STK1110 Høsten 2021

Mer om hypotesetesting og konfidensintervaller for to utvalg

Tilsvarer Avsnitt 10.3, 10.4 og 10.5

Ingrid Hobæk Haff
Matematisk institutt
Universitetet i Oslo

Hypotesetest for σ_1^2/σ_2^2

- Vi antar at $X_1, \ldots, X_m \overset{uif}{\sim} N(\mu_1, \sigma_1^2)$ og $Y_1, \ldots, Y_n \overset{uif}{\sim} N(\mu_2, \sigma_2^2)$, der μ_1, μ_2, σ_1 og σ_2 er ukjent.
- Vi vil teste hypotesene $H_0 : \sigma_1^2 = \sigma_2^2$ mot $H_a : \sigma_1^2 \neq \sigma_2^2$.
- Vi vet at $S_1^2 = \frac{1}{m-1}\sum_{i=1}^m (X_i - \bar{X})^2$ og $S_2^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$ er forventningsrette estimatorer for σ_1^2 og σ_2^2.
- Videre vet vi at $\frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}$ og $\frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n-1}$ er uavhengige.
- Da vet vi fra Kap. 6.4 at $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{m-1,n-1}$.
- Under H_0 er $\sigma_1^2 = \sigma_2^2$, og da er $F = \frac{S_1^2}{S_2^2} \sim F_{m-1,n-1}$.
- Da får vi en test med signifikansnivå α dersom vi forkaster H_0 når $F \leq F_{1-\alpha/2,m-1,n-1}$ eller $F \geq F_{\alpha/2,m-1,n-1}$.

Konfidensintervall for σ_1^2/σ_2^2

- For å lage et $100 \cdot (1-\alpha)\%$ konfidensintervall for σ_1^2/σ_2^2 bruker vi at
$$\mathrm{P}\left(F_{1-\alpha/2,m-1,n-1} \leq \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq F_{\alpha/2,m-1,n-1}\right) = 1-\alpha$$
som betyr at
$$\mathrm{P}\left(\frac{S_1^2}{S_2^2}F_{1-\alpha/2,m-1,n-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2}F_{\alpha/2,m-1,n-1}\right) = 1-\alpha.$$
- Et $100 \cdot (1-\alpha)\%$ konfidensintervall for σ_1^2/σ_2^2 er dermed gitt ved
$$\left(\frac{s_1^2}{s_2^2}F_{1-\alpha/2,m-1,n-1}, \frac{s_1^2}{s_2^2}F_{\alpha/2,m-1,n-1}\right)$$
- Metodene over er nokså følsomme overfor antakelsen om normalfordeling.

Eksempel

Vektøking for rotter.

Konfidensintervall for $p_1 - p_2$

- Vi antar at $X \sim Bin(m, p_1)$ og $Y \sim Bin(n, p_2)$ er uavhengige.
- Vi ønsker å lage et konfidensintervall for $p_1 - p_2$.
- La $\hat{p}_1 = \frac{X}{m}$ og $\hat{p}_2 = \frac{Y}{n}$, som vi vet er forventningsrette for p_1 og p_2.
- En naturlig estimator for $p_1 - p_2$ er dermed $\hat{p}_1 - \hat{p}_2$.
- Vi har:

$$\begin{aligned}
\mathrm{E}(\hat{p}_1 - \hat{p}_2) &= \mathrm{E}(\hat{p}_1) - \mathrm{E}(\hat{p}_2) = p_1 - p_2 \\
\mathrm{V}(\hat{p}_1 - \hat{p}_2) &\overset{uavh.}{=} \mathrm{V}(\hat{p}_1) + (-1)^2\mathrm{V}(\hat{p}_2) = \frac{p_1(1-p_1)}{m} + \frac{p_2(1-p_2)}{n}.
\end{aligned}$$

Konfidensintervall for $p_1 - p_2$ (forts.)

- Vi antar at m og n er store nok til at $\hat{p}_1$ og $\hat{p}_2$ tilnærmet normalfordelt.
- Da er også $\hat{p}_1 - \hat{p}_2$ tilnærmet normalfordelt, slik at

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{m} + \frac{p_2(1-p_2)}{n}}} \overset{tiln.}{\sim} N(0,1).$$

- Det betyr igjen at

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}} \overset{tiln.}{\sim} N(0,1).$$

Konfidensintervall for $p_1 - p_2$ (forts.)

