, 33.8, 12.4, 11,    4,  11, 61,
 891.708, 31, 45, 85,  7.3, 3.22, 11.4, 80.7, 1844, 11.5, 48.1, 18.5,  1,    1,   1, 53,
 871.338, 31, 24, 72,  9.0, 3.37, 10.9, 82.8, 3226,  5.1, 45.2, 12.3,  5,    3,  10, 61,
 971.122, 42, 40, 77,  6.1, 3.45, 10.4, 71.8, 2269, 22.7, 41.4, 19.5,  8,    3,   5, 53,
 887.466, 43, 27, 72,  9.0, 3.25, 11.5, 87.1, 2909,  7.2, 51.6,  9.5,  7,    3,  10, 56,
 952.529, 46, 55, 84,  5.6, 3.35, 11.4, 79.7, 2647, 21.0, 46.9, 17.9,  6,    5,   1, 59,
 968.665, 39, 29, 76,  8.7, 3.23, 11.4, 78.6, 4412, 15.6, 46.6, 13.2,  13,   7,  33, 60,
 919.729, 35, 31, 81,  9.2, 3.10, 12.0, 78.3, 3262, 12.6, 48.6, 13.9,   7,   4,   4, 55,
 844.053, 43, 32, 74, 10.1, 3.38,  9.5, 79.2, 3214,  2.9, 43.7, 12.0,  11,   7,  32, 54,
 861.833, 11, 53, 68,  9.2, 2.99, 12.1, 90.6, 4700,  7.8, 48.9, 12.3, 648, 319, 130, 47,
 989.265, 30, 35, 71,  8.3, 3.37,  9.9, 77.4, 4474, 13.1, 42.6, 17.7,  38,  37, 193, 57,
1006.490, 50, 42, 82,  7.3, 3.49, 10.4, 72.5, 3497, 36.7, 43.3, 26.4,  15,  10,  34, 59,
 861.439, 60, 67, 82, 10.0, 2.98, 11.5, 88.6, 4657, 13.6, 47.3, 22.4,   3,   1,   1, 60,
 929.150, 30, 20, 69,  8.8, 3.26, 11.1, 85.4, 2934,  5.8, 44.0,  9.4,  33,  23, 125, 64,
 857.622, 25, 12, 73,  9.2, 3.28, 12.1, 83.1, 2095,  2.0, 51.9,  9.8,  20,  11,  26, 50,
 961.009, 45, 40, 80,  8.3, 3.32, 10.1, 70.3, 2682, 21.0, 46.1, 24.1,  17,  14,  78, 56,
 923.234, 46, 30, 72, 10.2, 3.16, 11.3, 83.2, 3327,  8.8, 45.3, 12.2,   4,   3,   8, 58,
1113.156, 54, 54, 81,  7.4, 3.36,  9.7, 72.8, 3172, 31.4, 45.5, 24.2,  20,  17,   1, 62,
 994.648, 42, 33, 77,  9.7, 3.03, 10.7, 83.5, 7462, 11.3, 48.7, 12.4,  41,  26, 108, 58,
1015.023, 42, 32, 76,  9.1, 3.32, 10.5, 87.5, 6092, 17.5, 45.3, 13.2,  29,  32, 161, 54,
 991.290, 36, 29, 72,  9.5, 3.32, 10.6, 77.6, 3437,  8.1, 45.5, 13.8,  45,  59, 263, 56,
 893.991, 37, 38, 67, 11.3, 2.99, 12.0, 81.5, 3387,  3.6, 50.3, 13.5,  56,  21,  44, 73,
 938.500, 42, 29, 72, 10.7, 3.19, 10.1, 79.5, 3508,  2.2, 38.3, 15.7,   6,   4,  18, 56,
 946.185, 41, 33, 77, 11.2, 3.08,  9.6, 79.9, 4843,  2.7, 38.6, 14.1,  11,  11,  89, 54,
1025.502, 44, 39, 78,  8.2, 3.32, 11.0, 79.9, 3768, 28.6, 49.5, 17.5,  12,   9,  48, 53,
 874.281, 32, 25, 72, 10.9, 3.21, 11.1, 82.5, 4355,  5.0, 46.4, 10.8,   7,   4,  18, 60,
 953.560, 34, 32, 79,  9.3, 3.23,  9.7, 76.8, 5160, 17.2, 45.1, 15.3,  31,  15,  68, 57,
 839.709, 10, 55, 70,  7.3, 3.11, 12.1, 88.9, 3033,  5.9, 51.0, 14.0, 144,  66,  20, 61,
 911.701, 18, 48, 63,  9.2, 2.92, 12.2, 87.7, 4253, 13.7, 51.2, 12.0, 311, 171,  86, 71,
 790.733, 13, 49, 68,  7.0, 3.36, 12.2, 90.7, 2702,  3.0, 51.9,  9.7, 105,  32,   3, 71,
 899.264, 35, 40, 64,  9.6, 3.02, 12.2, 82.5, 3626,  5.7, 54.3, 10.1,  20,   7,  20, 72,
 904.155, 45, 28, 74, 10.6, 3.21, 11.1, 82.6, 1883,  3.4, 41.9, 12.3,   5,   4,  20, 56,
 950.672, 38, 24, 72,  9.8, 3.34, 11.4, 78.0, 4923,  3.8, 50.5, 11.1,   8,   5,  25, 61,
 972.464, 31, 26, 73,  9.3, 3.22, 10.7, 81.3, 3249,  9.5, 43.9, 13.6,  11,   7,  25, 59,
 912.202, 40, 23, 71, 11.3, 3.28, 10.3, 73.8, 1671,  2.5, 47.4, 13.5,   5,   2,  11, 60,
 967.803, 41, 37, 78,  6.2, 3.25, 12.3, 89.5, 5308, 25.9, 59.7, 10.3,  65,  28, 102, 52,
 823.764, 28, 32, 81,  7.0, 3.27, 12.1, 81.0, 3665,  7.5, 51.6, 13.2,   4,   2,   1, 54,
1003.502, 45, 33, 76,  7.7, 3.39, 11.3, 82.2, 3152, 12.1, 47.3, 10.9,  14,  11,  42, 56,
 895.696, 45, 24, 70, 11.8, 3.25, 11.1, 79.8, 3678,  1.0, 44.8, 14.0,   7,   3,   8, 56,
 911.817, 42, 83, 76,  9.7, 3.22,  9.0, 76.2, 9699,  4.8, 42.2, 14.5,   8,   8,  49, 54,
 954.442, 38, 28, 72,  8.9, 3.48, 10.7, 79.8, 3451, 11.7, 37.5, 13.0,  14,  13,  39, 58),
ncol=16,byrow=TRUE)

