10.9 Test på 2 forholdstal

p_1 = andelen Volvoeigarar i Noreg

p_2 = — " — " — i Sverige

X_1 = talet på Volvoeigarar i utvalet i Noreg

X_2 = — " — " — i Sverige.

$X_1 \sim B(n_1, p_1)$, $X_2 \sim B(n_2, p_2)$.

$\hat{p}_1 = \frac{X_1}{n_1}$, $\hat{p}_2 = \frac{X_2}{n_2}$, $Var\,\hat{p}_1 = \frac{p_1(1-p_1)}{n_1}$, $Var\,\hat{p}_2 = \frac{p_2(1-p_2)}{n_2}$.

$H_0: p_1 = p_2$ $\quad\quad H_1: \begin{cases} p_1 > p_2 \\ p_1 \neq p_2 \\ p_1 < p_2 \end{cases}$

For stor nok n_1 og n_2 er

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \approx N(0,1)$$

Sidan p_1 og p_2 er ukjende, er det vanleg å erstatte p_1 og p_2 med $\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$

og bruke at $$Z^0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx N(0,1)$$

Kap 11 Regresjonsanalyse

Føremål: Finne samanhengar mellom ein responsvariabel og ein eller fleire forklaringsvariable. Regresjonsanalyse er eit modelleringsverktøy.

11.2. Enkel lineær regresjon (lineær i koeffisientane)

Modell: $Y_i = \alpha + \beta x_i + \varepsilon_i \left\{ \begin{array}{l} \text{uavh.} \\ E[\varepsilon_i] = 0, \quad Var[\varepsilon_i] = \sigma^2 \end{array} \right.$

Y_i: respons, avh. variabel

x_i: regresjonsvariabel, uavh. variabel, forklaringsvariabel

x_i kan vere ein transformasjon av andre variable.

t.d. $x_i = \ln(t_i)$, $x_i = \sin(t_i)$ o.s.v.

Observerer $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ som skal oppfylle

$$y_1 = \alpha + \beta x_1 + \varepsilon_1$$
$$y_2 = \alpha + \beta x_2 + \varepsilon_2$$
$$\vdots$$
$$y_n = \alpha + \beta x_n + \varepsilon_n$$

Skal finne estimat, a og b for α, og β.

Minste kvadratsums metode

Finn den a og b som minimerer

$$Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a - b x_i)^2$$

$$\frac{\partial Q}{\partial a} = 0 \iff -2 \sum_{i=1}^{n} (y_i - a - b x_i) = 0$$

$$\frac{\partial Q}{\partial b} = 0 \iff -2\sum_{i=1}^{m}(y_i - a - bx_i)x_i = 0$$

Dette gjev dei 2 normallikningane

$$\sum_{i=1}^{m} a + \sum_{i=1}^{m} bx_i = \sum_{i=1}^{m} y_i \qquad (1)$$

$$\sum_{i=1}^{m} ax_i + \sum_{i=1}^{m} bx_i^2 = \sum_{i=1}^{m} x_i y_i \qquad (2)$$

Løyser vi (1) for a får vi:

$$ma + b\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i \Rightarrow a = \bar{y} - b\bar{x}$$

som innsett i (2) gjev:

$$(\bar{y} - b\bar{x})\sum_{i=1}^{m} x_i + b\sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i$$

$$\Rightarrow b\left(\sum_{i=1}^{m} x_i^2 - m\bar{x}^2\right) = \sum_{i=1}^{m} x_i(y_i - \bar{y})$$

$$\Rightarrow b = \frac{\sum_{i=1}^{m} x_i(y_i - \bar{y})}{\sum_{i=1}^{m} x_i^2 - m\bar{x}^2} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})y_i}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

Her er brukt.
$$\sum_{i=1}^{m} \bar{x}(y_i - \bar{y}) = \bar{x}\sum_{i=1}^{m}(y_i - \bar{y}) = 0$$

$$\sum_{i=1}^{m}(x_i - \bar{x})^2 = \sum_{i=1}^{m}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^{m} x_i^2 - 2m\bar{x}^2 + m\bar{x}^2 = \sum_{i=1}^{m} x_i^2 - m\bar{x}^2$$

$$\sum_{i=1}^{m} \bar{y}(x_i - \bar{x}) = \bar{y}\sum_{i=1}^{m}(x_i - \bar{x}) = 0$$

Estimators for β and α are given by

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

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

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

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

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

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

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

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

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