- Vi har

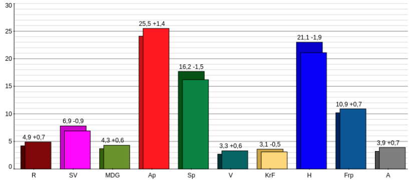
$$P\left(-z_{\alpha/2} \leq \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

- Et $100 \cdot (1-\alpha)\%$ konfidensintervall for $p_1 - p_2$ er dermed gitt ved

$$\left(\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}\right).$$

Eksempel

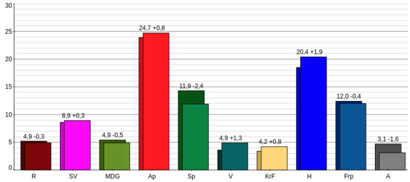
Norstat for NRK 12. august 2021



Fakta om meningsmålingen

Valg	Stortingsvalg
Område	Hele landet
Institutt	Norstat
Antall spurte	11400
Tatt opp	28/7 - 11/8

Norstat for NRK 9. september 2021



Fakta om meningsmålingen

Valg	Stortingsvalg
Område	Hele landet
Institutt	Norstat
Antall spurte	11500
Tatt opp	31/8 - 8/9

- I august var det 1330 av 8208 spurte som ville ha stemt på Sp om det var stortingsvalg dagen etter.
- I september var de tilsvarende tallene 985 av 8280.
- La p_1 og p_2 være oppslutningen om Sp i henholdsvis august og september.
- Vi vil lage et 95% konfidensintervall for $p_1 - p_2$.

Hypotesetest ang. p_1 og p_2

- Vi vil nå teste $H_0 : p_1 \leq p_2$ mot $H_a : p_1 > p_2$.
- Som vanlig holder det å beregne sannsynligheten for type I-feil i nullverdien, dvs. når $p_1 = p_2 = p$.
- En estimator for p er $\hat{p} = \frac{X+Y}{m+n} = \frac{m}{m+n}\hat{p}_1 + \frac{n}{m+n}\hat{p}_2$.
- Da er også $V(\hat{p}_1 - \hat{p}_2) = \frac{p(1-p)}{m} + \frac{p(1-p)}{n} = p(1-p)\left(\frac{1}{m} + \frac{1}{n}\right)$.
- Det betyr at

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m} + \frac{1}{n}\right)}} \overset{tiln.}{\sim} N(0,1).$$

Hypotesetest ang. p_1 og p_2 (forts.)

- Dermed får vi en test med tilnærmet signifikansnivå α hvis vi forkaster H_0 dersom

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m}+\frac{1}{n}\right)}} \geq z_\alpha.$$

- For tester av $H_0 : p_1 \geq p_2$ mot $H_a : p_1 < p_2$ får vi tilnærmet signifikansnivå α hvis vi forkaster H_0 dersom $Z \leq -z_\alpha$.
- For tester av $H_0 : p_1 = p_2$ mot $H_a : p_1 \neq p_2$ får vi tilnærmet signifikansnivå α hvis vi forkaster H_0 dersom $Z \leq -z_{\alpha/2}$ eller $Z \geq z_{\alpha/2}$.

Eksempel

Meningsmåling.

Hypotesetest ang. p_1 og p_2 (forts.)

- Vi vil nå finne styrkefunksjonen til tester for $H_0 : p_1 \leq p_2$ mot $H_a : p_1 > p_2$.
- La $\bar{p} = \frac{m}{m+n} p_1 + \frac{n}{m+n} p_2$ og $\sigma_{\hat{p}_1 - \hat{p}_2} = \text{sd}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1(1-p_1)}{m} + \frac{p_2(1-p_2)}{n}}$.
- Styrkefunksjonen er da gitt ved:

$$\gamma(p_1, p_2) \approx 1 - \Phi\left(\frac{z_\alpha \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{m} + \frac{1}{n}\right)} - (p_1 - p_2)}{\sigma_{\hat{p}_1 - \hat{p}_2}}\right).$$

- Vi vil bestemme utvalgsstørrelsene m og n, med $m = n$, slik at sannsynligheten for feil av type II blir høyst β for $p_1 = p_1'$ og $p_2 = p_2'$ med $p_1' > p_2'$.

Hypotesetest ang. p_1 og p_2 (forts.)

- Sannsynligheten for feil av type II er da

$$\beta(p_1', p_2') = 1 - \gamma(p_1', p_2') = \Phi\left(\frac{z_\alpha\sqrt{\bar{p}(1-\bar{p})\frac{2}{n}} - (p_1' - p_2')}{\sigma_{\hat{p}_1 - \hat{p}_2}}\right)$$

- Vi løser for n når $\beta(p_1', p_2') \leq \beta$, og får

$$n \geq \left(\frac{z_\alpha\sqrt{(p_1' + p_2')\left(1 - \frac{p_1'+p_2'}{2}\right)} + z_\beta\sqrt{p_1'(1-p_1') + p_2'(1-p_2')}}{p_1' - p_2'}\right)^2$$

Hypotesetest ang. p_1 og p_2 (forts.)

