Estimators for β and α are given by

$$\hat{\beta} = B = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \text{og} \quad \hat{\alpha} = A = \bar{Y} - B\bar{x}$$

$$E[B] = \frac{\sum_{i=1}^{n}(x_i-\bar{x})E[Y_i]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(\alpha+\beta x_i)}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \beta\frac{\sum_{i=1}^{n}(x_i-\bar{x})x_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$= \beta\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$E[A] = E[\bar{Y}] - \beta\bar{x} = \sum_{i=1}^{n}\frac{E[Y_i]}{n} - \beta\bar{x} = \sum_{i=1}^{n}\frac{(\alpha+\beta x_i)}{n} - \beta\bar{x}$$

$$= \alpha + \beta\bar{x} - \beta\bar{x} = \alpha.$$

$$Var[B] \overset{\text{p.g.a uavh.}}{=} \sum_{i=1}^{n}\frac{Var[(x_i-\bar{x})Y_i]}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sigma^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$Var[A] = Var[\overset{\text{uavh}}{\bar{Y} - B\bar{x}}] = \frac{\sigma^2}{n} + \frac{\bar{x}^2\cdot\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$$

Estimat for σ^2, $s^2 = \frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}$

$$S^2 = \sum_{i=1}^{n}\frac{(Y_i-\hat{Y}_i)^2}{n-2}, \quad \text{der } \hat{Y}_i = A + Bx_i = \hat{\alpha} + \hat{\beta}x_i$$

Konfidensintervall for koeffisientane

Modell : $Y = \alpha + \beta x + \varepsilon \sim \begin{cases} N(0,\sigma^2) \\ \text{uavh.} \end{cases}$

$\Rightarrow \quad B = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ og $A = \bar{Y} - B\bar{x}$ og vil vere normalfordelte som lineærkombinasjonar av uavh. og normalfordelte variablar.

La $S_{xx} = \sum_{i=1}^{n}(x_i-\bar{x})^2$

$$Z = \frac{B-\beta}{\frac{\sigma}{\sqrt{S_{xx}}}} \sim N(0,1)$$

Vidare vil $(n-2)\dfrac{S^2}{\sigma^2}$ vere χ^2-fordelt med $n-2$ fridomsgrader og $T = \dfrac{\frac{B-\beta}{\frac{\sigma}{\sqrt{S_{xx}}}}}{\frac{S}{\sigma}} = \dfrac{B-\beta}{\frac{S}{\sqrt{S_{xx}}}}$ er t-fordelt

med $n-2$ fridomsgrader

Vi får $P\left(-t_{\frac{\alpha}{2},n-2} < \dfrac{B-\beta}{\frac{S}{\sqrt{S_{xx}}}} < t_{\frac{\alpha}{2},n-2}\right) = 1-\alpha$

$\Leftrightarrow \quad P\left(B - t_{\frac{\alpha}{2},n-2}\dfrac{S}{\sqrt{S_{xx}}} < \beta < B + t_{\frac{\alpha}{2},n-2}\dfrac{S}{\sqrt{S_{xx}}}\right) = 1-\alpha$

som gjev $100(1-\alpha)\%$ konfidensintervallet for β:

$$\left(b - t_{\frac{\alpha}{2},n-2}\frac{s}{\sqrt{S_{xx}}},\ b + t_{\frac{\alpha}{2},n-2}\frac{s}{\sqrt{S_{xx}}}\right)$$

$$\mathrm{Var}[A] = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right] = \sigma^2\left[\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 + n\bar{x}^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\right] = \sigma^2\left[\frac{\sum_{i=1}^{n} x_i^2}{n S_{xx}}\right]$$

og eit $100(1-\alpha)\%$ konfidensintervall for α blir.

$$\left(a - t_{\frac{\alpha}{2},n-2}\, s\sqrt{\frac{\sum x_i^2}{n S_{xx}}},\quad a + t_{\frac{\alpha}{2},n-2}\, s\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n S_{xx}}}\right)$$

<u>Testar</u>

$$H_0: \beta = \beta_0,\quad H_1: \beta \underset{<}{\overset{>}{\neq}} \beta_0 \qquad H_0: \alpha = \alpha_0,\quad H_1: \alpha \underset{<}{\overset{>}{\neq}} \alpha_0$$

