Forelesning 15 — MET1190

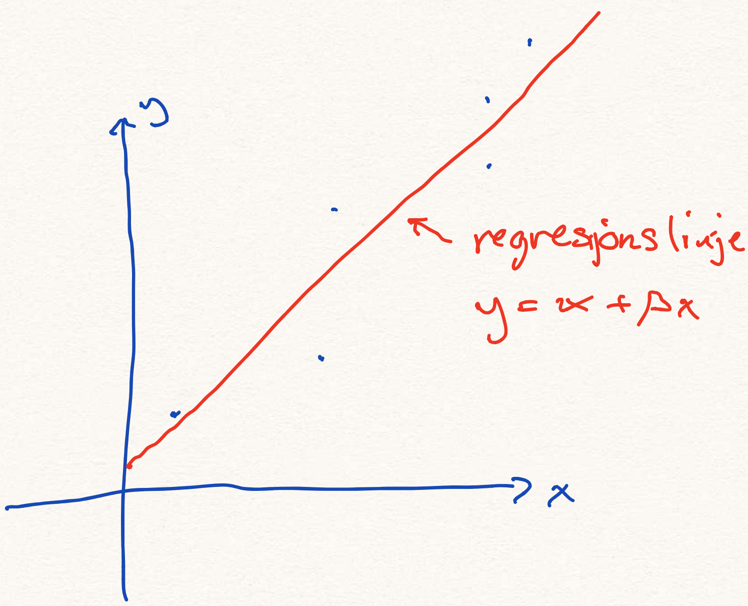
Plan:

1. Lineær regresjon: Hypotesetest
2. Lineær regresjon: Konfidensintervall

Repetisjon:

X	Y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

datasett
2 variable
n observasjoner



Modell: $Y = \alpha + \beta X + \varepsilon$

feilleddet

$\varepsilon \sim N(0, \sigma^2)$

Estimatorer: $\boxed{\hat{\beta} = r \cdot \frac{s_Y}{s_X} \qquad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}}$

Egenskaper til estimatorene:

$E(\hat{\alpha}) = \alpha$, $E(\hat{\beta}) = \beta$ forventningsrette estimatorer

$$SE(\hat{\alpha})^2 = Var(\hat{\alpha}) = \frac{\sigma^2 \cdot \sum x_i^2}{n \cdot (\sum x_i^2 - n\bar{x}^2)}$$

$$SE(\hat{\beta})^2 = Var(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^2 - n\bar{x}^2}$$

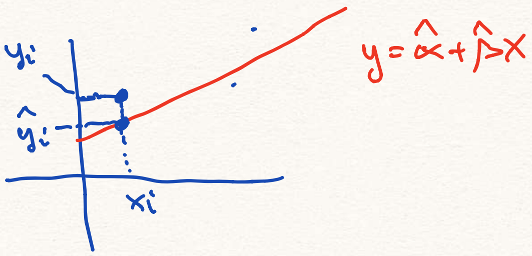
Husk: $\varepsilon \sim N(0, \sigma^2)$

Estimator for σ^2:

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{n-2} SS_E$$

y-verdien i datasettet

$\hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i$



Formler: (1) $\boxed{SS_T = SS_E + SS_R}$ $\Rightarrow SS_E = SS_T - SS_R$

$SS_T = \sum (y_i - \bar{y})^2$, $SS_E = \sum (y_i - \hat{y}_i)^2$, $SS_R = \sum (\hat{y}_i - \bar{y})^2$

(2) $\boxed{\frac{SS_R}{SS_T} = r^2}$ $\Rightarrow SS_R = r^2 \cdot SS_T$

$\Rightarrow SS_E = SS_T - r^2 \cdot SS_T$

$SS_E = SS_T \cdot (1 - r^2)$

r^2: andelen av variasjonen i Y som er forklart av regresjonen

$$SS_E = SS_T - SS_R = SS_T - r^2 \cdot SS_T$$

$$= SS_T (1-r^2)$$

$$= (n-1) s_y^2 (1-r^2)$$

$$s_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

$$= \frac{1}{n-1} SS_T$$

$$\sigma^2 \sim S^2 = \frac{1}{n-2} SS_E$$

$$\boxed{S^2 = \frac{n-1}{n-2} s_y^2 (1-r^2)}$$

Estimat for $SE(\hat{\alpha})$, $SE(\hat{\beta})$:

$$SE(\hat{\beta})^2 = \frac{\sigma^2}{\sum x_i^2 - n\bar{x}^2} = \frac{\sigma^2}{(n-1) \cdot s_x^2}$$

$$\sim \frac{\frac{\cancel{n-1}}{n-2} s_y^2 (1-r^2)}{\cancel{(n-1)} s_x^2} = \boxed{\frac{1}{n-2} \cdot \frac{s_y^2}{s_x^2} (1-r^2)}$$

$$SE(\hat{\alpha})^2 = \frac{\sigma^2 \sum x_i^2}{n \cdot (\sum x_i^2 - n\bar{x}^2)} \sim \frac{S^2 \cdot \sum x_i^2}{n \cdot (\sum x_i^2 - n\bar{x}^2)}$$

