

# Seminar 1 Econ3110

Bernhard Teodor Thodesen

2025-08-31

## Oppgave 1 - Dataframe

```
library(tidyverse)

car <- data.frame(mtcars)
```

a)

Last inn det innebygde datasettet 'mtcars' ved å bruke `data(mtcars)`. Finn gjennomsnitt, median og varians til alle variablene i datasettet.

```
deskriptiv_stat <- function(x) c(mean = round(mean(x),4),
                                  median = round(median(x),4),
                                  varians = round(var(x),4))

kable(sapply(car, deskriptiv_stat))
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mean	20.0906	6.1875	230.7219	146.6875	3.5966	3.2172	17.8487	0.4375	0.4062	3.6875	2.8125
median	19.2000	6.0000	196.3000	123.0000	3.6950	3.3250	17.7100	0.0000	0.0000	4.0000	2.0000
varians	36.3241	3.1895	15360.7998	4700.8669	0.2859	0.9574	3.1932	0.2540	0.2490	0.5444	2.6089

b)

Variabelen 'wt' viser vekten til hver observasjon i 1000 pund (lb). Eitt pund er 0.45 kg. Konverter vektvariabelen til tonn. Lag deretter en ny variabel som er hestekrefter (hp) per tonn.

```
car <- car |>
  mutate(wt = wt*0.45) |>
  mutate(hestekrefter = hp/wt)
```

c)

I datasettet er namna på radane bilmodellene. Lag ei eiga kolonne med namnet på bilmodellene. Endre radnamna i datasettet til radnummer

```
brands <- row.names(car)

car <- car |> mutate(brand = brands)

row.names(car) <- 1:length(brands)
```

d)

Finn namnet på dei tre bilane med høgast verdi av mpg. Finn deretter dei tre bilane med lågast vekt (wt)

```
car |> arrange(desc(wt)) |>
  slice(1:3)
```

```
##      mpg cyl disp  hp drat          wt  qsec vs am gear carb hestekrefter
## 1 10.4   8  460 215 3.00 2.44080 17.82  0  0   3   4      88.08587
## 2 14.7   8  440 230 3.23 2.40525 17.42  0  0   3   4      95.62416
## 3 10.4   8  472 205 2.93 2.36250 17.98  0  0   3   4      86.77249
##
##              brand
## 1 Lincoln Continental
## 2  Chrysler Imperial
## 3  Cadillac Fleetwood
```

```
#car |> slice_max(wt, n = 3)
```

```
car |> slice_min(wt, n = 3)
```

```
##      mpg cyl disp  hp drat          wt  qsec vs am gear carb hestekrefter
## 1 30.4   4  95.1 113 3.77 0.68085 16.90  1  1   5   2     165.96901
## 2 30.4   4  75.7  52 4.93 0.72675 18.52  1  1   4   2      71.55143
## 3 33.9   4  71.1  65 4.22 0.82575 19.90  1  1   4   1      78.71632
##
##              brand
## 1   Lotus Europa
## 2   Honda Civic
## 3 Toyota Corolla
```

e)

Kolonna du laga i c) viser kva både modell og produsent for bilen. Det første ordet i kvar rad viser produsent og resten kva modell det er. Lag ei kolonne 'producer' som viser produsenten til bilen. [Hint: ein måte er å bruke 'substr' og finne dei tre første bokstavane til kvar observasjon av 'brand'.

```
producer <- c()

for (i in 1:length(brands))
{
  splitta_navn <- str_split_1(brands[i], " ")
  producer[i] <- splitta_navn[1]
}

producer
```

```
## [1] "Mazda"      "Mazda"      "Datsun"      "Hornet"      "Hornet"      "Valiant"
## [7] "Duster"     "Merc"       "Merc"       "Merc"       "Merc"       "Merc"
## [13] "Merc"      "Merc"      "Cadillac"    "Lincoln"    "Chrysler"    "Fiat"
## [19] "Honda"     "Toyota"     "Toyota"     "Dodge"     "AMC"        "Camaro"
## [25] "Pontiac"   "Fiat"      "Porsche"    "Lotus"     "Ford"       "Ferrari"
## [31] "Maserati"  "Volvo"
```

