Linearis regresszio

Kekecs Zoltan

1 november 2019

A linearis regresszio alapjai

Ennek az oranak a celja hogy megismerkedjunk a linearis regresszioval, annak logikajaval, es az ertelmezesehez szukseges alapfogalmakkal.

A dokumentum legfrissebb valotzata itt talalhato: https://osf.io/rx4f5/

Package-ek betoltese

Betoltjuk a kovatkazo package-eket:

```
library(psych)  # for describe
library(dplyr)  # for data management
library(gsheet)  # to read data from google sheets
library(ggplot2)  # for ggplot
library(tidyverse)  # for tidy data
```

Sajat funckiok betoltese

Az alabbi funkcio csak az orai vizualizaciohoz kell, nem feltetlenul kell megerteni a tartalmat. Ezt a funkciot arra hasznaljuk majd hogy a linearis regresszioban fennmarado rezidualis hibat vizualizaljuk.

```
error_plotter <- function(mod, col = "black", x_var = NULL) {
    mod_vars = as.character(mod$call[2])
    data = as.data.frame(eval(parse(text = as.character(mod$call[3]))))
    y = substr(mod_vars, 1, as.numeric(gregexpr(pattern = "~",
        mod_vars)) - 2)
    x = substr(mod_vars, as.numeric(gregexpr(pattern = "~", mod_vars)) +
        2, nchar(mod_vars))

    data$pred = predict(mod)

    if (x == "1" & is.null(x_var)) {
        x = "response_ID"
        data$response_ID = 1:nrow(data)
    } else if (x == "1") {
        x = x_var
    }

    plot(data[, y] ~ data[, x], ylab = y, xlab = x)
    abline(mod)

    for (i in 1:nrow(data)) {
        clip(min(data[, x]), max(data[, x]), min(data[i, c(y,
            "pred")]), max(data[i, c(y, "pred")]))
        abline(v = data[i, x], lty = 2, col = col)
    }
```

```
}
```

Adatmenedzsment es adat bemutatasa

Adatok betoltese

Mondjuk, hogy egy turistak koreben gyakran latogatott cipoboltban dolgozunk, es mivel a vilagon sokfajta cipomeretet hasznalnak es az emberek gyakran nem tudjak a sajat europai cipomeretuket, szeretnenk a magassaguk alapjan megbecsulni, mekkora az europai cipomeretuk.

Az alabbi koddal betolthetjuk az adattablat, amiben a korabbi orakon felvett kerdoivekbol szerepelnek a magassag es cipomeret adatok.

```
mydata = as_tibble(gsheet2tbl("https://docs.google.com/spreadsheets/d/1GXx2YoktyIdXLqKdm4f_MMWHUzXYgxWR
```

Adatok ellenorzese

Szokas szerint az adatok ellenorzesevel kezdunk, pl. View(), describe(), es summary() funkciokkal.

```
# descriptive statistics
describe(mydata)
```

```
## Warning in describe(mydata): NAs introduced by coercion
```

```
##                   vars  n   mean    sd median trimmed  mad   min max
## jelige*              1 25  32.00    NA     32   32.00 0.00 32.00  32
## magassag             2 25 161.07 34.32    168  166.62 8.90  1.82 190
## cipomeret            3 25  38.94  3.34     38   38.60 2.97 35.00  47
## alvas_tegnap_1       4 25   7.92  1.26      8    7.90 1.48  6.00  10
## alvas_tegnap_3       5 23   7.61  1.03      8    7.68 1.48  5.00   9
## alvas_altalaban_1    6 25   7.04  1.02      7    7.10 1.48  5.00   9
## alvas_altalaban_3    7 23   7.13  0.63      7    7.16 0.00  6.00   8
## energiaszint_1       8 25   5.48  2.16      5    5.62 2.97  1.00   8
## energiaszint_3       9 23   6.17  1.92      7    6.26 1.48  3.00   9
##                    range  skew kurtosis   se
## jelige*             0.00    NA       NA   NA
## magassag          188.18 -3.95    15.59 6.86
## cipomeret          12.00  0.99    -0.05 0.67
## alvas_tegnap_1      4.00 -0.22    -1.12 0.25
## alvas_tegnap_3      4.00 -0.63    -0.11 0.22
## alvas_altalaban_1   4.00 -0.30    -0.71 0.20
## alvas_altalaban_3   2.00 -0.07    -0.63 0.13
## energiaszint_1      7.00 -0.27    -1.21 0.43
## energiaszint_3      6.00 -0.38    -1.35 0.40
```

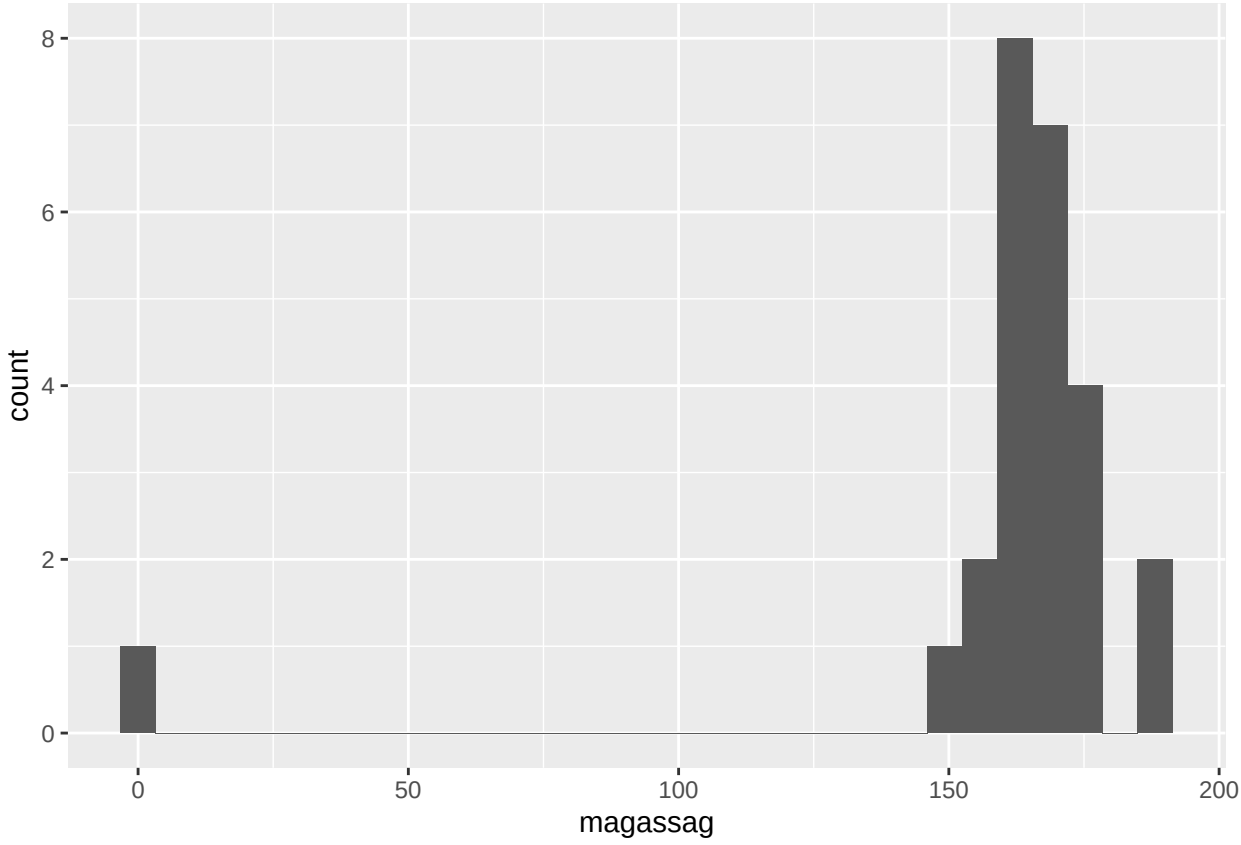
```
mydata %>% summary()
```

```
##     jelige            magassag        cipomeret     alvas_tegnap_1
##  Length:25          Min.   :  1.82   Min.   :35.00   Min.   : 6.00
##  Class :character   1st Qu.:162.00   1st Qu.:37.00   1st Qu.: 7.00
##  Mode  :character   Median :168.00   Median :38.00   Median : 8.00
##                     Mean   :161.07   Mean   :38.94   Mean   : 7.92
##                     3rd Qu.:170.00   3rd Qu.:40.00   3rd Qu.: 9.00
##                     Max.   :190.00   Max.   :47.00   Max.   :10.00
##
##  alvas_tegnap_3  alvas_altalaban_1 alvas_altalaban_3 energiaszint_1
##  Min.   :5.000   Min.   :5.00      Min.   :6.00      Min.   :1.00
```

```
##  1st Qu.:7.000   1st Qu.:6.00      1st Qu.:7.00      1st Qu.:4.00
##  Median :8.000   Median :7.00      Median :7.00      Median :5.00
##  Mean   :7.609   Mean   :7.04      Mean   :7.13      Mean   :5.48
##  3rd Qu.:8.000   3rd Qu.:8.00      3rd Qu.:7.50      3rd Qu.:8.00
##  Max.   :9.000   Max.   :9.00      Max.   :8.00      Max.   :8.00
##  NA's   :2                         NA's   :2
##  energiaszint_3
##  Min.   :3.000
##  1st Qu.:4.500
##  Median :7.000
##  Mean   :6.174
##  3rd Qu.:8.000
##  Max.   :9.000
##  NA's   :2
```

