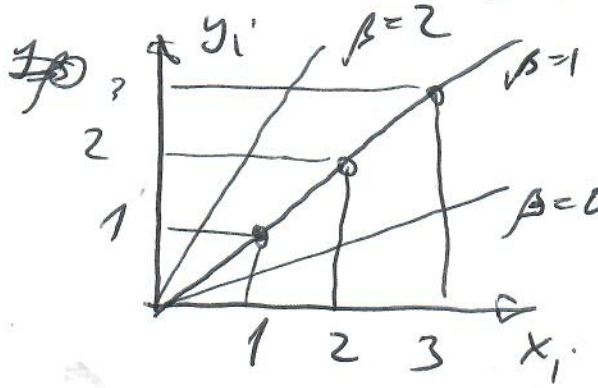
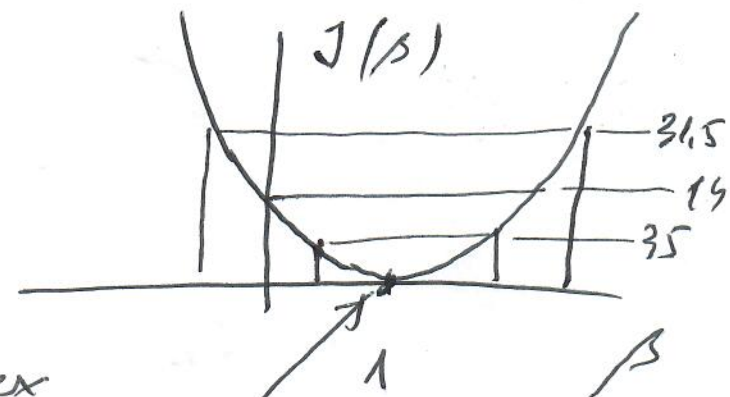
Theoretical question (Exam 2)

(1). $y_i = \beta x_i + \varepsilon_i \Rightarrow \boxed{J(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 \text{ (squared sum of residuals)}}$

$$J(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i) = (1-\beta\cdot 1)^2 + (2-2\beta)^2 + (3-3\beta)^2$$

$\beta = -0.5$	$J(\beta) = 31.5$
$\beta = 0$	$J(\beta) = 14$
$\beta = 0.5$	$J(\beta) = 3.5$
$\beta = 1$	$J(\beta) = 0$
$\beta = 1.5$	$J(\beta) = 3.5$
$\beta = 2$	$J(\beta) = 14$
$\beta = 2.5$	$J(\beta) = 31.5$





The slope of $J(\beta)$ is convex.

The minimum satisfies $J'(\beta) = 0$ $\min_{\beta} J(\beta) = 0$

(the slope of the cost function $J(\beta)' = 0$).

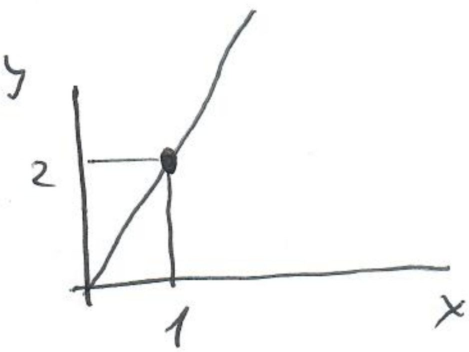
(2). $y_i = \beta x_i + \varepsilon_i$

Least square problem $\min\limits_{\beta} \sum\limits_{i=1}^{n} (y_i - \beta x_i)^2 = J(\beta)$

a) 1 point, e.g. (1,2)

$$\min_{\beta} \sum_{i=1}^{n=1} (2 - 1\cdot\beta)^2 \Rightarrow \boxed{\beta = 2}$$

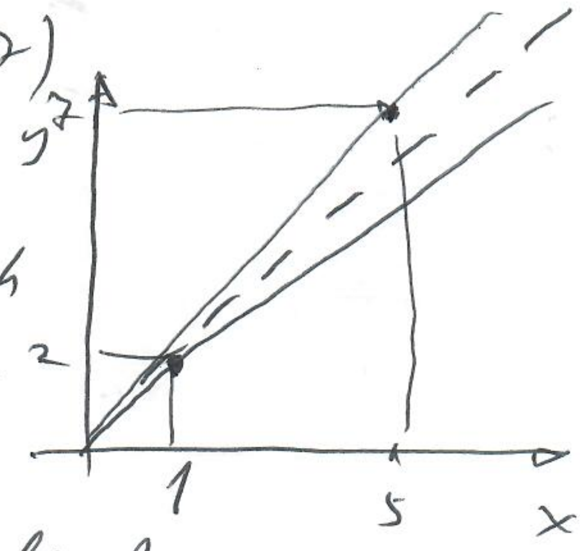
$J(\beta) = 0$



We pass the line with one coefficient through one point.