```
car <- car |> mutate(producer = producer)
```

```
car |> head()
```

```
##      mpg cyl disp  hp drat   wt  qsec vs am gear carb hestekrefter
## 1  21.0   6  160  110 3.90 1.17900 16.46  0  1   4    4     93.29941
## 2  21.0   6  160  110 3.90 1.29375 17.02  0  1   4    4     85.02415
## 3  22.8   4  108   93 3.85 1.04400 18.61  1  1   4    1     89.08046
## 4  21.4   6  258  110 3.08 1.44675 19.44  1  0   3    1     76.03249
## 5  18.7   8  360  175 3.15 1.54800 17.02  0  0   3    2    113.04910
## 6  18.1   6  225  105 2.76 1.55700 20.22  1  0   3    1     67.43738
##
##      brand producer
## 1      Mazda RX4    Mazda
## 2      Mazda RX4 Wag Mazda
## 3      Datsun 710    Datsun
## 4      Hornet 4 Drive Hornet
## 5 Hornet Sportabout Hornet
## 6      Valiant      Valiant
```

f)

```
#car |>
# group_by(producer) |>
# summarise(mpg = mean(mpg),
#           cyl = mean(cyl),
#           disp = mean(dis))
#

car |>
  select(-brand) |>
  group_by(producer) |>
  summarise_all(~round(mean(.),1)) |>
  kable()
```

producer	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb	hestekrefter
AMC	15.2	8.0	304.0	150.0	3.1	1.5	17.3	0.0	0.0	3.0	2.0	97.0
Cadillac	10.4	8.0	472.0	205.0	2.9	2.4	18.0	0.0	0.0	3.0	4.0	86.8
Camaro	13.3	8.0	350.0	245.0	3.7	1.7	15.4	0.0	0.0	3.0	4.0	141.8
Chrysler	14.7	8.0	440.0	230.0	3.2	2.4	17.4	0.0	0.0	3.0	4.0	95.6
Datsun	22.8	4.0	108.0	93.0	3.9	1.0	18.6	1.0	1.0	4.0	1.0	89.1
Dodge	15.5	8.0	318.0	150.0	2.8	1.6	16.9	0.0	0.0	3.0	2.0	94.7
Duster	14.3	8.0	360.0	245.0	3.2	1.6	15.8	0.0	0.0	3.0	4.0	152.5
Ferrari	19.7	6.0	145.0	175.0	3.6	1.2	15.5	0.0	1.0	5.0	6.0	140.4

producer	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb	hestekrefter
Fiat	29.9	4.0	78.8	66.0	4.1	0.9	19.2	1.0	1.0	4.0	1.0	71.2
Ford	15.8	8.0	351.0	264.0	4.2	1.4	14.5	0.0	1.0	5.0	4.0	185.1
Honda	30.4	4.0	75.7	52.0	4.9	0.7	18.5	1.0	1.0	4.0	2.0	71.6
Hornet	20.0	7.0	309.0	142.5	3.1	1.5	18.2	0.5	0.0	3.0	1.5	94.5
Lincoln	10.4	8.0	460.0	215.0	3.0	2.4	17.8	0.0	0.0	3.0	4.0	88.1
Lotus	30.4	4.0	95.1	113.0	3.8	0.7	16.9	1.0	1.0	5.0	2.0	166.0
Maserati	15.0	8.0	301.0	335.0	3.5	1.6	14.6	0.0	1.0	5.0	8.0	208.5
Mazda	21.0	6.0	160.0	110.0	3.9	1.2	16.7	0.0	1.0	4.0	4.0	89.2
Merc	19.0	6.3	207.2	134.7	3.5	1.6	19.0	0.6	0.0	3.6	3.0	82.9
Pontiac	19.2	8.0	400.0	175.0	3.1	1.7	17.0	0.0	0.0	3.0	2.0	101.1
Porsche	26.0	4.0	120.3	91.0	4.4	1.0	16.7	0.0	1.0	5.0	2.0	94.5
Toyota	27.7	4.0	95.6	81.0	4.0	1.0	20.0	1.0	0.5	3.5	1.0	83.1
Valiant	18.1	6.0	225.0	105.0	2.8	1.6	20.2	1.0	0.0	3.0	1.0	67.4
Volvo	21.4	4.0	121.0	109.0	4.1	1.3	18.6	1.0	1.0	4.0	2.0	87.1

## Oppgave 2 - bootstrapping

a)

Last inn 'mtcars' på nytt. Trekk 32 tilfeldige radnummer frå 'mtcars' med tilbakelegging, og lag ein ny data.frame du kallar 'df' med desse trekte observasjonane. [Du kan gjere det som beskrive ovanfor med 'sample', evt. bruk 'sample\_n'/'slice\_sample' frå 'dplyr'-pakken.]

```

bil <- data.frame(mtcars)

bil <- bil |>
  slice_sample(n = 32, replace = T)

```

a) ekstra

Lag ein funksjon i R som for ein kvar data.frame trekk n tilfeldige radnummer med tilbakelegging, der n er talet rader i data.frameen, og returnerer ein data.frame med desse n trekte observasjonane.