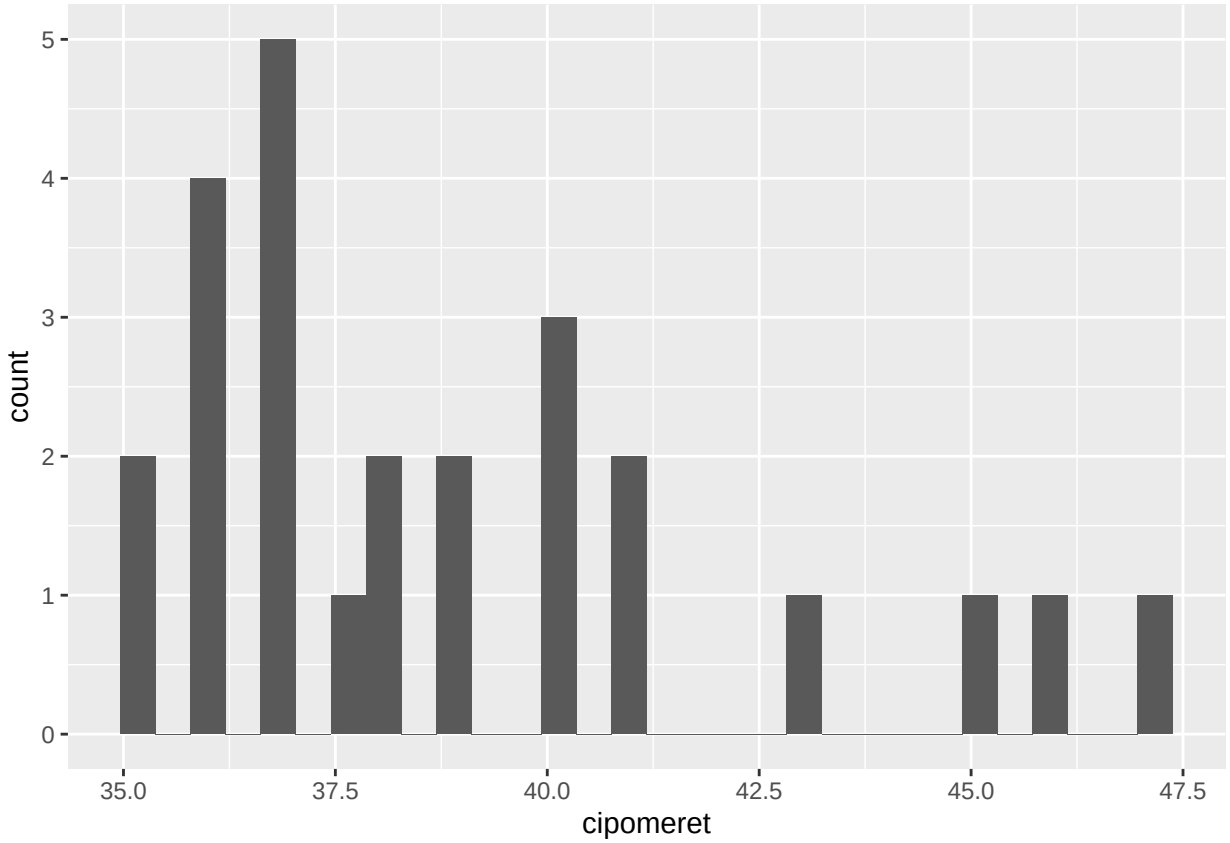
```
# histograms
mydata %>% ggplot() + aes(x = magassag) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

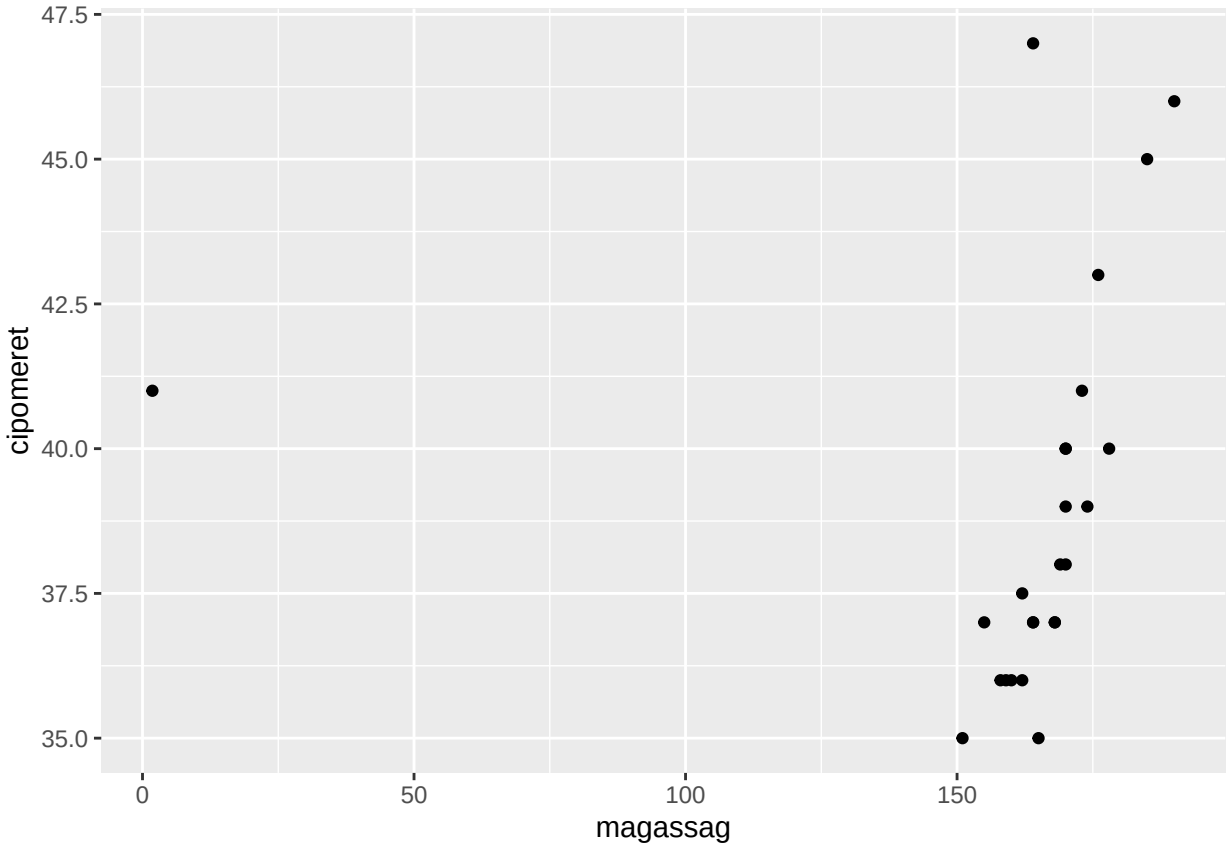



```
mydata %>% ggplot() + aes(x = cipomeret) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# scatterplot
mydata %>% ggplot() + aes(x = magassag, y = cipomeret) + geom_point()
```



Adattisztitas es ellenorzes

```
mydata_corrected = mydata %>% mutate(magassag = replace(magassag,
    magassag == 1.82, 182))