Teststatistikkar:

For test på β: $Z = \dfrac{B - \beta_0}{\frac{\sigma}{\sqrt{S_{xx}}}}$ evt. $T = \dfrac{B - \beta_0}{\frac{S}{\sqrt{S_{xx}}}} \sim t_{n-2}$

For test på α: $Z = \dfrac{A - \alpha_0}{\sigma\sqrt{\frac{\sum x_i^2}{n S_{xx}}}}$ evt. $T = \dfrac{A - \alpha_0}{S\sqrt{\frac{\sum x_i^2}{n S_{xx}}}} \sim t_{n-2}$

<u>Oppsplitting av variasjon</u> i dataane

Totalvariasjonen $SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \cancel{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2} = \sum_{i=1}^{n}(y_i - \hat{y}_i + \underbrace{\hat{y}_i}_{a+bx_i} - \bar{y})^2$$

$$= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + 2\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$$

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) = \underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)(a+bx_i)}_{=0} - \underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)}_{=0}\bar{y}$$

1. normallikning

1 og 2 normallikning

Dermed får vi:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i-\bar{y})$$

eller $SS_T = SS_E + SS_R$

Total variasjon =

Ein god modell bør ha $SS_T \approx SS_R$

$\frac{SS_R}{SS_T} = R^2$, (den multiple determinasjonskoeffisienten)

og brukast til å forklje kor mykje av den totale variasjonen i dataane som er forklart av modellen.

Estimering av forventing i x_0

$$Y_i = \alpha + \beta x_i \sim \begin{cases} N(0, \sigma^2) \\ \text{uavh} \end{cases}$$

$$\Rightarrow E[Y_i] = \alpha + \beta x_i = \mu_{Y_i}$$

$$\hat{\mu}_{Y_i} = A + B x_i, \quad E[\hat{\mu}_{Y_i}] = \alpha + \beta x_i$$

$$\hat{\mu}_{Y|x=x_0} = A + B x_0 = \bar{Y} - B\bar{x} + B x_0 = \bar{Y} + B(x_0 - \bar{x})$$

$$Var[\hat{\mu}_{Y_0}] \overset{\text{p.g.a uavh}}{=} Var(\bar{Y}) + Var[B](x_0 - \bar{x})^2 = \frac{\sigma^2}{n} + \frac{\sigma^2 (x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Difor er
$$\frac{\hat{\mu}_{Y_0} - (\alpha + \beta x_0)}{S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}} \sim t\text{-fordelt med } n-2 \text{ fridomsgrader}$$

s.a. eit $100(1-\alpha)\%$ konfidensintervall for μ_{Y_0} er gitt ved

$$\hat{\mu}_{Y_0} - t_{\frac{\alpha}{2}, n-2}\, S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} < \mu_{Y_0} < \hat{\mu}_{Y_0} + t_{\frac{\alpha}{2}, n-2}\, S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

Predikering av ny verdi i x_0, y_0

Estimator
$$\hat{Y}_0 = A + B x_0$$

$$E[Y_0 - \hat{Y}_0] = \alpha + \beta x_0 - \alpha - \beta x_0 = 0$$

$$Var[Y_0 - \hat{Y}_0] \overset{\text{p.g.a. uavh}}{=} Var[Y_0] + Var[\hat{Y}_0] = \sigma^2 + \frac{\sigma^2}{n} + \frac{\sigma^2 (x_0 - \bar{x})^2}{S_{xx}}$$

Difor er
$$\frac{Y_0 - \hat{Y}_0}{S\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}} \sim t\text{-fordelt med } n-2 \text{ fridomsgrader.}$$

Prediksjonsintervall for Y_0: $\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-2}\, S\sqrt{\frac{1}{n} + 1 + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$

Residuala.

$e_i = y_i - \hat{y}_i$, $i = 1, 2, \ldots, n$ blir kalla residuala. Desse kan fortelje om den tilpassa modellen er god. Det er vanleg å plotte residuala mot $\hat{y}_i$ for å sjekke konstant varians, evt om det er avvikande observasjonar i dataane. Ein bør og plotte dei mot rekkjefølgja som observasjonane er tekne i for å sjekke at det ikkje er mønster (raude) og lage normal-plott for å sjekke fordeling.