```

slicer <- function(df, n, tilbakelegging = 1) # Kun base r
  ### INput: df: dataframe, n: int ###
  ### funksjonen returnere en dataframe med n tilfeldige observasjoner fra df. ###
  ### Output: dataframe###
{
  rader <- sample(1:nrow(df), n, replace = tilbakelegging)
  return(df[rader,])
}

slicer(bil, 32) |> head()

```

```

##           mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Hornet 4 Drive...22 21.4   6 258.0 110 3.08 3.215 19.44 1 0   3   1
## AMC Javelin       15.2   8 304.0 150 3.15 3.435 17.30 0 0   3   2
## Merc 280C...24    17.8   6 167.6 123 3.92 3.440 18.90 1 0   4   4

```

```
## Merc 280C...25      17.8   6 167.6 123 3.92 3.440 18.90  1  0   4   4
## Dodge Challenger   15.5   8 318.0 150 2.76 3.520 16.87  0  0   3   2
## Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02  0  0   3   2
```

b)

Bruk ein for-loop til å repetere a) 1000 gonger. For kvar replikasjon, lagre gjennomsnittet til 'mpg' i ein vektor du kallar 'bs'

```
bs <- c()

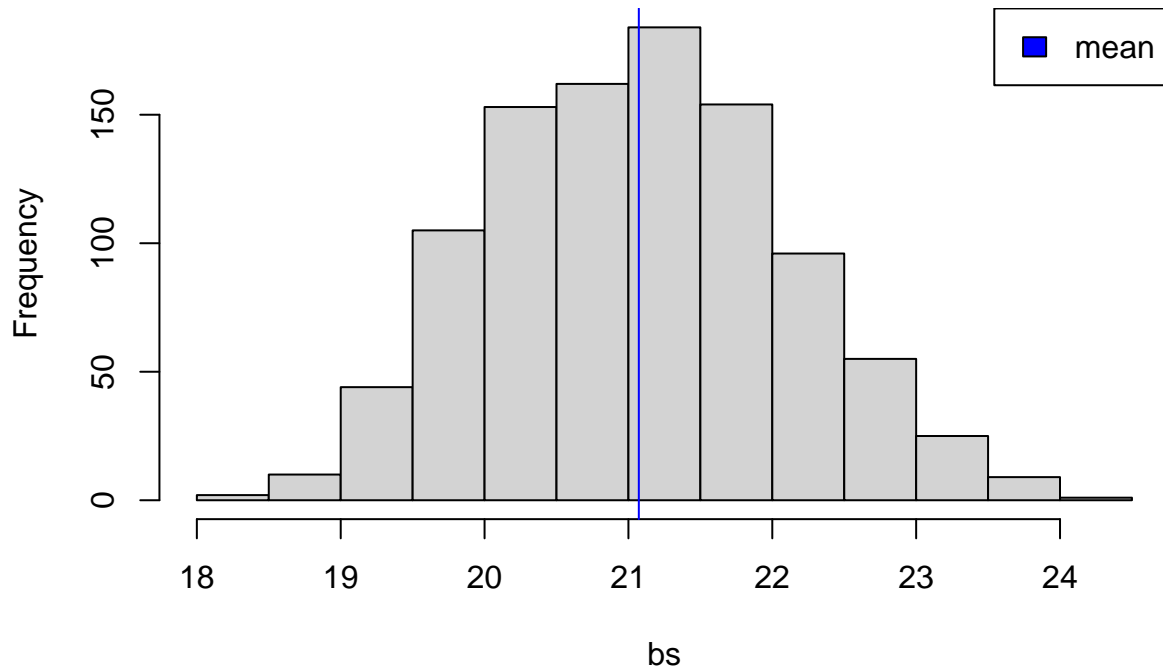
for(i in 1:1000)
{
  a <- slicer(bil, 32)
  bs[i] <- mean(a$mpg)
}
```

c)

Lag eit histogram over 'bs' og finn gjennomsnitt og varians. Samanlikn med gjennomsnitt og varians til 'mpg' i 'mtcars' (eller 'var(mtcars\$mpg)/32').