# descriptive statistics
describe(mydata_corrected)
```

```
## Warning in describe(mydata_corrected): NAs introduced by coercion
```

```
##                    vars  n   mean   sd median trimmed  mad min max range
## jelige*               1 25  32.00   NA     32   32.00 0.00  32  32     0
## magassag              2 25 168.28 9.22    168  167.90 8.90 151 190    39
## cipomeret             3 25  38.94 3.34     38   38.60 2.97  35  47    12
## alvas_tegnap_1        4 25   7.92 1.26      8    7.90 1.48   6  10     4
## alvas_tegnap_3        5 23   7.61 1.03      8    7.68 1.48   5   9     4
## alvas_altalaban_1     6 25   7.04 1.02      7    7.10 1.48   5   9     4
## alvas_altalaban_3     7 23   7.13 0.63      7    7.16 0.00   6   8     2
## energiaszint_1        8 25   5.48 2.16      5    5.62 2.97   1   8     7
## energiaszint_3        9 23   6.17 1.92      7    6.26 1.48   3   9     6
##                     skew kurtosis   se
## jelige*               NA       NA   NA
## magassag            0.43    -0.25 1.84
## cipomeret           0.99    -0.05 0.67
## alvas_tegnap_1     -0.22    -1.12 0.25
## alvas_tegnap_3     -0.63    -0.11 0.22
```

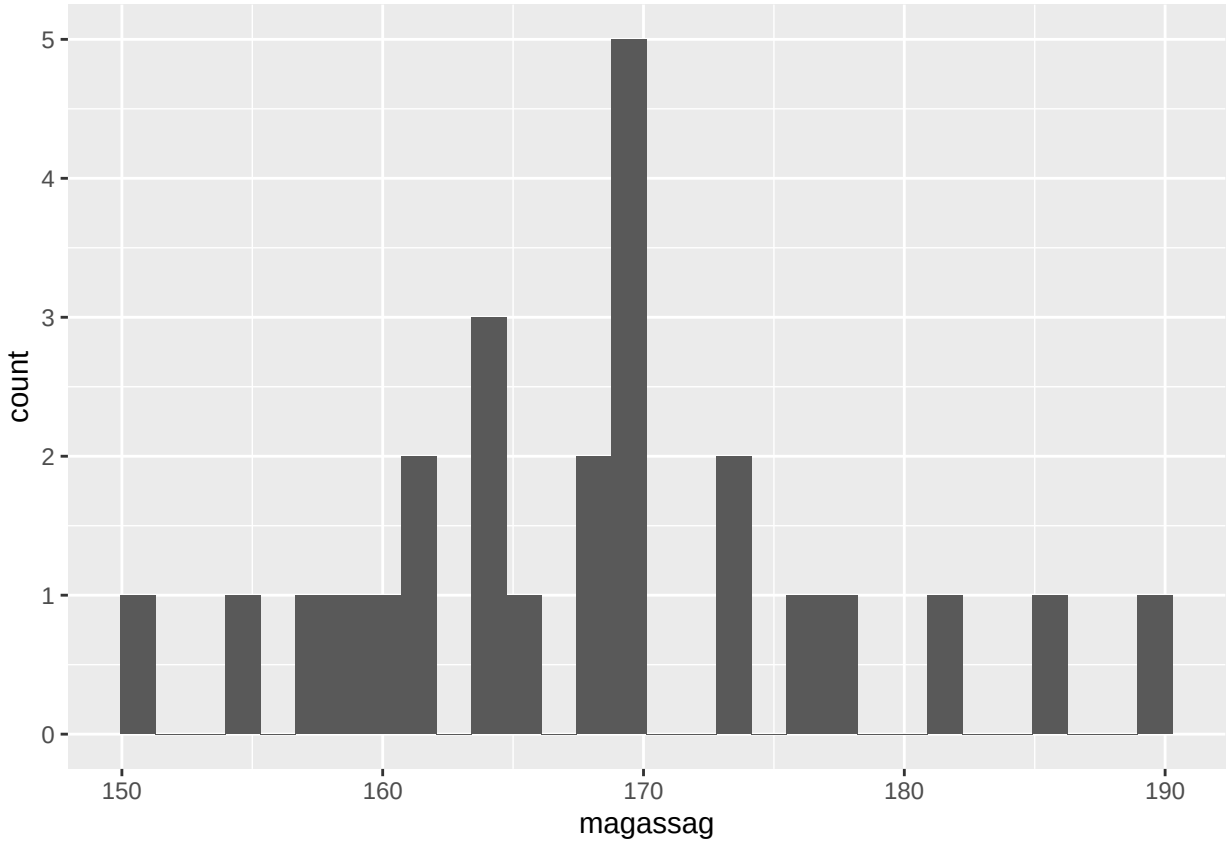
```
## alvas_altalaban_1 -0.30    -0.71 0.20
## alvas_altalaban_3 -0.07    -0.63 0.13
## energiaszint_1    -0.27    -1.21 0.43
## energiaszint_3    -0.38    -1.35 0.40
```

```
mydata_corrected %>% summary()
```

```
##     jelige             magassag       cipomeret     alvas_tegnap_1
##  Length:25          Min.   :151.0   Min.   :35.00   Min.   : 6.00
##  Class :character   1st Qu.:162.0   1st Qu.:37.00   1st Qu.: 7.00
##  Mode  :character   Median :168.0   Median :38.00   Median : 8.00
##                     Mean   :168.3   Mean   :38.94   Mean   : 7.92
##                     3rd Qu.:173.0   3rd Qu.:40.00   3rd Qu.: 9.00
##                     Max.   :190.0   Max.   :47.00   Max.   :10.00
##
##  alvas_tegnap_3  alvas_altalaban_1 alvas_altalaban_3 energiaszint_1
##  Min.   :5.000   Min.   :5.00      Min.   :6.00      Min.   :1.00
##  1st Qu.:7.000   1st Qu.:6.00      1st Qu.:7.00      1st Qu.:4.00
##  Median :8.000   Median :7.00      Median :7.00      Median :5.00
##  Mean   :7.609   Mean   :7.04      Mean   :7.13      Mean   :5.48
##  3rd Qu.:8.000   3rd Qu.:8.00      3rd Qu.:7.50      3rd Qu.:8.00
##  Max.   :9.000   Max.   :9.00      Max.   :8.00      Max.   :8.00
##  NA's   :2                         NA's   :2
##  energiaszint_3
##  Min.   :3.000
##  1st Qu.:4.500
##  Median :7.000
##  Mean   :6.174
##  3rd Qu.:8.000
##  Max.   :9.000
##  NA's   :2
```

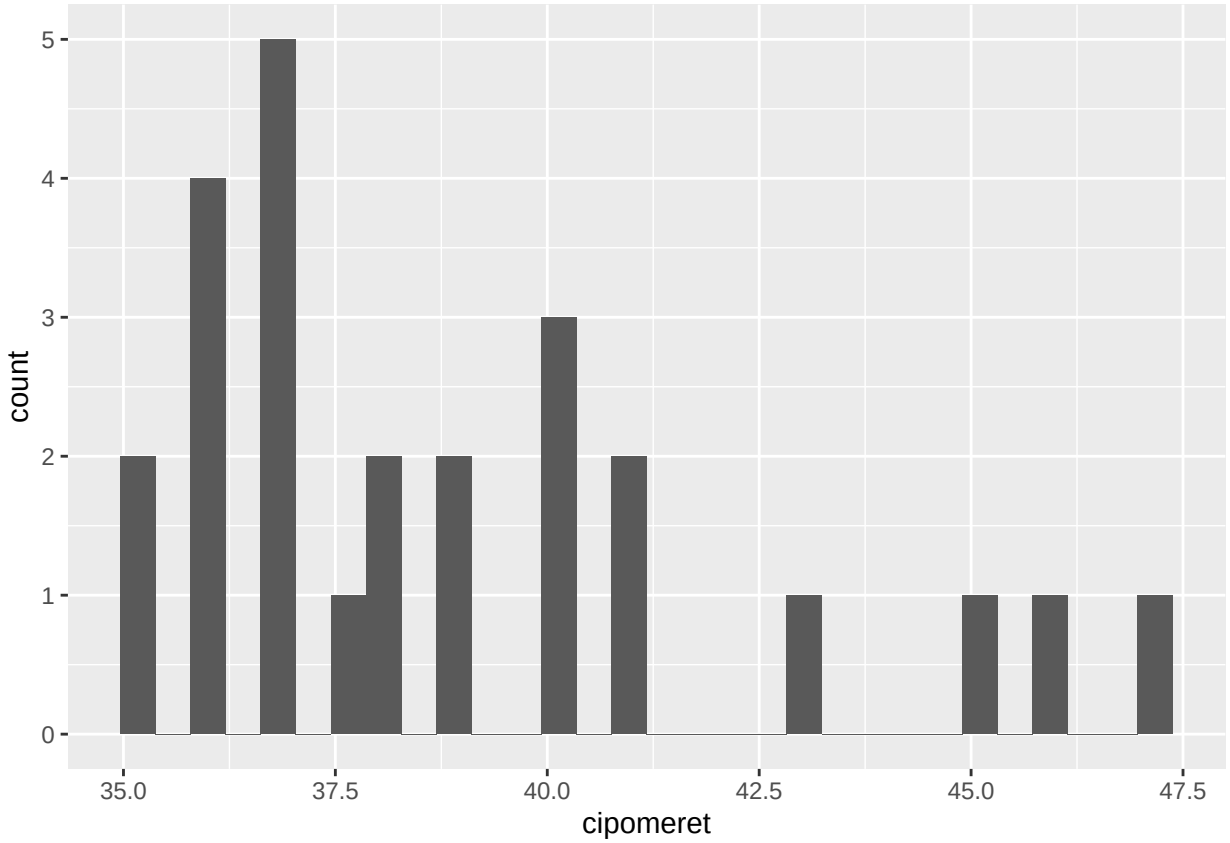
```
# histograms
mydata_corrected %>% ggplot() + aes(x = magassag) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

