

## Val av tal på observasjonar $n$ (kap. 10.6)

4. april 2018

Me vil fortsetje eksempelet frå forelesinga der me ynskjer å sjå på verkegraden til ein ny medisin i eit tilfeldig utval. Eksisterane medisin veit ein at verkar i 70% av høva, medan produsenten av den nye medisinen hevdar at den nye medisinen verkar i meir enn 70% av høva. Me ynskjer altså å utføre hypotesetesten

$$H_0 : p = p_0 = 0.7 \quad \text{mot} \quad H_1 : p > 0.7.$$

La  $X$  vere talet på pasientar der den nye medisinen har ein positivt effekt,  $X$  er altså binomisk fordelt med parametre  $n$  og  $p$ . Anta no at me tar del i planleggjinga av forsøket og kan bestemme talet på målingar me skal gjere, det vil seie at me kan bestemme  $n$ . Merk at  $p$  er eit ukjend tal som me er interessert i å anslå.

Ein rimeleg estimator for  $p$  er  $\hat{p} = X/n$ . Frå tidlegare (teorem 6.3 i læreboka) er det kjend at

$$\frac{X - np}{\sqrt{np(1-p)}} \approx n(z; 0, 1) \quad \text{alltid.} \quad (1)$$

Merk: denne standardiseringa er gjort med hensyn på den sanne verdien til  $p$  (som er ukjend).

Testobservatoren me har nytta tidlegare i forelesinga (når  $n$  er kjend) er:

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} \approx n(z; 0, 1) \quad \text{når } H_0 \text{ er riktig.} \quad (2)$$

Merk:  $Z$  er kun approksimativt standard normalfordelt dersom  $p = p_0$  er korrekt (altså når  $H_0$  er riktig).

Me forkastar  $H_0$  dersom

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} > z_\alpha$$

Me må fiksere/velje  $\alpha$  slik at

$$P(\text{Type I-feil}) = P(\text{forkast } H_0 \text{ når } H_0 \text{ riktig}) \leq \alpha$$

er liten. Itillegg ynskjer me å velje  $n$  slik at me kontrollerer

$$P(\text{Type II-feil}) = P(\text{ikkje forkast } H_0 \text{ når } H_1 \text{ riktig}) \leq \beta.$$

Det er ikkje særskild avlvorleg å gjere ein type II-feil dersom den sanne verdien til  $p$  er 0.701, men dersom den sanne verdien er  $p = 0.85$  er det meir alvorleg å gjere ein type II-feil.

Me vel derfor  $\delta > 0$  og ser på  $p = p_0 + \delta$ . Me vil kontrollere sannsynet for type II-feil, ynskjer altså å krevje  $n$  utifrå:

$$\begin{aligned} P(\text{ikkje forkast } H_0 \text{ når } p = p_0 + \delta) &\leq \beta \\ P\left(\frac{X - np_0}{\sqrt{np_0(1-p_0)}} \leq z_\alpha \text{ når } p = p_0 + \delta\right) &\leq \beta \end{aligned}$$

Merk: her er **IKKJE**  $H_0$  riktig og derfor er **IKKJE** uttrykket standard normalfordelt. Dette sidan (2) berre gjeld for  $p = p_0$ . Me isolerer no  $X$  for seg sjølve og får

$$P\left(X \leq np_0 + z_\alpha \sqrt{np_0(1-p_0)} \text{ når } p = p_0 + \delta\right) \leq \beta.$$

Me kan no nytte (1) til å få eit uttrykk som er standard normalfordelt:

$$P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{np_0 + z_\alpha \sqrt{np_0(1-p_0)} - np}{\sqrt{np(1-p)}} \text{ når } p = p_0 + \delta\right) \leq \beta$$

Dersom me sett inn  $p = p_0 + \delta$  får me

$$P\left(Z \leq \frac{np_0 + z_\alpha \sqrt{np_0(1-p_0)} - n(p_0 + \delta)}{\sqrt{n(p_0 + \delta)(1-(p_0 + \delta))}} \text{ når } p = p_0 + \delta\right) \leq \beta$$

$$P\left(Z \leq \frac{z_\alpha \sqrt{np_0(1-p_0)} - n\delta}{\sqrt{n(p_0 + \delta)(1-(p_0 + \delta))}} \text{ når } p = p_0 + \delta\right) \leq \beta$$

der  $Z$  er standard normalfordelt. Me får altså følgande krav

$$\frac{z_\alpha \sqrt{np_0(1-p_0)} - n\delta}{\sqrt{n(p_0 + \delta)(1-(p_0 + \delta))}} \leq -z_\beta$$

og løyser ulikskapen med hensyn på  $n$ :

$$\frac{z_\alpha \sqrt{p_0(1-p_0)} - \sqrt{n}\delta}{\sqrt{(p_0 + \delta)(1-(p_0 + \delta))}} \leq -z_\beta$$

$$-\sqrt{n}\delta \leq -z_\beta \sqrt{(p_0 + \delta)(1-(p_0 + \delta))} - z_\alpha \sqrt{p_0(1-p_0)} \quad . \quad (3)$$

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

Merk at  $n$  må vere eit heiltal, så me runder alltid opp til næraste heiltal.

La oss no anta at me vel å utføre testen med  $\alpha = 0.05$  som signifikansnivå, då er  $z_\alpha = 1.645$ . Tilsvarande krev me at dersom  $\delta = 0.15$  vil me at sannsynet for å gjere ein type II-feil skal vere høgst 5%, det vil seie at  $\beta = 0.05$  og  $z_\beta = 1.645$ . Ved å setje inn tala i (3) får me at  $n \geq 55$ , det vil seie at me må gjere minst  $n = 55$  observasjonar.