b) 2 points, e.g. (1,2), (5,7)

We cannot pass the line with one coefficient β through 2 different points. We must choose something in between (dashed line). This is done by least squares



$$\min_{\beta} (y_1 - \beta x_1)^2 + (y_2 - \beta x_2)^2 = (2 - \beta\cdot 1)^2 + (7 - 5\beta)^2 = J(\beta)$$

$$\frac{\partial J(\beta)}{\partial \beta} = 0 \ ? \Rightarrow$$

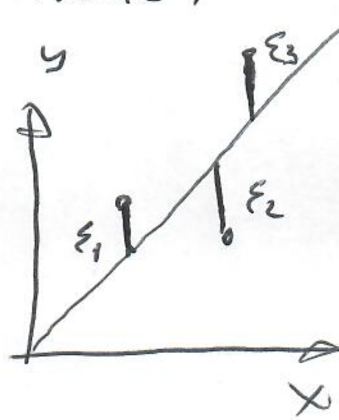
$J(\beta) = 4 - 4\beta + \beta^2 + 49 - 70\beta + 25\beta^2 =$ ~~[illegible]~~

$= 53 - 74\beta + 26\beta^2$

$\frac{\partial J(\beta)}{\partial \beta} = -74 + 52\beta = 0 \qquad \beta = \frac{74}{52} = \frac{37}{26}$

(c) 3 points - we get more simple derivation

$\min_{\beta} \sum_{i=1}^{3} (y_i - \beta x_i)^2 = J(\beta)$



$\frac{\partial J}{\partial \beta} = \sum_{i=1}^{3} 2(y_i - \beta x_i)(-x_i) = 0$

$\sum_{i=1}^{3} y_i x_i = \beta \sum_{i=1}^{n} x_i^2 \Rightarrow \beta = \frac{\sum_{i=1}^{3} y_i x_i}{\sum_{i=1}^{3} x_i^2}$

(d) the formula we derived $\beta = \frac{\sum y_i x_i}{\sum x_i^2}$ works for any number of points 1, 2, 3, or n.

For example, 1 point $x = 1, y = 2 \Rightarrow \beta = \frac{2 \cdot 1}{1 \cdot 1} = 2$

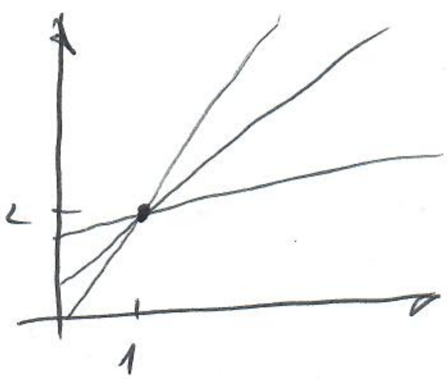
2 points $(1,2), (5,7) \Rightarrow \beta = \frac{1 \cdot 2 + 5 \cdot 7}{1 \cdot 1 + 5 \cdot 5} = \frac{37}{26}$

....

3. $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ -4

a) 1 data points $\Rightarrow$ we can draw infinitely many lines through this point

e.g. $(1, 2)$



$2 = \beta_0 + \beta_1 \cdot 1$, so we can use $\beta_0 = 0$ $\beta_1 = 2$

$\beta_0 = 1$ $\beta_1 = 1$

$\beta_2 = 2$ $\beta_1 = 0$

, ...