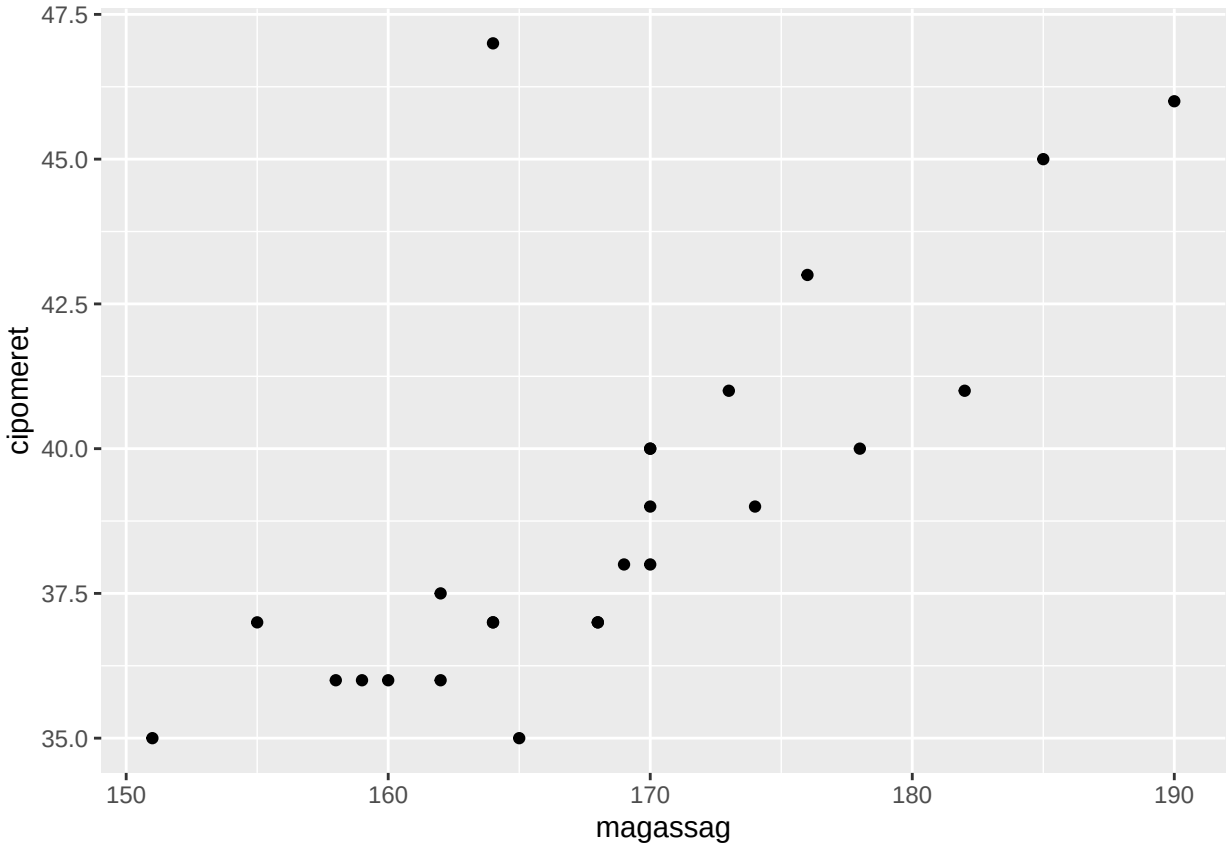



```
mydata_corrected %>% ggplot() + aes(x = cipomeret) + geom_histogram()
```