```
hist(bs,main = paste("Histogram av bs gj:",round(mean(bs),3), "varians: ", round(var(bs),3)))
abline(v = mean(bs), col = "blue")
legend("topright", legend = "mean", fill = "blue")
```

## Histogram av bs gj: 21.072 varians: 1.059



Estimator varians:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

$$\frac{S^2}{32} = \frac{\sum (x_i - \bar{x})^2}{n - 1} \cdot \frac{1}{32}$$

Hvorfor er vi interessert i dette???

## Oppgave 3 - OLS

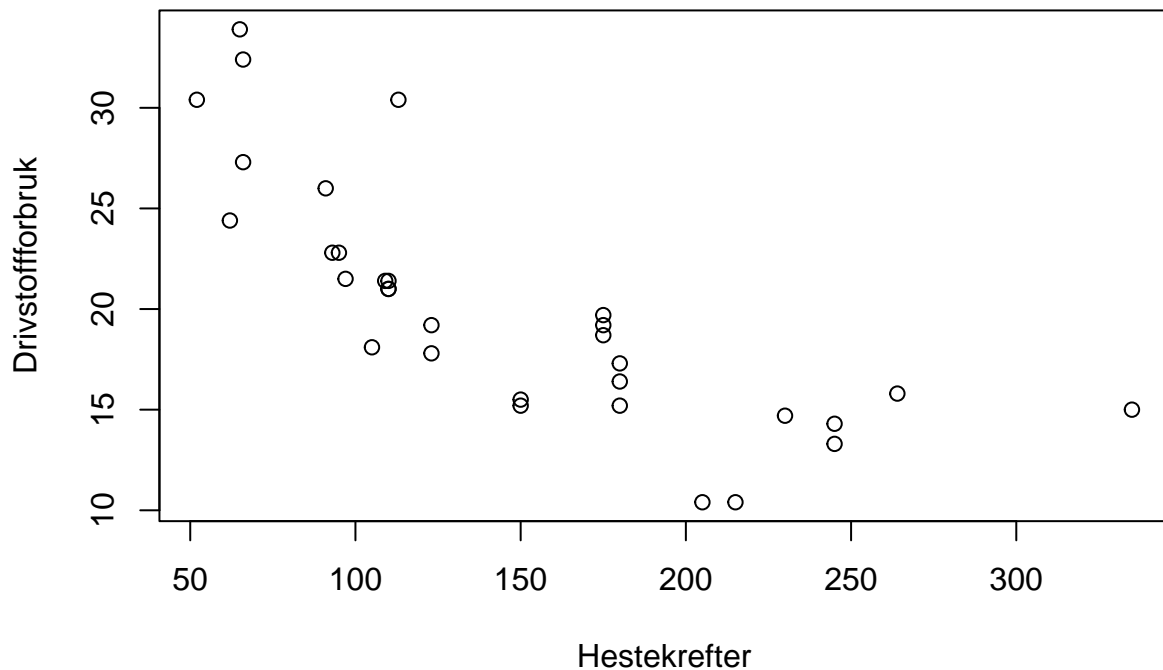
a)

Last inn 'mtcars'. Lag eit spredningsdiagram med drivstofforbruk (mpg) på y-aksen og hestekrefter (hp) på x-aksen. Lag passende tittel og namn på aksane. Er det ein positiv eller negativ samanheng mellom dei to variablane? Er samanhengen lineær? (Mpg er miles per gallon. Ein høg verdi av mpg tilseier derfor eit lågt drivstofforbruk.)

```
df <- data.frame(mtcars)
```

```
plot(df$hp, df$mpg, xlab = "Hestekrefter", ylab = "Drivstofforbruk", main = "Drivstofforbruk mot hestekrefter")
```

## Drivstoffbruk mot hestekrefter



Sammenhengen mellom variablene ser ut til å være negativ. Kan kanskje være lineær, ser kanskje litt mer ut som  $f(x) = \frac{1}{x}$  eller  $f(x) = -\ln(x)$

b)

Bruk OLS til å estimere  $b_0$  og  $b_1$  i likninga:

$$mpg = \beta_0 + \beta_1 hp + \epsilon$$

Bruk 'predict' til å finne dei predikerte verdiane av 'mpg' og plot desse som ei linje i spredningsdiagrammet frå a).

```
library(fixest)

mod1 <- feols(mpg ~ hp, data = df)