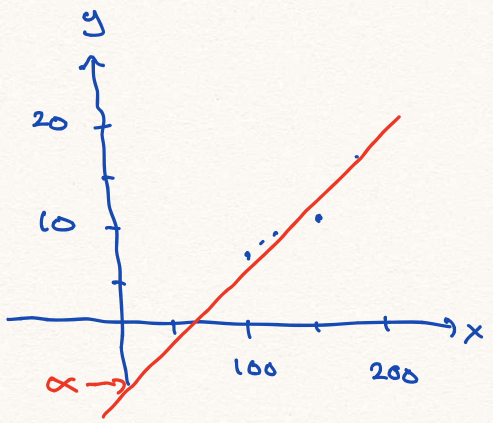
① Hypotesetest for β

Eks: Boligpris

x: ant. kvm. areal

y: pris

x	y
109	9.5
181	15.9
112	9.39
102	8.75
158	11.5



$y = \alpha + \beta x$

α: "pris ved 0 kvm"

β: "pris for en ekstra kvm."

Hypotesetest:

$H_0: \beta = 0$

$H_1: \beta \neq 0$

"det er sammenheng mellom x (areal i kvm) og y (pris)"

$$\hat{\beta} = r \cdot \frac{s_y}{s_x} \approx 0.079$$

Kalk: $\boxed{m}$

Det koster ca 79.000 kr for en ekstra kvm

Hypotesetest: (for β)

Vi undersøker om det er sammenheng mellom ant. kum. og pris.

$H_0: \quad \beta = 0 = \beta_0$

$H_1: \quad \beta \neq 0 = \beta_0$

Konfidensnivå: $\alpha = 0.05$

Testobservator:

$$T = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} = \frac{\hat{\beta}}{SE(\hat{\beta})} \qquad \beta_0 = 0$$

t-fordelt med $n-2$ frihetsgr. når H_0 er sann.

T-test for μ:

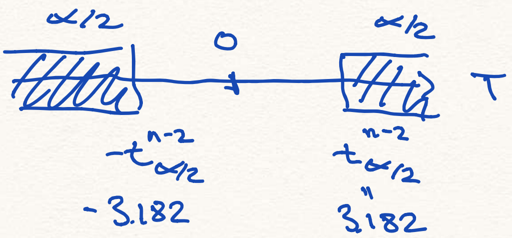
$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$SE(\bar{x})$

Forkastningsområde:

tosidig

$$|T| > t^{n-2}_{\alpha/2}$$



$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

$n = 5 \Rightarrow n-2 = 3$

$$t^{n-2}_{\alpha/2} = t^{3}_{0.025} = \underline{3.182}$$

Forkastningsområde: $|T| > 3.182$

Realisert t-verdi:

$$T = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

$$\hat{\beta} = \underline{0.079}$$

$$SE(\hat{\beta})^2 = \frac{1}{n-2} \frac{S_y^2}{S_x^2} \cdot (1-r^2)$$

$\boxed{r}$

$$r \simeq 0.946$$

$$1-r^2 \simeq 0.105$$

$\boxed{S_y}$

$\boxed{S_x}$

$$\frac{S_y^2}{S_x^2} \cdot (1-r^2) \simeq 0.000731$$

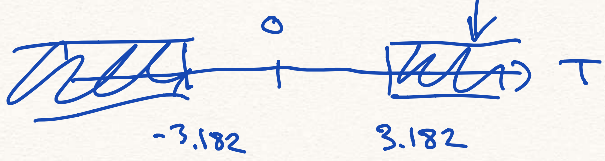
$$SE(\hat{\beta})^2 \simeq 0.000244$$

$\boxed{\sqrt{x}}$

$$SE(\hat{\beta}) \simeq \underline{0.0156}$$

$$\parallel$$

$$T = \frac{0.079}{0.0156} \simeq \underline{5.06}$$



$T > 3.182 \Rightarrow$ Vi forkaster H_0

Det er sammenheng mellom areal og pris.

2 Konfidensintervall for β

Eks: samme som før

$$\hat{\beta} = 0.079$$

95% Konfidensintervall for β:

$\alpha = 0.05$ (konfidensnivå)

Formel: $\hat{\beta} \pm t_{\alpha/2}^{n-2} \cdot SE(\hat{\beta})$

$$= 0.079 \pm 3.182 \cdot 0.0156$$

$$= 0.079 \pm 0.050$$

Konfidensintervall: $(0.029, 0.129)$

Vi kan være 95% sikre på at prisen for en ekstra kvm ligger mellom 29.000 kr og 129.000 kr