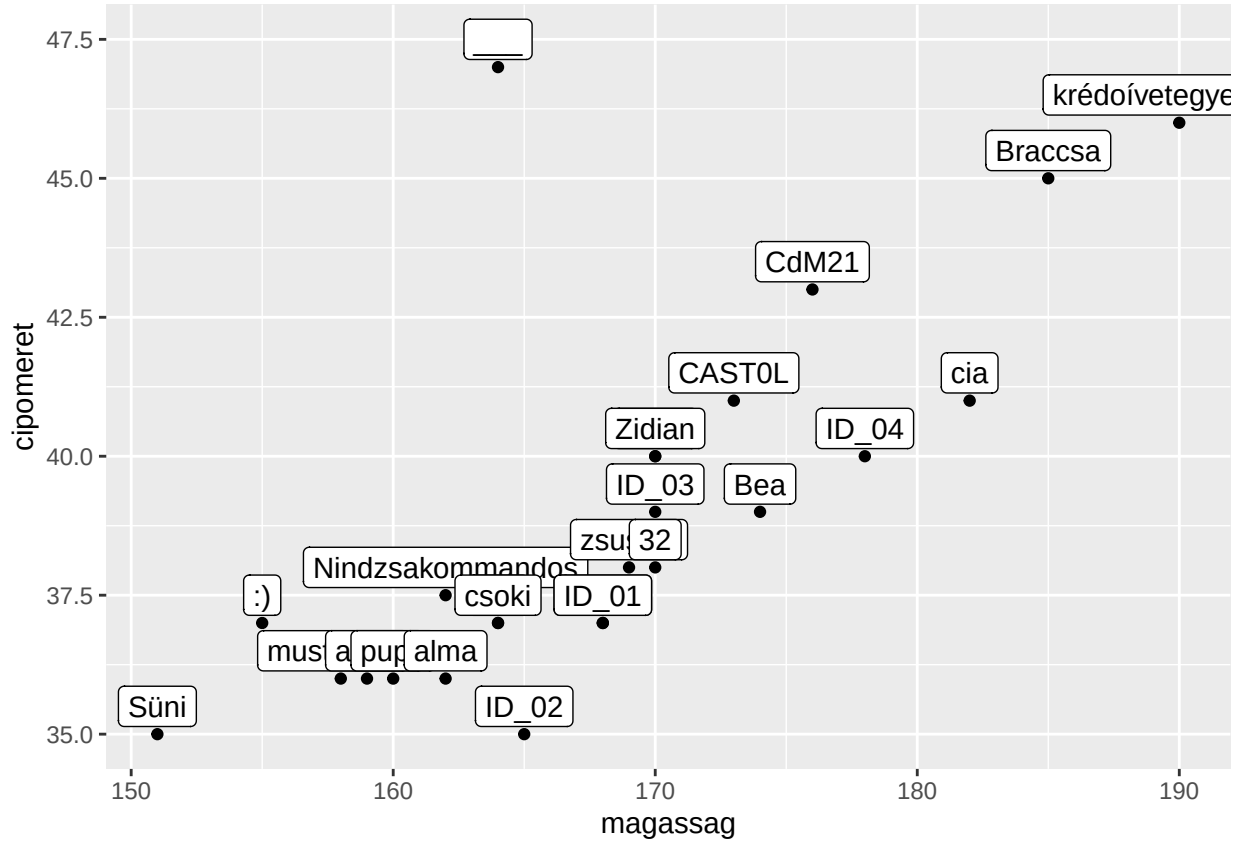
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# scatterplot
mydata_corrected %>% ggplot() + aes(x = magassag, y = cipomeret) +
    geom_point()
```

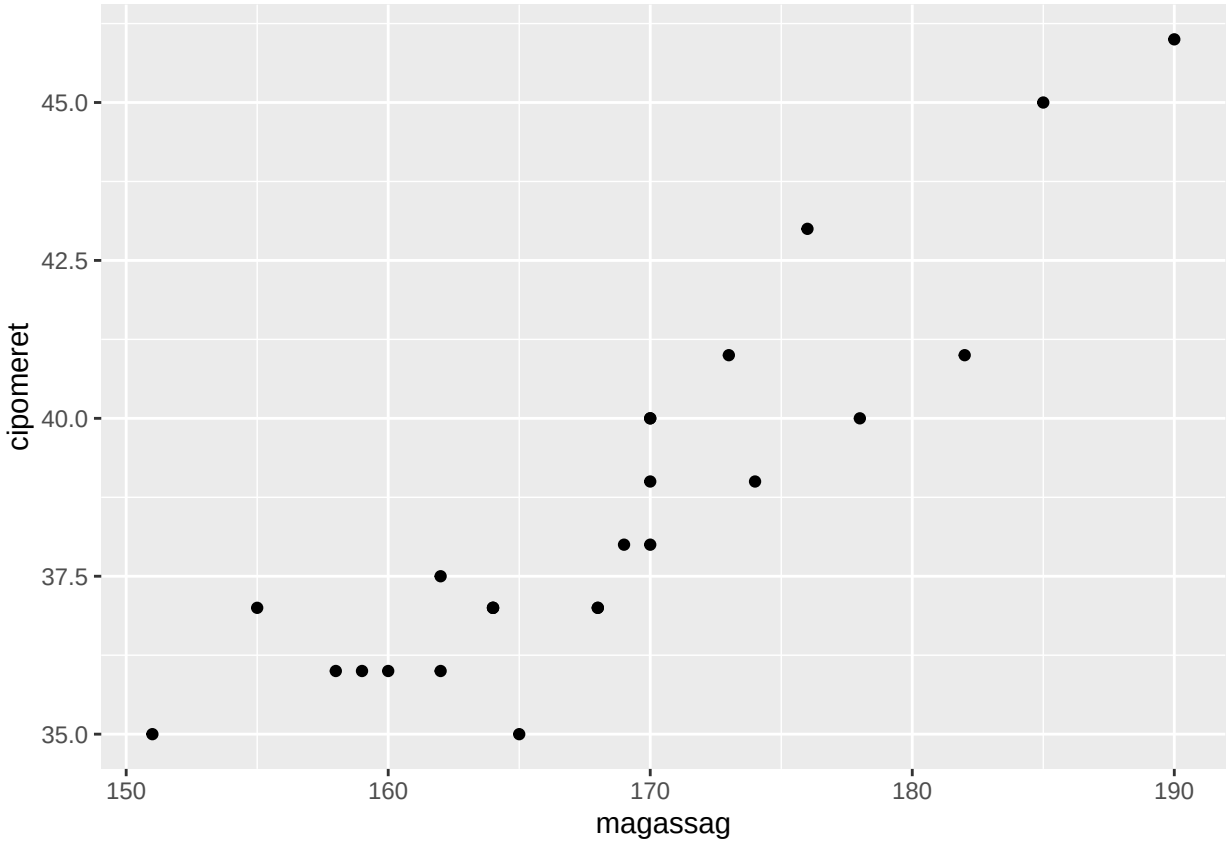



```
# scatterplot with labels
mydata_corrected %>% ggplot() + aes(x = magassag, y = cipomeret,
    label = jelige) + geom_point() + geom_label(nudge_y = 0.5)
```




```
mydata_corrected = mydata_corrected %>% mutate(cipomeret = replace(cipomeret,
    jelige == "___", 37))

# scatterplot
mydata_corrected %>% ggplot() + aes(x = magassag, y = cipomeret) +
    geom_point()
```



Bejoslas linearis modellel

Egyszeru linearis modell felepitese

A regreszio "bejoslasra" valo. Vagyis szeretnenk megtudni egy valtozo erteket (ezt altalaban bejosolt valtozonak vagy kimeneti valtozonak nevezzuk) mas bejoslo (prediktor) valtozok erteke alapjan.

Az alabbi peldaban szeretnenk megbecsulni (bejosolni/prediktalni) az egyes szemelyek EU cipomeretet a magassaguk (bejoslo valotozo) ismereteben. Ehhez eloszor az elozetes adataink hasznalataval felepitunk egy regresszios modellt.

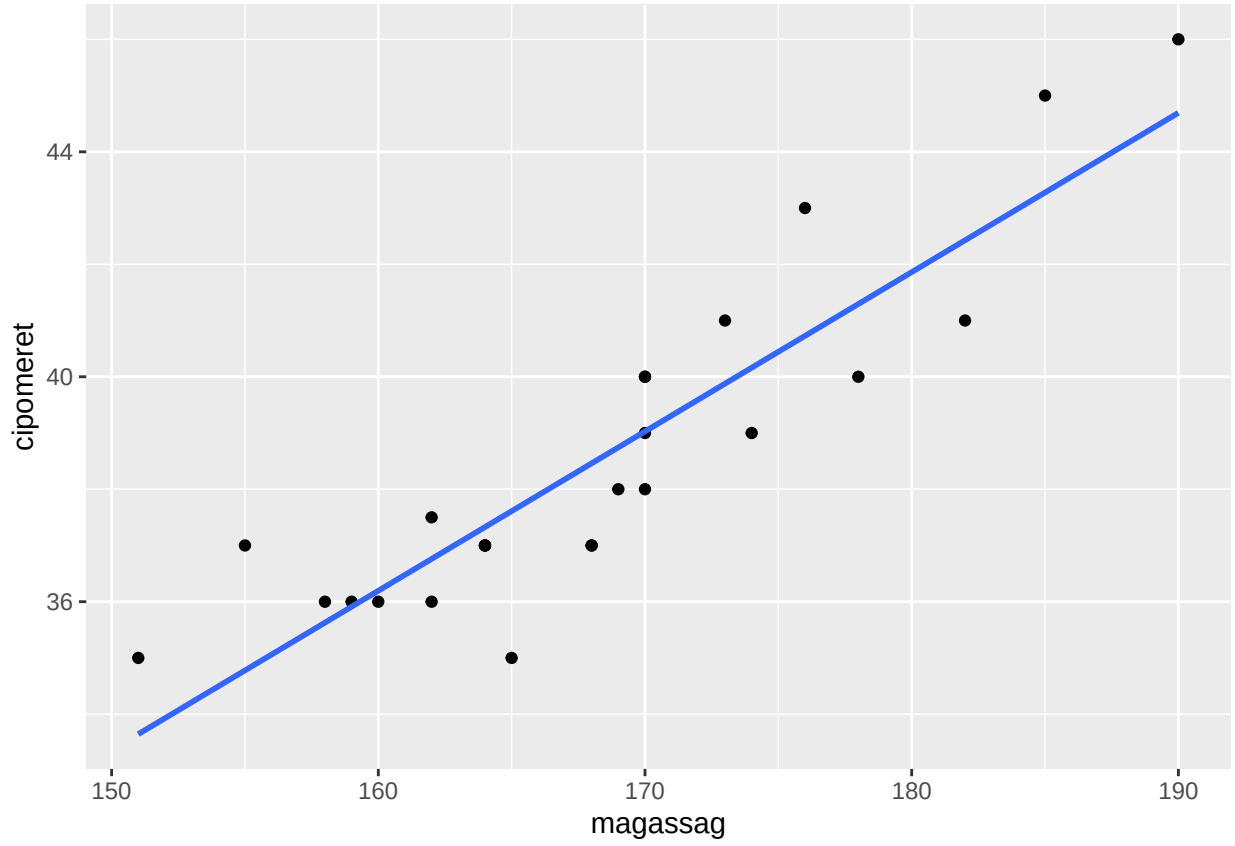
A linearis regresszios modellt az lm() funkcioval epitjuk. Mindig ugy kell felepiteni, hogy a bejosolni kivant valtozoval kezdunk (cipomeret), majd a ~ jel utan irjuk a bejoslo valtozot (magassag). A kod vegen pedig azt specifikaljuk, melzik adattablaban talalhatoak ezek a valtozok a data = parameterrel. A modellt elmenthetjuk egy objektumba (mod1).

(Az egyszeru linearis regresszional (simple linear regression) csak egy bejoslo valtozonk van, de barmennyi bejoslo valtozot hasznalhatunk, ilyenkor csak + jellel elvalasztva egymas utan irva be lehet mindet epiteni egyetlen modellbe, hogy javitsuk a modellunk bejoslo erejet.)