colnames(D)<-c("Y", "X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10", "X11", "X12", "X13", "X14", "X15")
dim(D) # 60 16
D<-data.frame(D)
```

```
# 1.modell
M1<-lm(Y~X11,data=D)
M1
summary(M1)# OK

M2<-lm(Y~X12+X13+X14+X15,data=D)
M2
summary(M2)# X14 es X15 gyenge


M2mod<-lm(Y~X12+X13,data=D)
M2mod
summary(M2mod)# OK


M3<-lm(Y~X4+X9+X11,data=D)
M3
summary(M3)# X11 gyenge

M3mod<-lm(Y~X4+X9,data=D)
M3mod
summary(M3mod)# OK

M4<-lm(Y~X1+X2+X3,data=D)
M4
summary(M4)# X2 X3 gyenge

M4mod<-lm(Y~X1,data=D)
M4mod
summary(M4mod)# OK

Mossz<-lm(Y~X1+X4+X9+X12+X13,data=D)
Mossz
summary(Mossz)# X1 X4 gyenge

Mossz2<-lm(Y~X9+X12+X13,data=D)
Mossz2
summary(Mossz2)# OK

Mtotal<-lm(Y~.,data=D)
Mtotal
summary(Mtotal)# sok gyenge

Mtotal2<-lm(Y~X1+X2+X6+X8+X9,data=D)
Mtotal2
summary(Mtotal2)# OK


Maut<-step(Mtotal)
Maut
summary(Maut)# OK


# ---
# "jo" modellek

#           éghajlat     | gazdasag-szocio                    | kor | lég
#
#             X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 + X13 + X14 + X15 #
#
# M1     Y ~                                                    X11                          # oreg .168
# M2     Y ~                                                          X12 + X13              # leg .310
# M3     Y ~                 X4 +                     X9                                     # gazd .508
# M4     Y ~ X1                                                                              # égh .260
# Mossz  Y ~                                          X9             + X12 + X13             # egyesített .607
# Mtotal Y ~ X1 + X2 +                X6      + X8 + X9                                       # az egeszből .721
#
# Maout  Y ~ X1 + X2 + X3 +?X4 +?X5 + X6 +      X8 + X9 +              X12 + X13             # lepesenkenti .792

# teljes ------------------------------------------------------------------------------------> .7985
# Y   halálozás
# X1  csapadék
# X2  januári hőmérséklet
# X3  júliusi hőmérséklet
# X4  öregek aránya
# X5  háztartás méret
# X6  iskolázottság
# X7  összkomfort aránya
# X8  népsűrűség
# X9  afroamerikai arány
# X10 nem fizikai munkás
# X11 szegények  aránya
# X12 CO2
```

```
# X13 NO
# X14 SO2
# X15 pára


summary(M1)# OK
summary(M2mod)# OK
summary(M3mod)# OK
summary(M4mod)# OK
summary(Mossz2)# OK
summary(Mtotal2)# OK
summary(Maut)# OK

c(formula(M1),formula(M2mod),formula(M3mod),formula(M4mod),formula(Mossz2),formula(Mtotal2),formula(Maut))

summary(M3mod)$r.sq,

c(summary(M1)$r.sq,summary(M2mod)$r.sq,summary(M3mod)$r.sq,summary(M4mod)$r.sq,summary(Mossz2)$r.sq,summary(Mtotal2)$r.sq,summary(Maut)$r.sq)

# ----
# fine
```