

# Sammenligninger af to regressionslinier

Sammenligningen af to regressionsligninger kan deles op i tre dele:

Først undersøges:

1. Om regressionslinierne har samme hældninger,  $\beta_{11}$  og  $\beta_{12}$ , dvs. bliver de to regressionsmodeller påvirket i samme grad af den forklarende variable  $x_i$ .

Man kan derefter undersøge:

2. Om de to regressionslinier har samme afskæringer  $\beta_{01}$  og  $\beta_{02}$
3. Om der er forskel på de (forventede) Y-værdier, når man har korrigeret for  $x$ 's indflydelse ("Alt andet lige beregninger")

## Ad. 1: Test af ens hældning

$$H_0: \beta_{11} = \beta_{12}$$

Før man tester for ens hældning, skal man sikre sig, at standardforudsætninger, er opfyldt. Dvs. er restleddene uafhængige og normalfordelte og er restledsvarianserne ens. De to første forudsætninger antages opfyldt, mens sidste forudsætning skal kontrolleres vha. en statistisk test ( $\sigma_1^2 = \sigma_2^2$ ):

$$H_1: \sigma_1^2 = \sigma_2^2 \text{ - versus - } \bar{H}_1: \sigma_1^2 \neq \sigma_2^2$$

Teststørrelse:

$$V = \frac{s_{\max}^2}{s_{\min}^2} \sim F(n_{\max} - 2, n_{\min} - 2)$$

$$H_1: \sigma_1^2 = \sigma_2^2$$

Godkendes	Forkastes
I (a)	I (b)

Hvis  $H_1$  godkendes så udregn **standard pooled varians** :

$$s^2 = \frac{(n_1 - 2)s_1^2 + (n_2 - 2)s_2^2}{n_1 + n_2 - 4}$$

**l(a):**

Teststørrelse for  $H_0$ :

$$t = \frac{\hat{\beta}_{11} - \hat{\beta}_{12}}{S \cdot \sqrt{\frac{1}{SAK_{x_1}} + \frac{1}{SAK_{x_2}}}} \sim t(n_1 + n_2 - 4)$$

Hvor  $S^2$  er standard pooled varians.

**l(b):**

Teststørrelse for  $H_0$  (Approksimativt):

$$U = \frac{\hat{\beta}_{11} - \hat{\beta}_{12}}{\sqrt{\frac{s_1^2}{SAK_{x_1}} + \frac{s_2^2}{SAK_{x_2}}}} \sim Nf(0,1)$$

Hvis  $H_0$  godkendes (uanset skæbnen ved  $H_1$  testet), så kan den fælles  $\hat{\beta}_1$  bestemmes :

$$\hat{\beta}_1 = \frac{SAK_{x_1}\hat{\beta}_{11} + SAK_{x_2}\hat{\beta}_{12}}{SAK_{x_1} + SAK_{x_2}}$$

Ad. 2: Test af ens afskæring

**$H_2: \beta_{01} = \beta_{02}$**

**$H_1: \sigma_1^2 = \sigma_2^2$**

<b><math>H_0: \beta_{11} = \beta_{12}</math></b>	Godkendes	Forkastes
Godkendes	II(a)	II(c)
Forkastes	II(b)	II(d)

**II(a):**

Teststørrelse for  $H_2$ :

$$t = \frac{\hat{\beta}_{01} - \hat{\beta}_{02}}{S' \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2)^2}{SAK_{x_1} + SAK_{x_2}}}} \sim t(n_1 + n_2 - 3)$$

$$(\hat{\beta}_{01} - \hat{\beta}_{02} = \bar{y}_1 - \bar{y}_2 - \hat{\beta}_1(\bar{x}_1 - \bar{x}_2))$$

I teststørrelsen for  $H_2$  indgår den **superpoolede varians**,  $S'^2$ :

$$S'^2 = \frac{1}{n_1 + n_2 - 3} \sum (y_i - \hat{\beta}_{01} - \hat{\beta}_1 x_i)^2 + \sum (y_i - \hat{\beta}_{02} - \hat{\beta}_1 x_i)^2 \approx \sigma^2$$

hvor  $\hat{\beta}_1$  er det fælles estimat for hældningen .

Hvis  $H_2$  godkendes (dvs., hvis  $H_0$ ,  $H_1$  og  $H_2$  alle godkendes), slå da gruppe 1 og gruppe 2 sammen til et fælles datasæt. Estimer herefter én værdi af  $\beta_0$ ,  $\beta_1$  og  $\sigma^2$  på baggrund af de  $n_1 + n_2$  observationer.