```
mod1 <- lm(cipomeret ~ magassag, data = mydata_corrected)
```

Az linearis regresszioban a kimeneti valtozo es a prediktor kozotti kapcsolatot egy egyenessel modellezzuk. A modell az az ehgyenes lesz ami a legkozelebb van a pont diagram pontjaihoz.

```
mydata_corrected %>% ggplot() + aes(x = magassag, y = cipomeret) +
    geom_point() + geom_smooth(method = "lm", se = F)
```



A regresszios modell megad egy matematikai egyenletet, amibe a prediktor valtozo erteket behelyettesitve megkaphatjuk a legjobb becslest a kimeneti valtozo ertekere. Ezt az egyenletet regresszios egyenletnek (regression equation) nevezzuk.

A regresszios egyenletet igy formalizaljuk: Y = b0 + b1*X1, amelyben Y a kimeneti (bejosolt) valtozo becsult erteke, a b0 egy konstans ertek, amit legtobbszor intercept-nek neveznek, a b1 a regresszios egyutthato, az x1 pedig a bejososlo (prediktor) erteke az adott szemelynel.

Vagyis ugy kaphatunk egy becslest az Y bejosolt valotozo ertekere (magassag), ha a konstanshoz hozzaadjuk a regresszios egyutthato es a prediktor ertekenek szorzatat.

Ha kilistazzuk a modell objektumot (mod1), akkor megkaphatjuk a regresszios egyenletet erre a modellre amit most epitettunk.

```
mod1
```

```
##
## Call:
## lm(formula = cipomeret ~ magassag, data = mydata_corrected)
##
## Coefficients:
## (Intercept)     magassag
##     -9.1332       0.2833
```

Ez azt jelenti, hogy a cipomeretet bejoslo regresszios egyenlet a kovetkezo:

cipomeret = -9.13 + 0.28 * magassag

vagyis egy 170 cm magas ember eseten a modell altal becsult cipomeret:

-9.13 + 0.28 * 170 = 38.47

Ezt a szamitast nem kell kezzel vagy fejben megcsinalni, ehelyett hasznalhatod az R predict() funciojat a bejosolt ertek kiszamitasara barmilyen, vagy akar tobb prediktor-ertekre is.

A predict() funkcio hasznalatahoz meg kell adnunk egy adattablat (data.frame vagy tibble-t) ami a prediktor ertekeit tartalmazza, amit a kimeneti valtozo megbeszlesere, bejoslasara szeretnenk hasznalni.

```
magassag = c(150, 160, 170, 180, 190)
magassag_df = as_tibble(magassag)
```

```
## Warning: Calling `as_tibble()` on a vector is discouraged, because the behavior is likely to change i
## This warning is displayed once per session.
```

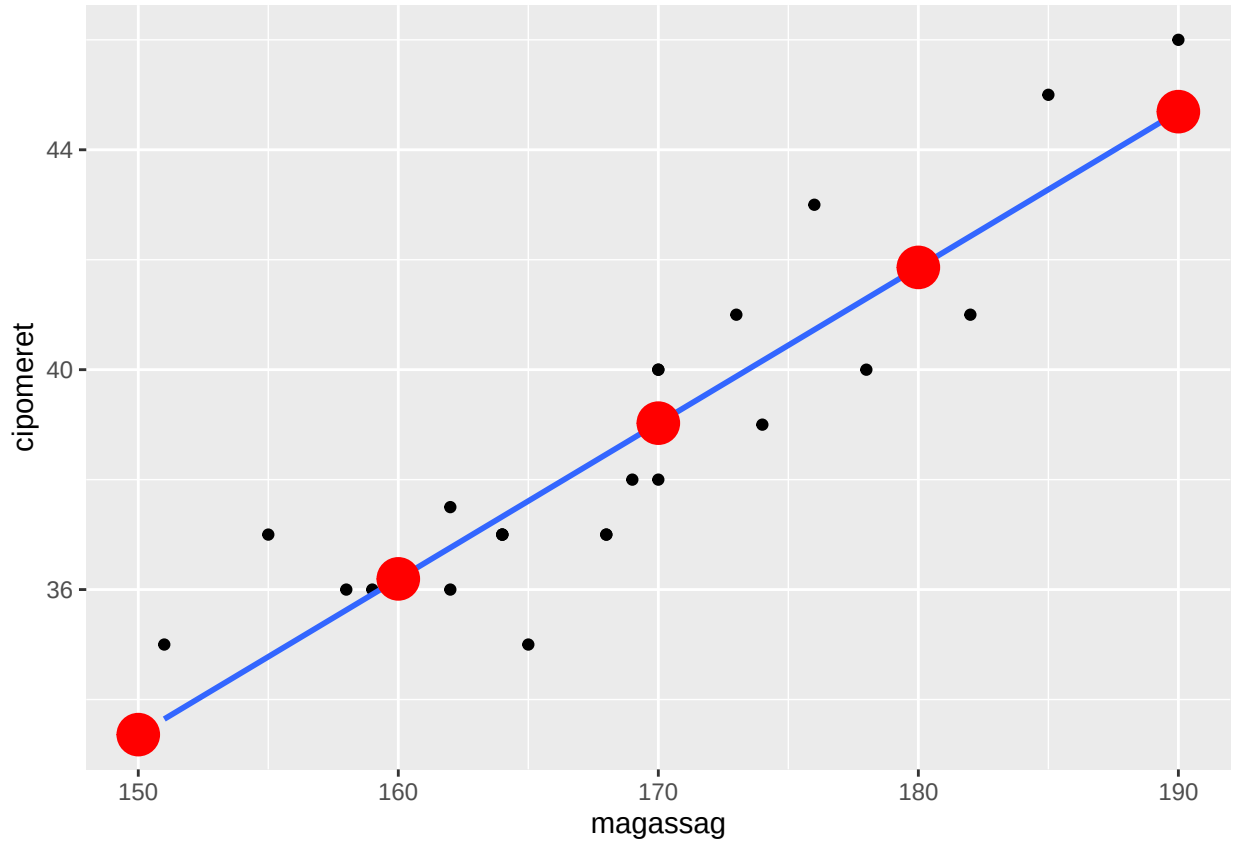
```
predictions = predict(mod1, newdata = magassag_df)

magassag_df_with_predicted = cbind(magassag_df, predictions)
magassag_df_with_predicted
```

```
##   value predictions
## 1   150    33.36134
## 2   160    36.19430
## 3   170    39.02727
## 4   180    41.86024
## 5   190    44.69321
```

Predicted values all fall on the regression line

