NTNU
Kunnskap for ei betre verd

To-utval og prediksjonsintervall

Torstein Fjeldstad
Institutt for matematiske fag, NTNU
30.10.2018

I dag

- Repetisjon
- To-utval
- Parvise observasjonar
- Prediksjonsintervall

Repetisjon

Utledning av konfidensintervall

Situasjon: Anta $X_1, X_2, \ldots, X_n$ tilfeldig utval frå $f(x;\theta)$-populasjonen. Ynskjer eit $(1-\alpha)\cdot 100\%$ konfidensintervall for θ.

1. Finn ein estimator for θ, $\hat{\theta}$ (t.d. SME)
2. La $Z = h(\hat{\theta}, \theta)$ der $h(\cdot,\cdot)$ er ein funksjon s.a. Z har ei kjend fordeling.
3. Då har me at
$$P(z_{1-\alpha/2} \leq h(\hat{\theta}, \theta) \leq z_{\alpha/2}) = 1-\alpha$$
4. Løys ulikskapane (mhp. θ) kvar for seg og finn eit uttrykk med θ i midten
$$P(\hat{\theta}_L(X_1, X_2, \ldots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, X_2, \ldots, X_n)) = 1-\alpha$$
5. Et $(1-\alpha)\cdot 100$ % konfidensintervall for θ er
$$\left[\hat{\theta}_L(X_1, X_2, \ldots, X_n), \hat{\theta}_U(X_1, X_2, \ldots, X_n)\right]$$

Einsidig konfidensintervall

Situasjon: $X_1, X_2, \ldots, X_n$ tilfeldig utval med $X_i \sim n(x_i; \mu, \sigma)$ der σ er kjend. Dei einsidige $(1-\alpha)100\%$ konfidensgrensene er

øvre grense: $$\bar{x} + z_\alpha \sqrt{\frac{\sigma^2}{n}}$$

nedre grense: $$\bar{x} - z_\alpha \sqrt{\frac{\sigma^2}{n}}$$

To-utval

To-utval

Situasjon (X og Y uavhengig):

— $X_1, X_2, \ldots, X_n$ uif $n(x; \mu_x, \sigma_x)$

— $Y_1, Y_2, \ldots, Y_m$ uif $n(y; \mu_y, \sigma_y)$

Mål: $(1 - \alpha)100$ % konfidensintervall for differanse $\mu_x - \mu_y$.

To-utval

Situasjon (X og Y uavhengig):

- $X_1, X_2, \ldots, X_n$ uif $n(x; \mu_X, \sigma_X)$
- $Y_1, Y_2, \ldots, Y_m$ uif $n(y; \mu_Y, \sigma_Y)$

Mål: $(1-\alpha)100$ % konfidensintervall for differanse $\mu_X - \mu_Y$.

To situasjonar

- σ_X og σ_Y kjend:

$$\left[\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right]$$

To-utval

Situasjon (X og Y uavhengig):

— $X_1, X_2, \ldots, X_n$ uif $n(x; \mu_x, \sigma_x)$

— $Y_1, Y_2, \ldots, Y_m$ uif $n(y; \mu_y, \sigma_y)$

Mål: $(1-\alpha)100$ % konfidensintervall for differanse $\mu_x - \mu_y$.

To situasjonar

— σ_x og σ_y kjend:

$$\left[\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}, \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right]$$

— σ_x og σ_y ukjend:

$$\left[\bar{X} - \bar{Y} - t_{\alpha/2,v}\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}, \bar{X} - \bar{Y} + t_{\alpha/2,v}\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}\right]$$

Korreksjon fridomsgrader

$$v = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{\left(\frac{S_X^2}{n}\right)^2}{n-1} + \frac{\left(\frac{S_Y^2}{m}\right)^2}{m-1}}$$

Parvise observasjonar

Parvise observasjonar

Situasjon ($X_{1,i}$ og $X_{2,i}$ avhengig):

— $X_{1,1}, X_{1,2}, \ldots, X_{1,n}$ uif $n(x_1; \mu_1, \sigma_1)$

— $X_{2,1}, X_{2,2}, \ldots, X_{2,n}$ uif $n(x_2; \mu_2, \sigma_2)$

Parvise observasjonar

Situasjon ($X_{1,i}$ og $X_{2,i}$ avhengig):