- For tester av $H_0 : p_1 \geq p_2$ mot $H_a : p_1 < p_2$ er styrkefunksjonen

$$\gamma(p_1, p_2) \approx \Phi\left(\frac{-z_\alpha\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma_{\hat{p}_1 - \hat{p}_2}}\right),$$

  og den $m = n$ som gir sannsynlighet for feil av type II høyst lik β for $p_1 = p_1'$ og $p_2 = p_2'$ med $p_1' < p_2'$ er gitt ved samme formel som over.
- For tester av $H_0 : p_1 = p_2$ mot $H_a : p_1 \neq p_2$ er styrkefunksjonen

$$\begin{aligned}\gamma(p_1, p_2) \approx & 1 - \Phi\left(\frac{z_{\alpha/2}\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma_{\hat{p}_1 - \hat{p}_2}}\right) \\ & + \Phi\left(\frac{-z_{\alpha/2}\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma_{\hat{p}_1 - \hat{p}_2}}\right).\end{aligned}$$

Hypotesetest ang. p_1 og p_2 (forts.)

- Den $m = n$ som gir sannsynlighet for feil av type II høyst lik β for $p_1 = p_1'$ og $p_2 = p_2'$ med $p_1' \neq p_2'$ er gitt ved

$$n \geq \left(\frac{z_{\alpha/2}\sqrt{(p_1' + p_2')\left(1 - \frac{p_1'+p_2'}{2}\right)} + z_\beta \sqrt{p_1'(1-p_1') + p_2'(1-p_2')}}{p_1' - p_2'} \right)^2$$

Eksempel

Eks. 10.12 i boka.

Parforsøk

- Så langt har vi antatt at X_i-ene og Y_i-ene i de to utvalgene er uavhengige.
- Det er naturlig dersom de er knyttet til to forskjellige populasjoner.
- I en del tilfeller gjelder imidlertid ikke denne antakelsen, f.eks. hvis X_i og Y_i er to målinger på samme person i et medisinsk forsøk.
- Da kan en i stedet organisere målingene parvis.

Parforsøk (forts.)

- Anta at vi har par av observasjoner $(X_1, Y_1), \dots, (X_n, Y_n)$, med $\mathrm{E}(X_i) = \mu_1$ og $\mathrm{E}(Y_i) = \mu_2$.
- La $D_i = X_i - Y_i$, $i = 1, \dots, n$.
- Anta at $D_1 \dots, D_n \overset{uif}{\sim} N(\mu_D, \sigma_D^2)$, der $\mu_D = \mathrm{E}(D_i) = \mathrm{E}(X_i) - \mathrm{E}(Y_i) = \mu_1 - \mu_2$ og $\sigma_D^2 = \mathrm{V}(D_i)$.
- Vi kan da bruke metodene for ett normalfordelt utvalg til å lage konfidensintervall (Kap. 8) og tester (Kap. 9) for $\mu_D = \mu_1 - \mu_2$.

Eksempel

Eks. 10.9 i boka.

Parforsøk eller to uavhengige grupper?

- Når bør en bruke t-test for pardata og når er det bedre å bruke en t-test for to uavhengige, like store utvalg?
- Testobservator for parforsøk: $T = \frac{\bar{D}}{S_D/\sqrt{n}} = \frac{\bar{X}-\bar{Y}}{S_D/\sqrt{n}}$, der $S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$ er en estimator for σ_D^2.
- Testobservator for to uavhengige grupper: $T = \frac{\bar{X}-\bar{Y}}{\sqrt{(S_1^2+S_2^2)/n}}$, der $S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ og $S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ er estimatorer for henholdvis $\sigma_1^2 = \mathrm{V}(X_i)$ og $\sigma_2^2 = \mathrm{V}(Y_i)$.
- Forskjellen på disse to er nevnerne.

Parforsøk eller to uavhengige grupper? (forts.)

- For pardata har vi:

$$\sigma_D^2 = \mathrm{V}(X_i - Y_i) = \sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho,$$

  der $\rho = \mathrm{Corr}(X_i, Y_i)$.

- For uavhengige grupper har vi

$$\mathrm{V}(X_i - Y_i) \stackrel{uavh.}{=} \sigma_1^2 + \sigma_2^2.$$

- Hvis ρ er positiv og stor, er dermed σ_D^2 mye mindre enn $\sigma_1^2 + \sigma_2^2$, og da er det best å bruke framgangsmåten for parforsøk, fordi det da er lettere å oppdage forskjeller mellom gruppene.