```
mydata_corrected %>% ggplot() + aes(x = magassag, y = cipomeret) +
    geom_point() + geom_smooth(method = "lm", se = F) + geom_point(data = magassag_df_with_predicted,
    aes(x = value, y = predictions), col = "red", size = 7)
```



_____Gyakorlas_____

1. Epits egy egyszeru linearis regresszio modellt az lm() fugvennyel amiben az **energiaszint_1** a kimeneti valtozo es az **alvas_altalaban_1** a prediktor. A modell eredmenyet mentsd el egy objektumba mentsd.
2. Ird le a regresszios fuggvenyt amivel bejosolhato az energiaszint.
3. Ertelmezd a regresszios fuggvenyt. Aki tobbet alszik annak magasabb vagy alacsonyabb az energiaszintje? (Egy abra segithet)
4. Ertelmezd a regresszios fuggvenyt. Aki egy oraval tobbet alszik mint masok, annak mennyivel varhato hogy magasabb lesz az energiaszintje?
5. Ennek a modellnek a segitsegevel becsuld meg az energiaszintjet olyan embereknek akik altalaban 5, 7, vagy 9 orat alszanak.

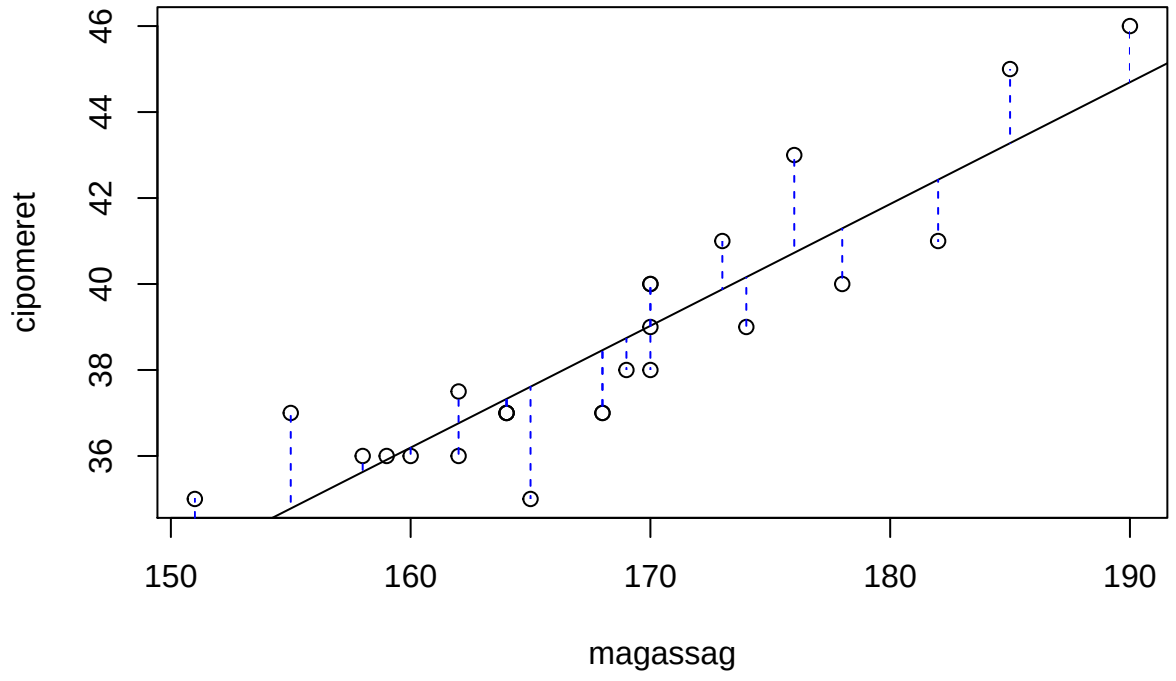
Milyen jo a modellem? (modellilleszkedes)

Hogyan merheto a becslesi/bejoslasi hatekonysag?

A modell becslesi hatekonysagat tobb fele keppen lehet merni. A legkezenfekvobb modszer, hogy meghatarozzuk, a modell becslese mennyire esett tavol a valos bejosolni kivant ertekektol. Vagyis megmerjuk a modell figyelembevetele utan fennmarado "hibat".

Ezt konnyen megtehetjuk egy olyan adatbazisban, ahol rendelkezesunkre all a bejosolni kivant valtozo valos erteke, ugy hogy kivonjuk egymasbol a valos erteket es a modell altal becsult erteket. Ez a rezidualis (fennmarado) hiba, masneven **residual error**.

```
error_plotter(mod1, col = "blue")
```



Ha vesszuk az osszes ilyen hiba ertek abszoluterteket, es osszeadjuk oket, megkapjuk a modell rezidualis

abszolut hiba (residual absolute difference - RAD) erteket.

Ennel azonban joval gyakoribb hogy a rezidualis hiba negyzetosszeget hasznaljak (residual sum of squared - RSS) a statisztikaban. Vagyis az egyes rezidualis hiba ertekeket negyzetre emelik, majd osszeadjak oket.

Az alabbi peldaban a mod1 eredeti adattablajanak magassagertekeit hasznaljuk a cipomeret becsult ertekenek kiszamitasara (predict(mod1)), es ezt vonjuk ki az ugyan ezen adattablaban szereplo valos cimpomeret ertekekbol, igy kapjuk meg a rezidualis hibaertekeket. Majd egyenkent vagy az abszolutertekeuket (RAD), vagy a negyzetuket vesszuk (RSS), es osszeadjuk oket a sum() fugvennyel.

```
RAD = sum(abs(mydata_corrected$cipomeret - predict(mod1)))
RAD
```

```
## [1] 26.29851
```

```
RSS = sum((mydata_corrected$cipomeret - predict(mod1))^2)
RSS
```

```
## [1] 39.15215
```

Hasznos a modellunk?

Azt, hogy mennyire hasznos a modellunk (mennyit nyerunk azzal, hogy ezt a modellt hasznaljuk), meghatarozhatjuk ugy, hogy osszehasonlitjuk a rezidualis hibat abban az esetben amikor a modellunket hasznaljuk (vagyis amikor figyelembe vesszuk a prediktoraink erteket) egy olyan esettel, amikor a prediktorokat egyalatalan nem vesszuk figyelembe, csak a bejosolni kivant valtozo atlagat hasznaljuk a becslesre.

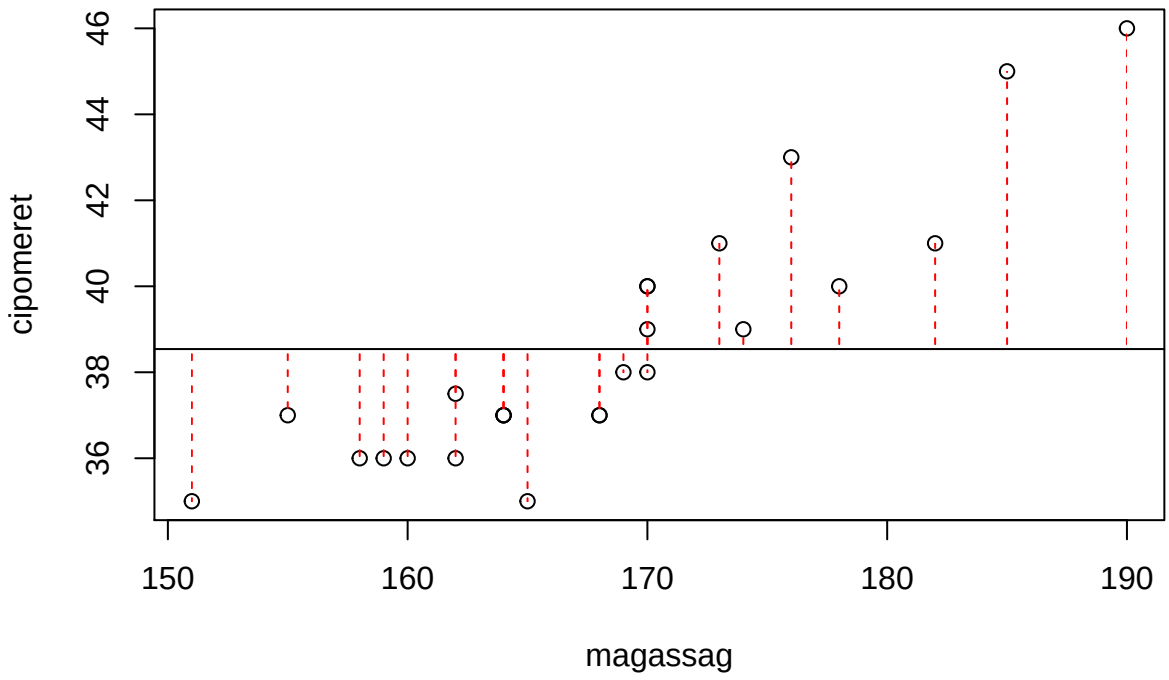
Az alabbi kodban epitunk egy olyan uj modellt, ahol nem veszunk figyelembe semmilyen masik valtozot, csak a cipomeret atlagat, es azt hasznaljuk fel a cipomeret becslesekent. (pl. ha tudjuk, hogy a populacioban az atlagos cipomeret 38, akkor mindenkinek ezt a cipomeretet becsuljuk majd, fuggetlenul attol, hogy milyen magas az illeto). Ezt a modellt **null modellnek** nevezzuk. Azt, hogy a bejosolt valtozo atlagat akarjuk becslesre hasznalni, ugy adhatjuk meg, hogy a ~ utan csak egy 1-est rakunk, nem irunk mas valtozonevet.

Ez persze nagy rezidualis hibahoz vezet (hiszen bar ez a populacioban az atlagos, megis a legtobb embernek nem pont 38-as a laba). A null modell altal predukalt rezidualis hibat ugyan ugy szamoljuk ki, mint a tobbi modellnel a residual sum of squared-et, viszont ennek van egy specialis neve is az irodalomban, ezt ugy hivjak, hogy total sum of squared (TSS), mert ez a lehetseges legegyszerubb meg ertelmes modell, ami altalaban azert nagy hibaval jar, igy ezt vesszuk a “teljes” hiba mennyisegnek, es ehhez viszonyitjuk a tobbi modell altal elert hibat.

Alabb kiszamoljuk a TSS-t. Lathato hogy a formula ugyan az mint az RSS eseten.

