Formelsamling

Grundläggande statistik för ekonomer

Beskrivande mått:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

$$Cov(x,y) = s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

$$Corr(x,y) = r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

Laspeyres index:

$$P_{0-t}^{L} = \frac{\sum p_t q_0}{\sum p_0 q_0}$$

Paasches index:

$$P_{0-t}^{L} = \frac{\sum p_t q_t}{\sum p_0 q_t}$$

Additionssatsen:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Multiplikationssatsen

$$P(A \cap B) = P(A \mid B)\, P(B) = P(B \mid A)\, P(A)$$

Satsen om total sannolikhet:

$$P(B) = \sum_{i=1}^{k} P(A_i) P(B|A_i)$$

Bayes' Sats:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Väntevärde för en diskret stokastisk variabel X:

$$E(X) = \mu_X = \sum_{\text{alla } x} x \cdot P(x)$$

Varians för en diskret stokastisk variabel X:

$$Var(X) = E(X^2) - \mu_X^2 = \left(\sum_{\text{alla } x} x^2 P(x) \right) - \mu_X^2 = \sum_{\text{alla } x} (x - \mu_X)^2 \cdot P(x)$$

Kovarians mellan två diskreta stokastiska variabler X och Y:

$$Cov(X, Y) = E(XY) - \mu_X \mu_Y = \sum_{\text{alla x}} \sum_{\text{alla y}} xy P(x, y) - \mu_X \mu_Y$$

Räkneregler för väntevärden och varianser (a, b och c är konstanter, X och Y är stokastiska variabler):

$$\begin{aligned}
E(c) &= c \\
E(cX) &= cE(X) \\
E(c+X) &= c+E(X) \\
E(aX+bY) &= aE(X)+bE(Y) \\
Var(c) &= 0 \\
Var(cX) &= c^2Var(X) \\
Var(c+X) &= Var(X) \\
Var(aX+bY) &= a^2Var(X)+b^2Var(Y)+2abCov\,(X,Y)
\end{aligned}$$

Binomialfördelningen:

$$P\,(x) = \binom{n}{x} P^x(1-P)^{n-x} = \frac{n!}{x!\,(n-x)!} P^x(1-P)^{n-x}$$

$$E(X) = nP \qquad Var(X) = nP(1-P)$$

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

$$\text{om} \quad X \sim N(\mu,\sigma)$$

Väntevärde och varians för urvalsmedelvärdet

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

där alla $X_1, X_2, ..., X_n$ är oberoende och har väntevärde μ och varians σ^2:

$$\begin{aligned}
E\left(\overline{X}\right) &= \mu \\
Var\left(\overline{X}\right) &= \frac{\sigma^2}{n}
\end{aligned}$$

Stratifierat urval:

$$\overline{X}_{str} = \sum_{i=1}^{L} W_i \overline{X}_i$$

$$Var(\overline{X}_{str}) = \sum_{i=1}^{L} W_i^2 Var(\overline{X}_i)$$

$$\text{där} \quad W_i = \frac{N_i}{N}$$

Ändlighetskorrektion (OSU utan återläggning, liten population):

$$\frac{N-n}{N-1} = \left(1 - \frac{n}{N}\right)$$

Konfidensintervall för μ:

$n \geq 30$ oavsett om populationen är normalfördelad eller ej	σ^2 känd: $\overline{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ σ^2 okänd: $\overline{x} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$
$n < 30$ populationen är normalfördelad	σ^2 känd: $\overline{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ σ^2 okänd: $\overline{x} \pm t_{v,\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$
$n < 30$ populationen är ej normalfördelad	Konfidensintervall kan inte beräknas

Konfidensintervall för P:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Konfidensintervall för $\mu_x - \mu_y$ (Oberoende stickprov):

$n_x \geq 30$, $n_y \geq 30$ oavsett om populationerna är normalfördelade eller ej	σ_x^2, σ_y^2 känd: σ_x^2, σ_y^2 okänd:	$(\overline{x} - \overline{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ $(\overline{x} - \overline{y}) \pm z_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$
$n_x < 30$, $n_y < 30$ populationerna är normalfördelade	σ_x^2, σ_y^2 känd: σ_x^2, σ_y^2 okänd: (men antas lika)	$(\overline{x} - \overline{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ $(\overline{x} - \overline{y}) \pm t_{v,\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}$ där $s_p^2 = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2}$
$n_x < 30$, $n_y < 30$ populationerna är ej normalfördelade	Konfidensintervall kan inte beräknas	

Konfidensintervall för $P_x - P_y$:

$$(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$$

Testvariabler vid test av hypotes om μ:

$n \geq 30$ oavsett om populationen är normalfördelad eller ej	σ^2 känd: $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ σ^2 okänd: $Z = \frac{\overline{X}-\mu_0}{s/\sqrt{n}}$
$n < 30$ populationen är normalfördelad	σ^2 känd: $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ σ^2 okänd: $t = \frac{\overline{X}-\mu_0}{s/\sqrt{n}}$

Testvariabel vid test av hypotes om P:

$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

Testvariabler vid test av hypotes om $\mu_x - \mu_y$ (Oberoende stickprov):

$n_x \geq 30$, $n_y \geq 30$ oavsett om populationerna är normalfördelade eller ej	σ_x^2, σ_y^2 känd: σ_x^2, σ_y^2 okänd:	$Z = \frac{(\overline{x}-\overline{y})-D_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}$ $Z = \frac{(\overline{x}-\overline{y})-D_0}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}}$
$n_x < 30$, $n_y < 30$ populationerna är normalfördelade	σ_x^2, σ_y^2 känd: σ_x^2, σ_y^2 okänd: (men antas lika)	$Z = \frac{(\overline{x}-\overline{y})-D_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}$ $t = \frac{(\overline{x}-\overline{y})-D_0}{\sqrt{s_p^2\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}}$ där $s_p^2 = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2}$
$n_x < 30$, $n_y < 30$ populationerna är ej normalfördelade	Konfidensintervall kan inte beräknas	

Testvariabel vid test av hypotes om $P_x - P_y$:

$$Z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}} \qquad \text{där} \qquad \hat{p}_0 = \frac{n_x\hat{p}_x + n_y\hat{p}_y}{n_x+n_y}$$

Testvariabler för χ^2-test:

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \qquad \text{där} \qquad E_i = nP_i$$

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad \text{där} \qquad E_{ij} = \frac{R_i C_j}{n}$$

Linjär regression:

$$Y_i = \beta_0 + \beta_1 x_i + \mathcal{E}_i \qquad\qquad \hat{y}_i = b_0 + b_1 x_i$$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{s_{xy}}{s_x^2} = r_{xy}\frac{s_y}{s_x}$$

$$b_0 = \overline{y} - b_1 \overline{x}$$

$$\text{SST} = \text{SSR} + \text{SSE}$$

$$\text{SST} = \sum_{i=1}^{n}(y_i - \overline{y})^2 \qquad \text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \qquad \text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$\text{MSR} = \frac{\text{SSR}}{K} \qquad\qquad \text{MSE} = \frac{\text{SSE}}{n - K - 1}$$

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} \qquad\qquad F = \frac{\text{MSR}}{\text{MSE}}$$

Konfidensintervall för β_j: $\quad b_j \pm t_{v,\alpha/2} \cdot s_{b_j}$

Testvariabel vid test av hypotes om β_j^*: $\quad t = \dfrac{b_j - \beta_j^*}{s_{b_j}}$

Testvariabel vid test av hypotes om ρ: $\quad t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$