Statistikk

Jo Eidsvik

Matematiske fag, NTNU

Hypotesetesting for ett-utvalg

Data: $X_1, \dots, X_{n_1}$

- Normalfordelt, kjent varians : Normaltest: $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- Normalfordelt, ukjent varians : T-test: $T = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim t_{n-1}$
- Binomisk fordeling: Eksakt test, Normalapproksimasjon, med eller uten halvkorreksjon. $Z = \frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n}} \approx N(0,1)$
- Normalfordelt og test på varians σ: $\frac{(n-1)s^2}{\sigma_0^2} \sim \chi_{n-1}$

Hypotesetesting for to-utvalg

- Populasjon 1: $X_1, \ldots, X_{n_1}$
- Populasjon 2: $Y_1, \ldots, Y_{n_2}$.

H_0 : De to populasjonene kommer fra en fordeling med samme parametre.
H_1 : De to populasjonene kommer fra like fordelingsklasser men med ulike parametre.

Tre tilfeller

1. Normalfordeling $X_i \sim N(\mu_1, \sigma_1^2)$, $Y_j \sim N(\mu_2, \sigma_2^2)$. H_0: $\mu_1 = \mu_2$, kjente varianser.
2. Normalfordeling $X_i \sim N(\mu_1, \sigma_1^2)$, $Y_j \sim N(\mu_2, \sigma_2^2)$. H_0: $\mu_1 = \mu_2$, $\sigma_1 = \sigma_2 = \sigma$, ukjent varians.
3. Binomisk fordeling med indikatorvariable $X_i \sim Bin(p_1, 1)$. $Y_j \sim Bin(p_2, 1)$. H_0: $p_1 = p_2$.

1 gir Normaltest. 2 gir t-test (frihetsgrader kan være vanskelig). 3 gir approksimativ normaltest. For tilfelle 1 og 2 har vi i tillegg parring. Der $n_1 = n_2$, og $d_i = X_i - Y_i$. H_0: $\mu_d = 0$.

Eksempel binomisk

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \approx N(0,1) \tag{1}$$

Under H_0: $p_1 = p_2$.

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \approx N(0,1) \tag{2}$$

Forkaster om Z er signifikant større enn $\alpha/2$ percentiler i $N(0,1)$.

Eksempel: Kirurgi, infeksjoner (Josepsh Lister, tidlig 19hundre)

$n_1 = 40 =$ antall amputasjoner utført med desinfeksjonsmiddel. Andel overlevende $\hat{p}_1 = 34/40 = 0.85$.
$n_2 = 35 =$ antall amputasjoner utført uten desinfeksjonsmiddel. Andel overlevende $\hat{p}_2 = 19/35 = 0.54$.
H_0: $p_1 = p_2$. H_1: $p_1 > p_2$. Signifikansnivå $\alpha = 0.01$

Eksempel: Kirurgi, infeksjoner (Joseph Lister, tidlig 19hundre)

Under H_0: Forkast dersom $Z > z_{0.01} = 2.33$. Vi får:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = \frac{0.85 - 0.54}{\sqrt{\frac{0.85 \cdot 0.15}{40} + \frac{0.54 \cdot 0.46}{35}}} = 3.05 \qquad (3)$$

Forkast H_0. Desinfeksjonsmiddel hjelper under kirurgiske inngrep.
P-verdi: $P(N(0,1) > 3.05) = 0.0011$. Liten. Forkast H_0.