— $X_{1,1}, X_{1,2}, \ldots, X_{1,n}$ uif $n(x_1; \mu_1, \sigma_1)$

— $X_{2,1}, X_{2,2}, \ldots, X_{2,n}$ uif $n(x_2; \mu_2, \sigma_2)$

La

$$D_i = X_{1,i} - X_{2,i} \sim n\left(d; \mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2 - 2\mathrm{Cov}(X_{1,i}, X_{2,i})}\right)$$

Parvise observasjonar

Situasjon ($X_{1,i}$ og $X_{2,i}$ avhengig):

— $X_{1,1}, X_{1,2}, \ldots, X_{1,n}$ uif $n(x_1; \mu_1, \sigma_1)$

— $X_{2,1}, X_{2,2}, \ldots, X_{2,n}$ uif $n(x_2; \mu_2, \sigma_2)$

La

$$D_i = X_{1,i} - X_{2,i} \sim n\left(d; \mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2 - 2\mathrm{Cov}(X_{1,i}, X_{2,i})}\right)$$

Mål: $(1-\alpha)100$ % konfidensintervall for differanse $\mu_D = \mu_1 - \mu_2$.
Estimator for $\mu_1 - \mu_2 = \bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$.

Parvise observasjonar

Situasjon ($X_{1,i}$ og $X_{2,i}$ avhengig):

— $X_{1,1}, X_{1,2}, \ldots, X_{1,n}$ uif $n(x_1; \mu_1, \sigma_1)$

— $X_{2,1}, X_{2,2}, \ldots, X_{2,n}$ uif $n(x_2; \mu_2, \sigma_2)$

La

$$D_i = X_{1,i} - X_{2,i} \sim n\left(d; \mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2 - 2\mathrm{Cov}(X_{1,i}, X_{2,i})}\right)$$

Mål: $(1-\alpha)100$ % konfidensintervall for differanse $\mu_D = \mu_1 - \mu_2$.
Estimator for $\mu_1 - \mu_2 = \bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$.
Eit $(1-\alpha)100$ % konfidensintervall for $\mu_1 - \mu_2$:

$$\left[\bar{D} - t_{\alpha/2,n-1}\sqrt{\frac{S_D^2}{n}}, \bar{D} + t_{\alpha/2,n-1}\sqrt{\frac{S_D^2}{n}},\right]$$

Prediksjonsintervall

Prediksjonsintervall

Situasjon: $X_1, X_2, \ldots, X_n$ tilfeldig utval med $X_i \sim n(x_i; \mu, \sigma)$ der σ er kjend og μ ukjend. La $X_0 \sim n(x_0; \mu, \sigma)$ vere ei framtidig måling, uavhengig av $X_1, X_2, \ldots, X_n$.

— Ynskjer eit $(1-\alpha)100\%$ prediksjonsintervall for X_0

— Estimator for μ, $\hat{\mu} = \bar{X}$

— Start med

$$\bar{X} - X_0 \sim n\left(z; 0, \sqrt{\frac{\sigma^2}{n} + \sigma^2}\right)$$

— Observator

$$Z = \frac{\bar{X} - X_0}{\sqrt{\sigma^2\left(\frac{1}{n}+1\right)}} \sim n(z; 0, 1)$$

— $(1-\alpha)100\%$ prediksjonsintervall for X_0

$$\left[\bar{X} - z_{\alpha/2}\sqrt{\sigma^2\left(\frac{1}{n}+1\right)}, \bar{X} + z_{\alpha/2}\sqrt{\sigma^2\left(\frac{1}{n}+1\right)}\right]$$

Skilnad konfidensintervall og prediksjonsintervall

Konfidensintervall:

$$P\left(\bar{X} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha$$

Prediksjonsintervall:

$$P\left(\bar{X} - z_{\alpha/2}\sqrt{\sigma^2\left(\frac{1}{n}+1\right)} \leq X_0 \leq \bar{X} + z_{\alpha/2}\sqrt{\sigma^2\left(\frac{1}{n}+1\right)}\right) = 1 - \alpha$$

Torsdag

— Hypotesetesting