TMA 321

Generella idén bakom K.I. $\boxed{F17}$ ①

<u>Steg 1:</u> Hitta en referensvariabel R där $X_1, \ldots, X_n$ och θ ingår. R väljs så att dess fördelning är känd.

<u>Steg 2:</u> Hitta lämpliga kvantiler (z_α, t_α ...)

<u>Steg 3:</u> Lös ut θ ur uttrycket

$$1-2\alpha = \mathbb{P}(x_{1-\alpha} \leq R \leq x_\alpha)$$

eller motsvarande.

<u>Normalapproximation för K.I.</u>

Vi betraktar proportions-skattning. Principen densamma i andra fall.

Antag att proportionen p har en viss egenskap och ta ett stickprov av storlek n. Låt $X = \#$ med egenskapen och antag att $X \sim Bin(n,p)$ (ok? oberoende?).

Vi skattar p med $\hat{p}(X) = \frac{X}{n}$, men vad blir K.I.?

<u>Alt 1:</u> Använd exakt fördelning ($X \sim Bin(n,p)$) (typ steg 1 ovan)

Steg 2: Hitta k_α s.a. $\mathbb{P}(|\frac{X}{n}-p|>k_\alpha) = 1-2\alpha$

Steg 3: $\mathbb{P}(|\frac{X}{n}-p|>k_\alpha) = \mathbb{P}(-k_\alpha < \frac{X}{n}-p < k_\alpha)$

$$= \mathbb{P}\left(\frac{X}{n}-k_\alpha < p < \frac{X}{n}+k_\alpha\right).$$

Ok? Njaej. Värdet på k_α kommer bero på p (som är okänd), X är ej symmetrisk, bör då K.I. vara det?

Alt 2: Om n är "stort" ger CGS att

$$X \approx N(E[X], Var(X)) = N(np, np(1-p))$$

$$\Rightarrow \quad \frac{X-np}{\sqrt{np(1-p)}} \approx N(0,1). \qquad \text{(steg 1)}$$

Steg 2: Kvantiler har vi.

Steg 3: Lös ut p ur $\mathbb{P}\left(-z_\alpha \leq \frac{X-np}{\sqrt{np(1-p)}} \leq z_\alpha\right)$

Steg 3': Lös ut p ur $\mathbb{P}\left(-z_\alpha \leq \frac{X-np}{\sqrt{n\hat{p}(1-\hat{p})}} \leq z_\alpha\right)$

$$= \mathbb{P}\left(-z_\alpha\sqrt{n\hat{p}(1-\hat{p})} \leq X-np \leq z_\alpha\sqrt{n\hat{p}(1-\hat{p})}\right)$$

$$= \mathbb{P}\left(\frac{X}{n}-z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \frac{X}{n}+z_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

Anm: Principen densamma, men detaljer olika i andra fall.

K.I. för σ^2 i normalfallet.

Låt $X_1, X_2, \ldots$ vara iid $N(\mu, \sigma^2)$ och låt

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Def: En s.v. Y är chi-kvadrat fördelad med n frihetsgrader om

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

där $Z_1, \ldots, Z_n$ är iid. $N(0,1)$. $(Y \sim \chi^2(n))$

Sats: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

B: Lång räkning och mgf $\square$

Låt $\chi^2_{n-1}(\alpha)$ vara α-kvantilen för $\chi^2(n-1)$-förd.

Vi kan hitta ett K.I. med konf.grad α genom att

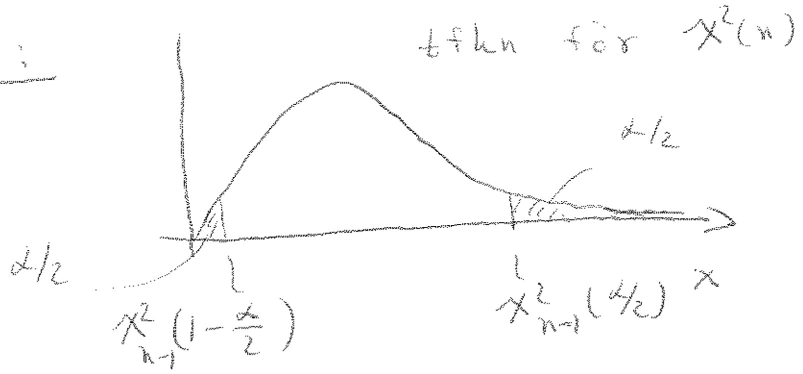
$$1-\alpha = \mathbb{P}\left(\chi^2_{n-1}(1-\alpha/2) \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1}(\alpha/2) \right)$$

$$= \mathbb{P}\left(\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)} \right)$$

s.a. $I_{\sigma^2} = \left[\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)} \right]$

är vårt K.I. med konf. grad $1-\alpha$.

