TMA 321

F18 (1)

Två stickprov forts

Antag att $X_1, \dots, X_n$ och $Y_1, \dots, Y_m$ är oberoende med $X_i \sim N(\mu_1, \sigma^2)$ och $Y_i \sim N(\mu_2, \sigma^2)$ (samma varians!). Vi får att

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma^2}{n}\right) \text{ och } \bar{Y} \sim N\left(\mu_2, \frac{\sigma^2}{m}\right)$$

så att
$$\hat{\Theta} = \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$$

OBS: $\mathrm{Var}(\bar{X} - \bar{Y}) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(-\bar{Y}) + 2\,\mathrm{Cov}(\bar{X}, -\bar{Y})$

$= \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) - 2\,\mathrm{Cov}(\bar{X}, \bar{Y})$

$\underset{\bar{X}, \bar{Y} \text{ ober.}}{=} \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y})$

Här är σ^2 okänd och skattas m.h.a

$$S_p^2 = \frac{(n-1)S_{\bar{X}}^2 + (m-1)S_{\bar{Y}}^2}{n+m-2}.$$

Vi använder
$$R = \frac{\hat{\Theta}(\bar{X}, \bar{Y}) - \Theta}{S_p(\bar{X}, \bar{Y})\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

Sats: Om R är som ovan gäller att

$$R \sim t(m+n-2)$$

Vi får att

$$1-\alpha = \mathbb{P}\left(|\hat{\theta} - \theta| \leq t_{m+n-2}(\alpha)\, S_p(\bar{X},\bar{Y}) \sqrt{\tfrac{1}{n}+\tfrac{1}{m}} \right)$$

Så att ett $(1-\alpha)\cdot 100\ \%$ K.I. för θ

blir $\quad I_\theta = \bar{X}-\bar{Y} \pm t_{m+n-2}(\alpha)\, S_p(\bar{X},\bar{Y})\sqrt{\tfrac{1}{n}+\tfrac{1}{m}}$

Tal: (Som förra gången) Två sjöar $n=20$, $m=29$

$\hat{\mu}_1 = \bar{x} = 5.3 \quad \hat{\mu}_2 = \bar{y} = 6.1$ och

$s_x^2 = 1.52$ och $s_y^2 = 1.83$ så

Hitta ett 95 % K.I. för $\mu_1 - \mu_2$.

L: Vi slår ihop skattningarna av σ^2:

$$s_p^2 = \frac{19}{47}\cdot 1.52 + \frac{28}{47}\cdot 1.83 \approx 1.705$$

Vi har att $\quad t_{47}(0.025) \approx 2.01$

[OBS: Tabell anger $t_{40}(0.025)$ och $t_{60}(0.025)$
så man får interpolera...]

så ett 95 % numeriskt K.I. blir

$$I_{\mu_1-\mu_2} = -0.8 \pm 2.01\cdot\sqrt{1.705}\sqrt{\frac{1}{20}+\frac{1}{29}} \approx [-1.563, -0.037]$$

Vi har fallet $\sigma_1 \neq \sigma_2$ båda okända

hvar, Vi skippar detta.

Man kan förstås tänka sig två stickprov

där data ej kommer från en normalfördelning.

<u>Tal:</u> Två väljarundersökningar genomfördes

med 4 månaders mellanrum. Resultatet

för parti Q blev som följer:

1) 143 av $n_1 = 1217$ tillfrågade stöttar Q

2) 139 " $n_2 = 1078$ " " Q

Skapa ett 99 % K.I. för $p_2 - p_1$

där p_i = väljarandel vid undersökning i.

<u>1:</u> Låt $\hat{p}_1 = \frac{X_1}{n_1}$ och $\hat{p}_2 = \frac{X_2}{n_2}$ där

X_i = # anhängare vid undersökning i.