b) 2 data points $(1, 2), (5, 7)$

We can solve $y_i = \beta_0 + \beta_1 x_i$

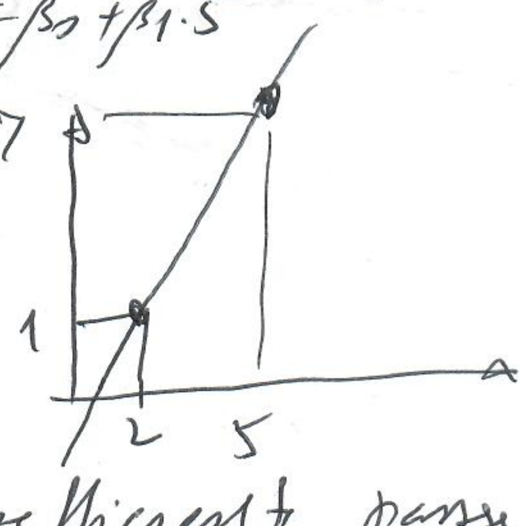
$2 = \beta_0 + \beta_1 \cdot 1 \Rightarrow \beta_0 = 2 - \beta_1 \Rightarrow 7 = 2 - \beta_1 + \beta_1 \cdot 5$

$7 = \beta_0 + \beta_1 \cdot 5$

$\Rightarrow 5 = 4\beta_1$

$\boxed{\beta_1 = \frac{5}{4}}$

$\beta_0 = 2 - \frac{5}{4} = \frac{3}{4}$



2 coefficients passe through 2 points exactly!

−5

c) 3 data points, we derive OLS since we cannot pass through 3 points.

$$J(\beta_0,\beta_1) = \sum_{i=1}^{n} (y_i - \beta_1 x_i - \beta_0)^2$$

$$\frac{\partial J}{\partial \beta_0} = \sum_{i=1}^{n} 2(y_i - \beta_1 x_i - \beta_0)(-1) = 0$$

$$\sum y_i - \beta_1 \sum x_i - \beta_0 \cdot n = 0 \Rightarrow$$

$$\boxed{\beta_0 = \bar{y} - \beta_1 \bar{x}}$$

$$\frac{\partial J}{\partial \beta_1} = \sum_{i=1}^{n} 2(y_i - \beta_1 x_i - \beta_0)(-x_i) = \sum y_i x_i - \beta_1 \sum x_i^2 - \beta_0 \sum x_i = 0$$

$$\sum y_i x_i - \beta_1 \sum x_i^2 - \sum x_i (\bar{y} - \beta_1 \bar{x}) = 0$$

$$\frac{\sum y_i x_i}{n} - \frac{\bar{y} \sum x_i}{n} = \beta_1 \frac{\sum x_i^2}{n} - \beta_1 \frac{\sum x_i \bar{x}}{n}$$

$$\boxed{\beta_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2}}$$

So we can find β_1 and later β_0

This formula works for 3 points and for any other larger number of points as well as for 2 points.

4.) 5) Consider a regression line

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \dots - \beta_k x_{ki}\right)^2 = J(\beta_0, \beta_1, \dots \beta_k)$$

$$\frac{\partial J}{\partial \beta_0} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_{1i} - \dots - \beta_k x_{ki}\right)(-1) = 0$$

$$\frac{\partial J}{\partial \beta_1} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_{1i} - \dots - \beta_k x_{ki}\right)(-x_{1i}) = 0$$

$$\frac{\partial J}{\partial \beta_2} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_{1i} - \dots - \beta_k x_{ki}\right)(-x_{2i}) = 0$$

$$\dots\dots\dots\dots$$

$$\frac{\partial J}{\partial \beta_k} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1 x_{1i} - \dots - \beta_k x_{ki}\right)(-x_{ki}) = 0$$

We have a linear system of equations with $(k+1)$ equations and $(k+1)$ unknowns $(\beta_0, \dots \beta_k)$. We can solve it as in 3) but it will be long and tedious. We will not do it because such solution follows trivially with linear algebra!

(6) $y = X\beta + \varepsilon$

Least square problem:

$$\min_{\beta} \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) = y'y - \beta'X'y - y'X\beta + \beta'X'X\beta = J(\beta)$$

$$\frac{\partial J}{\partial \beta} = \frac{\partial}{\partial \beta}\left(y'y - \beta'X'y - y'X\beta + \beta'X'X\beta\right) =$$

$$= -2X'y + 2X'X\beta = 0 \Rightarrow \boxed{(X'X)\beta = X'y}$$

$$\Rightarrow \underbrace{(X'X)^{-1}(X'X)}_{I}\beta = (X'X)^{-1}X'y \Rightarrow \beta = (X'X)^{-1}X'y$$

Please see details of β derivation in the handout I provided.

$$\begin{pmatrix} 1 & 1 \\ 4 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 16 \end{pmatrix}$$