```
mod_mean <- lm(cipomeret ~ 1, data = mydata_corrected)

error_plotter(mod_mean, col = "red", x_var = "magassag")  # visualize error
```




```
TSS = sum((mydata_corrected$cipomeret - predict(mod_mean))^2)
TSS
```

```
## [1] 202.96
```

Azt, hogy mennyi informaciot nyertunk a kimeneti valtozo valtozekonysagarol (variance) a prediktorok figyeklembevetelevel ahhoz kepest ha a null modellt vettuk volna figyelembe, az R^2 statisztika mutatja meg. Ennek a formulaja: 1-(RSS/TSS)

```
R2 = 1 - (RSS/TSS)
R2
```

```
## [1] 0.8070942
```

Ez azt jelenti, hogy a prediktorok figyelembevetelevel (a mi esetunkben ez a magassag), a cipomeret valtozekonysaganak 80.71%-at tudjuk megmagyarazni.

R^2 = 1 azt jelenti, hogy a kimeneti valtozo variabilitasat teljesen meg tudjuk magyarazni a prediktorok ismereteben.

R^2 = 0 azt jelenti, hogy a kimeneti valtozo variabilitasat egyaltalan nem magyarazzak meg a prediktorok

A prediktorokat tartalmazo modell szigifikansan jobb mint a null modell a kimeneti valtozo becslesere?

A ket modell altal produkalt rezidualis hibat az anova() funkcioval hasonlithatjuk ossze, melybe a null modell es a prediktorokat tartalamzo modell objektumot kell beletenni (akar tobbet modellt is lehet egyszerre). Az anova() szignifikancia erteket is ad.

```
anova(mod_mean, mod1)
```

```
## Analysis of Variance Table
##
## Model 1: cipomeret ~ 1
## Model 2: cipomeret ~ magassag
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     24 202.960
## 2     23  39.152  1    163.81 96.229 1.097e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Az egyszeru megoldas

A modell summary() kikeresevel mindez a fenti informacio megkaphato, es meg tobb is. Itt megtalalod az R^2 erteket, az RSS-t, a modell null modellel valo osszehasonlitasanak F teszt statisztikajat es szignifikanciajat, es meg a regresszios egyenletet is.

```
summary(mod1)
```

```
##
## Call:
## lm(formula = cipomeret ~ magassag, data = mydata_corrected)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6108 -1.0273 -0.1943  0.9727  2.2729
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -9.13318    4.86683  -1.877   0.0733 .
## magassag     0.28330    0.02888   9.810  1.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.305 on 23 degrees of freedom
## Multiple R-squared:  0.8071, Adjusted R-squared:  0.7987
## F-statistic: 96.23 on 1 and 23 DF,  p-value: 1.097e-09
```

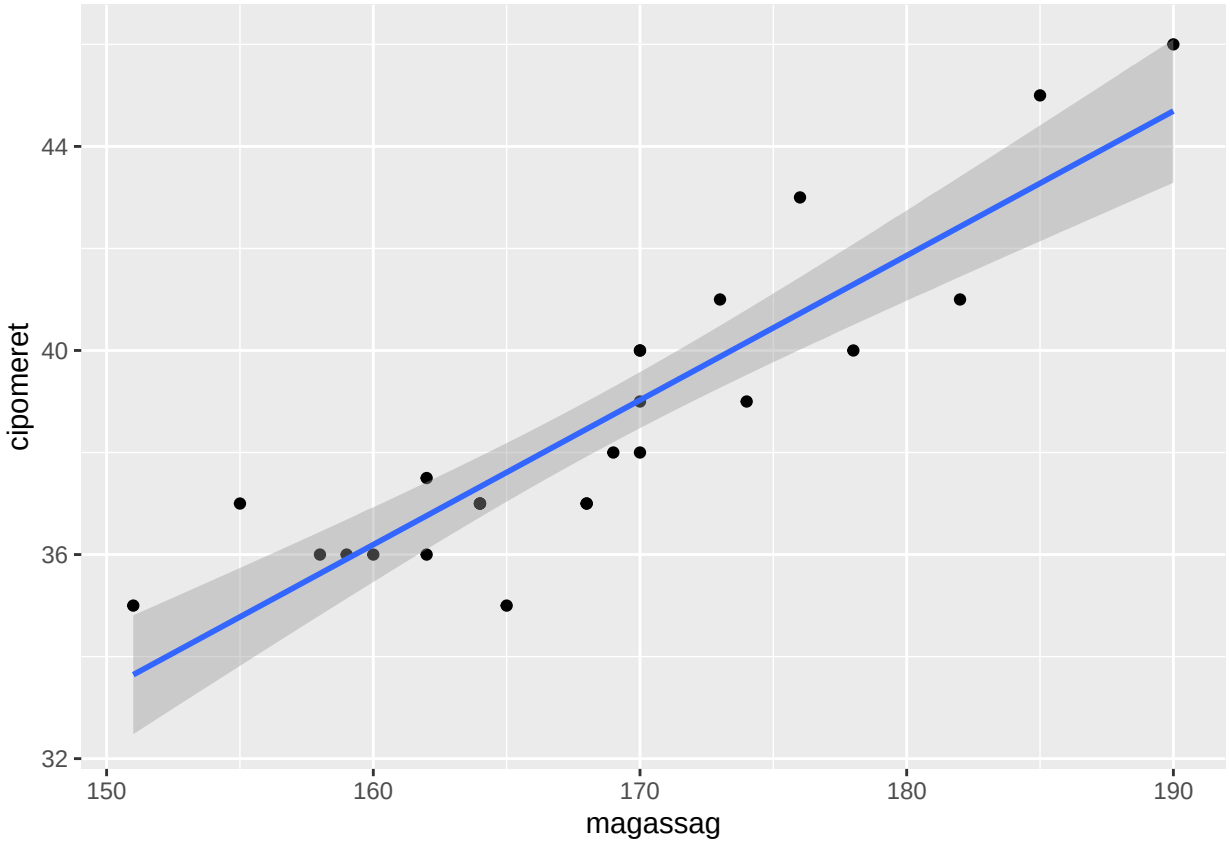
A regresszios egyutthatok (regression coefficients) konfidencia intervallumat a confint() paranccsal lehet kilistazni.

```
confint(mod1)
```

```
##                   2.5 %    97.5 %
## (Intercept) -19.2009731 0.9346187
## magassag      0.2235552 0.3430383
```

A regresszios becsles konfidencia intervallumat pedig a geom_smooth()-al lehet vizualizalni.

```
ggplot(mydata_corrected, aes(x = magassag, y = cipomeret)) +
    geom_point() + geom_smooth(method = "lm", formula = y ~ x)
```



_____Gyakorlas_____

1. (Ezt nem kell megtenned ha ezt mar megtetted az elozo gyakorlasban, csak hasznald ugyan azt a model objektumot) Epits egy egyszeru linearis regresszio modellt az lm() fugvennyel amiben az **energiaszint_1** a kimeneti valtozo es az **alvas_altalaban_1** a prediktor. A modell eredmenyet mentsd el egy objektumba mentsd.
2. Listazd ki a model summary-t a summary() fugvennyel
3. Olvasd le a model RSS-jet
4. Olvasd le hogy a model ami tartalmazza az alvas_altalaban_1 prediktort szignifikansan jobb bejosloja-e az energiaszintnek mint a null modell.
5. Hatarozd meg a regresszios egyutthatok konfidencia intervallumat a confint() fuggvennyel

_____