STK1110 Høsten 2021

Hypotesetesting om en populasjonsandel

Tilsvarer Avsnitt 9.3

Ingrid Hobæk Haff
Matematisk institutt
Universitetet i Oslo

Hypotesetesting om en populasjonsandel p

- Anta at vi har et utvalg $X_1, \ldots, X_n$ fra en stor populasjon, der en andel p har en viss egenskap, med $X_i = 1$ hvis individ/element i har egenskapen og $X_i = 0$ ellers.
- Videre antar vi at $X_1, \ldots, X_n \overset{uif}{\sim} Bernoulli(p)$.
- Da er $Y = \sum_{i=1}^{n} X_i \sim Bin(n, p)$.
- Vi er interessert i hypoteser av typen
  - $H_0 : p \leq p_0$ mot $H_a : p > p_0$
  - $H_0 : p \geq p_0$ mot $H_a : p < p_0$
  - $H_0 : p = p_0$ mot $H_a : p \neq p_0$.

Store utvalg

- Anta først at utvalgsstørrelsen n er stor.
- Vi vet at $\hat{p} = \frac{Y}{n}$ er forventningsrett for p og at $\mathrm{V}(\hat{p}) = \frac{p(1-p)}{n}$.
- Hvis $np_0 \geq 10$ og $n(1-p_0) \geq 10$, så er

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \overset{tiln.}{\sim} N(0,1)$$

  når $p = p_0$.
- En test med tilnærmet signifikansnivå α for å teste $H_0 : p \leq p_0$ mot $H_a : p > p_0$ forkaster da H_0 dersom

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \geq z_\alpha.$$

Store utvalg (forts.)

- Styrkefunksjonen er da

$$\gamma(p) = \mathrm{P}(\text{Forkaste} H_0 | p) \approx 1 - \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right).$$

- Vi ønsker å finne n slik at sannsynligheten for feil av type II blir høyst β når $p = p_1 > p_0$.
- Med

$$\begin{aligned}
\mathrm{P}(\text{Feil av type II} | p = p_1) =& 1 - \gamma(p_1) \\
\approx& \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \leq \beta
\end{aligned}$$

får vi

$$n \geq \left(\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)}}{p_1 - p_0}\right)^2.$$

Store utvalg (forts.)

- For tester av typen $H_0 : p \geq p_0$ mot $H_a : p < p_0$ forkaster en H_0 ved signifikansnivå tilnærmet α dersom $Z \leq -z_\alpha$.
- Styrkefunksjonen er da

$$\gamma(p) \approx \Phi\left(\frac{p_0 - p - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right).$$

- Den n som gir sannsynlighet for feil av type II høyst lik β når $p = p_1 < p_0$ er da som på forrige foil.

Store utvalg (forts.)

- For tester av typen $H_0 : p = p_0$ mot $H_a : p \neq p_0$ forkaster en H_0 ved signifikansnivå tilnærmet α dersom $Z \leq -z_{\alpha/2}$ eller $Z \geq z_{\alpha/2}$.
- Styrkefunksjonen er da

$$\begin{aligned}\gamma(p) \approx & \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) \\ & - \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right).\end{aligned}$$

- Den n som gir sannsynlighet for feil av type II høyst lik β når $p = p_1 < p_0$ er da tilnærmet

$$n \geq \left(\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)}}{p_1 - p_0}\right)^2.$$

Eksempel

Forsøk med salve.

Små utvalg

- Dersom n ikke er stor nok til at vi kan bruke tilnærmingen med normalfordeling, kan vi i stedet bruke den binomiske fordelingen direkte.
- Anta at vi vil teste $H_0 : p \leq p_0$ mot $H_a : p > p_0$.
- En forkaster da H_0 for $Y \geq k$ for en passende k.
- Da har vi

$$\mathrm{P}(\text{Feil av type I}) \leq 1 - B(k-1; n, p_0),$$

der $B(\cdot; n, p_0)$ er fordelingsfunksjonen til den binomiske fordelingen med parametere (n, p_0).

Små utvalg (forst.)

- Det er sjelden mulig å finne en k slik at sannsynligheten for feil av type I blir akkurat lik signifikansnivået.
- I stedet velger en det minste heltallet k slik at

$$1 - B(k-1; n, p_0) \leq \alpha.$$

- For $p_1 < p_0$ får vi sannsynlighet for feil av type II lik $\beta(p_1) = B(k-1; n, p_1)$.
- Tester av typen $H_0 : p \geq p_0$ mot $H_a : p < p_0$ finner en på samme måte.
- En forkaster da H_0 ved signifikansnivå α når $Y \leq k$, der k er det største heltallet slik at

$$\text{P(Feil av type I)} \leq B(k; n, p_0) \leq \alpha.$$

Små utvalg (forst.)

- For tester av typen $H_0 : p = p_0$ mot $H_a : p \neq p_0$ forkaster en H_0 ved signifikansnivå α når $Y \leq k_1$ eller $Y \geq k_2$, med $k_1 < k_2$, der de utgjør det minste forkastningsområdet slik at

$$\mathrm{P}(\text{Feil av type I}) \leq B(k_1; n, p_0) + 1 - B(k_2 - 1; n, p_0) \leq \alpha.$$

dvs. slik at

$$B(k_2 - 1; n, p_0) - B(k_1 - 1; n, p_0) \geq 1 - \alpha.$$