(7). a) $\beta = (X'X)^{-1}(X'y) = \left(\begin{pmatrix} 1 & 1 \\ 4 & 6 \end{pmatrix}\begin{pmatrix} 1 & 4 \\ 1 & 6 \end{pmatrix}\right)^{-1} \cancel{\begin{pmatrix} 1 & 4 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix}} =$

$$= \begin{pmatrix} 2 & 10 \\ 10 & 52 \end{pmatrix}^{-1}\begin{pmatrix} 3 \\ 16 \end{pmatrix} = \frac{\begin{pmatrix} 52 & -10 \\ -10 & 2 \end{pmatrix}}{\det = 104 - 100}\begin{pmatrix} 3 \\ 16 \end{pmatrix} = \begin{pmatrix} \frac{52}{4} & -\frac{10}{4} \\ -\frac{10}{4} & \frac{2}{4} \end{pmatrix}\begin{pmatrix} 3 \\ 16 \end{pmatrix} =$$

$$= \begin{pmatrix} 13 & -2{,}5 \\ -2{,}5 & 0{,}5 \end{pmatrix} \begin{pmatrix} 3 \\ 16 \end{pmatrix} = \begin{pmatrix} -1 \\ 0{,}5 \end{pmatrix}$$

b)

$$\beta = (X'X)^{-1}X'y = \left(\begin{pmatrix} 1 & 1 & 1 \\ 4 & 6 & 8 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 1 & 6 \\ 1 & 8 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & 1 \\ 4 & 6 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 18 \\ 18 & 116 \end{pmatrix}^{-1} \begin{pmatrix} 6 \\ 40 \end{pmatrix} = \frac{\begin{pmatrix} 116 & -18 \\ -18 & 3 \end{pmatrix}}{\det} \begin{pmatrix} 6 \\ 40 \end{pmatrix} =$$

$$\det = 116 \cdot 3 - 18^2 = 24$$

$$= \begin{pmatrix} \frac{116}{24} & -\frac{18}{24} \\ -\frac{18}{24} & \frac{3}{24} \end{pmatrix} \begin{pmatrix} 6 \\ 40 \end{pmatrix} = \begin{pmatrix} \frac{116}{4} - \frac{18 \cdot 10}{6} \\ -\frac{18}{4} + \frac{3 \cdot 10}{6} \end{pmatrix} = \begin{pmatrix} -1 \\ 0{,}5 \end{pmatrix}$$

c)

$$\beta = (X'X)^{-1}X'y = \left(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 4 & 6 & 8 & 10 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 1 & 6 \\ 1 & 8 \\ 1 & 10 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 4 & 6 & 8 & 10 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} -1 \\ 0{,}5 \end{pmatrix}$$ (please, check it by doing all multiplications and inversions like in a, b)

8) $y_i = \beta_1 x_i + \varepsilon_i \quad , \quad \varepsilon_i \sim N(0, \sigma^2)$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

a) $$E[\hat{\beta}_1] = E\left[\frac{\sum_{i=1}^{n} x_i(\beta_1 x_i + \varepsilon_i)}{\sum_{i=1}^{n} x_i^2}\right] = E\left[\frac{\beta_1 \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2}\right] =$$

$$= E[\beta_1] + E\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) = \beta_1 + \cancel{\frac{\sum x_i \overset{0}{E[\varepsilon_i]}}{\sum x_i^2}} = \beta_1$$

(unbiased)

b) $$var(\beta_1) = var\left(\frac{\sum x_i y_i}{\sum x_i^2}\right) = var\left(\frac{\sum x_i(\beta_1 x_i + \varepsilon_i)}{\sum x_i^2}\right) =$$

$$= \frac{1}{(\sum x_i^2)^2} var(\beta_1 \sum x_i^2 + \sum x_i \varepsilon_i) = \frac{1}{(\sum x_i^2)^2} var(\sum x_i \varepsilon_i) =$$

$$= \frac{\sum x_i^2 \, var(\varepsilon_i)}{(\sum x_i^2)^2} = \frac{\sigma_\varepsilon^2 \sum x_i^2}{(\sum x_i^2)^2} = \frac{\sigma_\varepsilon^2}{\sum x_i^2}$$

c) $\beta \sim N\left(0, \frac{\sigma^2}{\sum x_i^2}\right)$ - assymptotic distribution of β