ST301, 1995, nr 1.

a) Y = Fuel a_{ij}

1) ? ? x_1 = weight - linearly ?

2) increase with x_2 = disp (but increase is lower down for large x_2 ?)

Can not find table of correlations ?

3) There may well be differences in fuel among different types of cars - as well as in x_1 and x_2.

b) There are 5 estimates for type of car since one car type is a reference value. Estimates differences compared to this car type.

The third column $t\text{-ratio} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{n-p-1}$ under $H_0: \beta_j = 0$

$$p\text{-value} = P\left(|t_{n-p-1}| > \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}\right)$$

c) R^2 increases since new variable $Disp^2 = x_2^2$ is included.

$$R^2 = 1 - \frac{RSS}{TSS} = \text{explained part of variation.}$$

$$= \frac{MSS}{TSS}$$

Since p-value of x_2^2 equals $0.4499 > 0.05$ this variable is not significant - and can be excluded from model (also R^2 hardly changes)

d) Cross validated R^2

$$R^2_{cross} = 1 - \frac{\sum (y_i - \hat{y}_i^{(-i)})^2}{\sum (y_i - \bar{y})^2} \, ,$$

$\hat{y}_i^{(-i)}$ pred. of μ_i when leaving y_i out.

R^2_{cross} will not necc. increase with p] More reasonable than R^2

Prediktion of y_i is done without y_i] at least for model selection.

Would choose model with the largest R^2_{cross}.

Disp + Type $\quad R^2_{cross} = 0.8097$

e) Plott $(\hat{y}_i, r_i)$ No apparent trend, linearity OK? Maybe variation decreases as $\hat{y}_i$ increases. Some slight heteroscedaicity – prob. not important.

Plot (x_{1i}, r_i) No apparent trend.] Maybe CPR $(x_{ji}, \hat{\beta}_j x_{ji} + r_i)$ better?

(x_{2i}, r_i) — " —

Plot (i, r_i) No app. trend. – Not time data. But no indication of dependency.] Piecame in Part II of STK4900

f) $x_{1k} = 2745$

$x_2 = 125$

Compact

$$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\gamma}_1$$

$$= \cancel{3.8871 + 3.[illegible] \cdot 10^{-5} \cdot 2745 +}$$

$$= 3.9181 + 4.164 \cdot 10^{-5} \cdot 2745 + 0.0075941 \cdot 125 - 1.477 = 4.533$$

For å beregne $Var(\hat{\mu})$ må vi også vite $Cov(\hat{\beta}_l, \hat{\beta}_k)$.

ST301, 2000, nr 1 (a - e)

y_i = price suggestion

$x_{1i}, \ldots, x_{7i}$

a) x_{1i} = area, x_{2i} = rooms

$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i \Rightarrow \beta_1$ sign $\neq 0$

$y_i = \beta_0 + \beta_2 x_{2i} + \varepsilon_i \Rightarrow \beta_2$ — " —

but $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ gives only β_{12} sign.

x_{1i} and x_{2i} pos. correlated, conveys same information.

b) $\hat{\beta}_0 = 524$ = est. price for 0 km², 0 room etc. appartment!

$\hat{\beta}_1 = 20.2$, price increases 20000 with increase 1 m²

$\hat{\beta}_2 = 8$, 8000 one room

$\hat{\beta}_3 \approx 90$, App. with garage 90.000 more

$\hat{\beta}_4 = 115$ balcony 116.000 more

$\hat{\beta}_5 = -6.15$, Price goes down 150.000 with rent up 1000 kr.

~~(...)~~

$\hat{\beta}_6 = -106$. Moving one km east reduces price 100.000

$\hat{\beta}_7 = \cdot 7.25$ Moving north then reduce 7050

c) Sign. effect. Size (m^2)
Balcony
Rent
East-west.

Advantage: Simpler model when removing covariates
But these could be reduced to much,
For instance it could make sense including garage.

d) Hist. & QQ plot $r_i = y_i - \hat{y}_i$ Normal assumption
outliers.

Plot $(\hat{y}_i, r_i)$ $\left\{ \begin{array}{l} \text{Overall linear model} \\ \text{Constant variance} \end{array} \right.$

Plott (x_{ji}, r_i) Non linearity
(Better ?, CPR $(x_{ji}, \hat{\beta}_j x_{ji} + r_i)$

e) Cross val. $R^2_{cross} = 0.873$ Full
$R^2_{cross} = 0.883$ Reduced] Reduced is better.