<u>Anm:</u>



Ej symmetrisk! $\chi^2_{n-1}(1-\alpha/2) \neq -\chi^2_{n-1}(\alpha/2)$

<u>Tal (forts):</u> Data $N(\mu, \sigma^2)$ och gav

$\bar{x} \approx 4.84$, $s^2(x) \approx 6.21$

c) Hitta ett 95 % K.I. för σ^2

<u>L:</u> Här är $\alpha = 0.05$ och $n = 13$. Tabell ger

$\chi^2_{12}(0.025) \approx 23.34$ och $\chi^2_{12}(0.975) \approx 4.40$

(tabell sid A8) OBS: Tabellen visar $\mathbb{P}(\cdot \leq \chi^2_{n-1}(\alpha))$

ej $\mathbb{P}(\cdot \geq \chi^2_{n-1}(\alpha))$!)

S.a.

$$I_{\sigma^2} = \left[\frac{12 \cdot 6.21}{23.34}, \frac{12 \cdot 6.21}{4.4}\right] \approx [3.19, 16.94] //$$

<u>Anm:</u> Om man blandar ihop $\chi^2_{n-1}(\alpha)$ och $\chi^2_{n-1}(1-\alpha)$

blir $I_{\sigma^2} = [16.94, 3.19]$ och man ser att det

är fel.

Två stickprov

Vi har två oberoende stickprov $X = (X_1, \dots, X_n)$ och $Y = (Y_1, \dots, Y_m)$ där

$X_i \sim N(\mu_1, \sigma_1^2)$ och $Y_i \sim N(\mu_2, \sigma_2^2)$.

Vi vill analysera $\theta = \mu_1 - \mu_2$ som skattas med $\hat{\theta} = \bar{X} - \bar{Y}$.

Fall 1: σ_1, σ_2 kända. Vi har att $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n})$

$\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m})$ s.a. $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m})$

Låt $D^2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$ och bilda

$$R = \frac{\bar{X} - \bar{Y} - \theta}{D} \sim N(0,1).$$

$$\Rightarrow \quad 1 - \alpha = \mathbb{P}(-z_{\alpha/2} \leq R \leq z_{\alpha/2}) = \mathbb{P}(\bar{X} - \bar{Y} - z_{\alpha/2} D \leq \theta \leq \bar{X} - \bar{Y} + z_{\alpha/2} D)$$

$$\Rightarrow \quad I_{\mu_1 - \mu_2} = \bar{X} - \bar{Y} \pm z_{\alpha/2} D$$

Tal: I en undersökning uppmättes pH-värdet i två olika sjöar. Den första kontrollerades på $n=20$ platser och den andra på $m=29$. Resultaten anses vara oberoende och $N(\mu_1, \sigma_1^2)$ resp. $N(\mu_2, \sigma_2^2)$ fördelade. Om $\sigma_1^2 = 1.4$, $\sigma_2^2 = 1.9$, hitta ett 95 % K.I. för $\mu_1 - \mu_2$ om data gav $\bar{x} = 5.3$ och $\bar{y} = 6.1$

__1:__ Vi får att $D = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \approx 0.368$.

Tabell ger $z_{0.025} = 1.96$ så ett 95 % K.I.

för $\mu_1 - \mu_2$ blir

$$I_{\mu_1-\mu_2} = [-0.8 - 1.96 \cdot 0.368, -0.8 + 1.96 \cdot 0.368]$$

$$= [-1.52, -0.079] //$$

__Anm:__ Har vi fog att hävda att de faktiska pH-värdena är olika?

Antag nu att $X_i \sim N(\mu_1, \sigma^2)$ och $Y_i \sim N(\mu_2, \sigma^2)$

$$\Rightarrow \quad \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right).$$

Hur skall σ^2 skattas?

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \text{och} \quad S_Y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2$$

skattar båda σ^2. Man kan visa att

$$S_p^2 = \frac{(n-1) S_X^2 + (m-1) S_Y^2}{n+m-2}$$ "p för pooled"

är VVR och har minst varians av alla skattare på formen $\beta S_X^2 + (1-\beta) S_Y^2$.