[OBS: X_i s.v. "innanperspektiv"]

Vi har att (?) $X_i \sim Bin(n_i, p_i)$

S.a. $\hat{p}_i \approx N(E[\hat{p}_i], Var(\hat{p}_i)) = N\left(p_i, \frac{p_i(1-p_i)}{n_i}\right)$

$$\Rightarrow \quad \hat{p}_1 - \hat{p}_2 \approx N\Big(p_1 - p_2,\ \underbrace{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}_{= \cancel{s^2_{\hat{p}_1-\hat{p}_2}}\ Var(\hat{p}_1-\hat{p}_2)}\Big)$$

Vi ersätter variansen (dvs dess

skattning $\quad s^2_{\hat{p}_1-\hat{p}_2} = \dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}$

Tabell ger oss att

$$0.99 = \mathbb{P}(-2.575 \le Z \le 2.575)$$

$$\approx \mathbb{P}\left(-2.575 \le \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{s_{\hat{p}_1-\hat{p}_2}} \le 2.575\right)$$

$$= \mathbb{P}\Big(\hat{p}_1 - \hat{p}_2 - 2.575 \cdot s_{\hat{p}_1-\hat{p}_2} < p_1 - p_2$$

$$< \hat{p}_1 - \hat{p}_2 + 2.575\, s_{\hat{p}_1-\hat{p}_2}\Big)$$

Ett 99 % numeriskt K.I. blir:

$$I_{p_1-p_2} \approx -0.0114 \pm 2.575 \cdot 0.0138 \approx [-0.047, 0.024]$$

<u>Anm:</u> Har vi fog att hävda att sympatierna har ändrats (dvs $p_1 \ne p_2$)?

Stickprov i par

Dyker upp naturligt t. ex. när samma experiment upprepas vid olika tidpunkter eller med olika metoder

Tal: Halten bly upmäts i 20 vattendrag som ligger i anslutning till en nyetablerad fabrik. Mätning genomfördes innan, och ett år efter fabriken startades. Resultatet av mätningarna blev som följer:

Drag nr:	1	2	3	4	......	19	20
Innan:	1.7	1.6	2.7	4.3		0.7	2.9
Efter:	1.9	1.5	3.1	3.9		0.8	3.4

Under antagandet att mätningarnas resultat kommer från normalfördelningar, bestäm ett 99 % K.I. för skillnaden i halten bly.

Anm: Låt μ_k = halten bly i vattendrag k innan mätning.

Det är rimligt att förvänta sig skillnader mellan vattendragen, dvs $\mu_1 \neq \mu_2$ etc.

Vi antar dock att skillnaden Δ innan/efter är densamma för alla vattendragen, dvs $\mu_1 \Rightarrow \mu_1+\Delta$ $\mu_2 \Rightarrow \mu_2+\Delta, \dots, \mu_{20} \Rightarrow \mu_{20}+\Delta$.

Vi vill skatta Δ!

Vad innebär normalantagandet?

Jo: X_k = uppmätt halt innan

Y_k = " " efter

$\Rightarrow$ $X_k \sim N(\mu_k, \sigma_1^2)$ och $Y_k \sim N(\mu_k+\Delta, \sigma_2^2)$

och därmed blir $Z_k := Y_k - X_k \sim N(\Delta, \sigma_1^2+\sigma_2^2)$

1: Vi skattar Δ med $\hat{\Delta} = \bar{Z}$ där $Z_k \sim N(\Delta, \sigma^2)$ $(\sigma^2 = \sigma_1^2+\sigma_2^2)$. Vi får att

$$R = \frac{\hat{\Delta} - \Delta}{S(Z)/\sqrt{n}} \sim t(n-1).$$

Data ger oss $\bar{z} = 0.42$ och $S_z = 0.32$

s.a.
$$I_\Delta = \bar{z} \pm t_{19}(0.005)\frac{S_z}{\sqrt{20}} = 0.42 \pm 2.86 \cdot \frac{0.32}{\sqrt{20}}$$
$$= 0.42 \pm 0.205 //$$

Anm: Parat stickprov reduceras till ett stickprov!