etable(mod1, tex = T, se.below = F)
```

Dependent Variable:	mpg
Model:	(1)
<i>Variables</i>	
Constant	30.10*** (1.634)
hp	-0.0682*** (0.0101)
<i>Fit statistics</i>	
Observations	32
R <sup>2</sup>	0.60244
Adjusted R <sup>2</sup>	0.58919
<i>IID standard-errors in parentheses</i>	
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>	

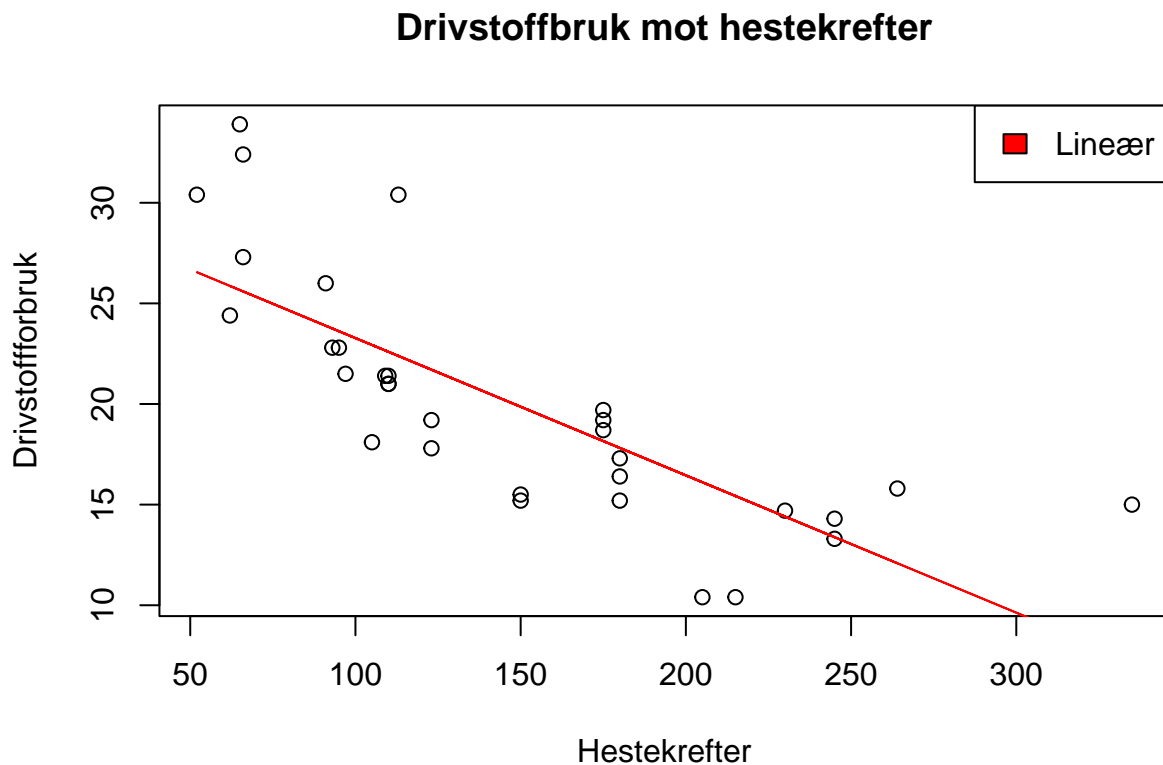
Bruke predict:

```

preds <- data.frame(x_val = df$hp, lin = predict(mod1))

plot(df$hp, df$mpg, xlab = "Hestekrefter", ylab = "Drivstoffforbruk", main = "Drivstoffforbruk mot hestekrefter",
     lines(preds$x_val, preds$lin, col = "red")
     legend("topright", legend = "Lineær", fill = "red"))

```





c)

Bruk OLS til å estimere  $b_0, b_1$  og  $b_2$  likninga

$$mpg = \beta_0 + \beta_1 hp + \beta_2 hp^2 + \epsilon$$

Bruk 'predict' til å finne dei predikerte verdiane av 'mpg' og plot desse som ei linje i spredningsdiagrammet frå a). 4

```
mod2 <- feols(mpg ~ hp + hp^2, data = df)

etable(mod2, tex = T, se.below = F)
```

Dependent Variable:	mpg
Model:	(1)
<i>Variables</i>	
Constant	40.41*** (2.741)
hp	-0.2133*** (0.0349)
hp square	0.0004*** ( $9.84 \times 10^{-5}$ )
<i>Fit statistics</i>	
Observations	32
R <sup>2</sup>	0.75611
Adjusted R <sup>2</sup>	0.73929
<i>IID standard-errors in parentheses</i>	
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>	

```
pred2 <- predict(mod2)

preds2 <- data.frame(x = df$hp, y = pred2)

preds2 <- preds2 |>
  arrange(x)

plot(df$hp, df$mpg, xlab = "Hestekrefter", ylab = "Drivstoffforbruk", main = "Drivstoffbruk mot hestekrefter",
     lines(preds$x_val, preds$lin, col = "red")
     lines(preds2$x, preds2$y, col = "blue")
     legend("topright", legend = c("lineær", "kvadrert"), fill = c("red", "blue"))
```

## Drivstoffbruk mot hestekrefter