**II(b):**

Teststørrelse for  $H_2$ :

$$t = \frac{\hat{\beta}_{01} - \hat{\beta}_{02}}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{\bar{x}_1^2}{SAK_{X_1}} + \frac{\bar{x}_2^2}{SAK_{X_2}}}} \sim t(n_1 + n_2 - 4)$$

$$(\hat{\beta}_{01} = \bar{y}_1 - \hat{\beta}_{11}\bar{x}_1)$$

$$(\hat{\beta}_{02} = \bar{y}_2 - \hat{\beta}_{12}\bar{x}_2)$$

hvor  $S^2$  er den standard poolede varians.

Hvis  $H_2$  godkendes, så gå videre i analysen med to datasæt (to sæt hældninger  $\beta_{11}$  og  $\beta_{12}$ )

**II(c):**

Dvs. ens hældning  $\beta_1$ , men forskellig varians. To varianser  $S_1^2$  og  $S_2^2$  opretholdes og  $H_2$  testes approksimativt:

$$U = \frac{\hat{\beta}_{01} - \hat{\beta}_{02}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + (\bar{x}_1 - \bar{x}_2)^2 \frac{SAK_{X_1}S_1^2 + SAK_{X_2}S_2^2}{(SAK_{X_1} + SAK_{X_2})^2}}} \sim NF(0,1)$$

$$(\hat{\beta}_{01} - \hat{\beta}_{02} = \bar{y}_1 - \bar{y}_2 - \hat{\beta}_1(\bar{x}_1 - \bar{x}_2))$$

**III(d)**

Dvs. forskellig hældning og forskellig varians.  $H_2$  testes approksimativt:

$$U = \frac{\hat{\beta}_{01} - \hat{\beta}_{02}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} + \frac{\bar{x}_1^2 S_1^2}{SAK_{X_1}} + \frac{\bar{x}_2^2 S_2^2}{SAK_{X_2}}}} \sim NF(0,1)$$

$$(\hat{\beta}_{01} = \bar{y}_1 - \hat{\beta}_{11}\bar{x}_1)$$

$$(\hat{\beta}_{02} = \bar{y}_2 - \hat{\beta}_{12}\bar{x}_2)$$

**Ad. 3: "Alt andet lige" sammenligninger af totale forventede y-værdier**

Det undersøges nu, om der er forskel på de (forventede) Y-værdier, når man har korrigeret for x's indflydelse. Vi tester altså hypotesen:

$$H_3 : \mathbf{E}(y_1 | \mathbf{x} = \mathbf{x}^0) = \mathbf{E}(y_2 | \mathbf{x} = \mathbf{x}^0)$$

hvor  $x^0$  er den fælles x-værdi.

SAS anvender automatisk flg.  $x^0$ :

$$x^0 = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

		H1: $\sigma_1^2 = \sigma_2^2$	
H0: $\beta_{11} = \beta_{12}$	Godkendes	Forkastes	
Godkendes	III(a)	-	
Forkastes	III(b)	-	

### III(a)

Beregn først **forudsagte værdier**:

$$\hat{E}_1 = \hat{\beta}_{01} + \hat{\beta}_1 x^0$$

$$\hat{E}_2 = \hat{\beta}_{02} + \hat{\beta}_1 x^0$$

hvor  $\hat{\beta}_1$  er det fælles estimat for hældningen (se I(b)).

Teststørrelse for  $H_3$ :

$$t = \frac{\hat{E}_1 - \hat{E}_2}{S' \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2)^2}{SAK_{x_1} + SAK_{x_2}}}} \sim t(n_1 + n_2 - 3)$$

$S^2$  er den **superpooled varians** (se II(a)). SAS anvender imidlertid standard pooled varians  $S^2$  (se I(a)) og t-fordeling for dermed  $n_1 + n_2 - 4$  frihedsgrader.

### III(b)

Beregn først **forudsagte værdier**:

$$\hat{E}_1 = \hat{\beta}_{01} + \hat{\beta}_{11} x^0$$

$$\hat{E}_2 = \hat{\beta}_{02} + \hat{\beta}_{12} x^0$$

Teststørrelse for  $H_3$ :

$$t = \frac{\hat{E}_1 - \hat{E}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{(x^0 - \bar{x}_1)^2}{SAK_{x_1}} + \frac{(x^0 - \bar{x}_2)^2}{SAK_{x_2}}}} \sim t(n_1 + n_2 - 4)$$

hvor  $S^2$  er den standard pooled varians