

Data-analyysi R-ohjelmistolla

Tommi Härkänen

Terveiden ja hyvinvoinnin laitos (THL), Helsinki
E-mail: tommi.harkanen@helsinki.fi

Helsingin yliopisto, 8.4.2014

Hypoteesit

- ▶ Tutkija muodostaa tutkimuskysymyksensä perusteella väittämiä tai käsityksiä (hypoteesit).
- ▶ Tutkija hankkii sopivan havaintoaineiston.
- ▶ **Kysymys** Tukeeko havaintoaineisto tutkijan hypoteesia?
- ▶ Esim. lantinkeitto ja kaksisuuntainen testi:
 - Nollahypoteesi H_0 "Kruunan todennäköisyys on 0.5" on yksinkertainen yksinkertainen hypoteesi ts. yksi piste $p = 0.5 \in \Theta = [0, 1]$.
 - Vaihtoehtohypoteesi H_1 "Kruunan todennäköisyys ei ole 0.5", $p \neq 0.5$.

Sisältö

Tilastolliset testit

Testin voima

Erilaisia testejä

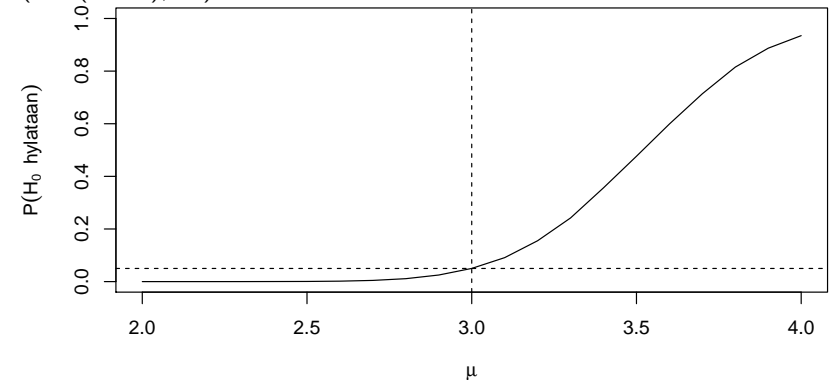
Yhdistetty hypoteesi

Yksisuuntainen testi, normaalijakauma, σ tunnettu

Oletetaan $\mu_0 = 3$, $\sigma = 1$, $n = 10$ ja $\alpha = 0.05$.

Hypoteesit: $H_0 : \mu \leq 3$, $H_1 : \mu > 3$.

Simuloidut hylkäystodennäköisyydet, kun kriittinen alue on $(\Phi^{-1}(1 - \alpha), \infty)$:



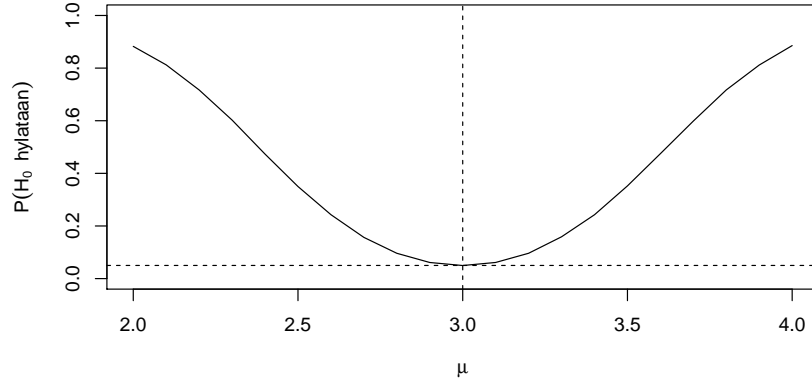
Yhdistetty hypoteesi

Kaksisuuntainen testi, normaalijakauma, σ tunnettu

Oletetaan $\mu_0 = 3$, $\sigma = 1$, $n = 10$ ja $\alpha = 0.05$.

Hypoteesit: $H_0 : \mu = 3$, $H_1 : \mu \neq 3$.

Simuloidut hylkäystodennäköisyydet, kun kriittinen alue on $(-\infty, \Phi^{-1}(\alpha/2)) \cup (\Phi^{-1}(1 - \alpha/2), \infty)$:



Päätös

Todellisuus	H_0 hyväksytään	H_0 hylätään
H_0 tosi	oikea päätös	hylkäämisvirhe I lajin virhe
H_0 epätosi	hyväksymisvirhe II lajin virhe	oikea päätös

Table : Testaamiseen liittyviä virheitä.

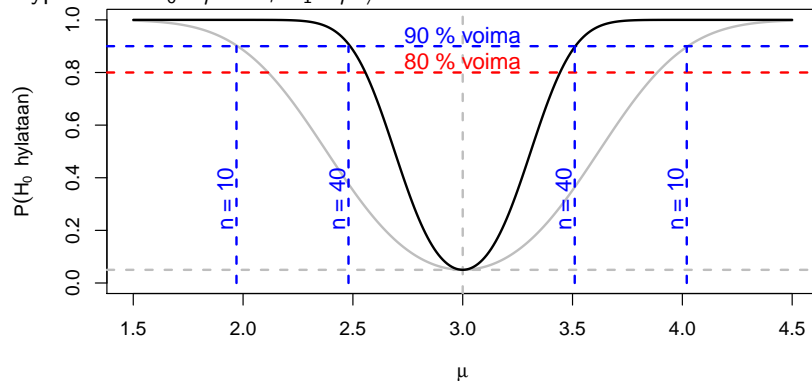
- ▶ I lajin virheen todennäköisyys yleensä kiinnitetään (esim. $\alpha = 0.05$).
- ▶ II lajin virheen todennäköisyys riippuu
 - ▶ otoskooosta n
 - ▶ oikean ja nollahypoteesin mukaisten parametriarvojen erotuksesta $\theta - \theta_0$.
- ▶ Usein voimalaskelmia käytetään koesuunnittelussa: "Miten suuri otoskoko tarvitaan tietyn eron saamiseksi esiin?"
 Suurempi otoskoko paljastaa eron varmemmin, mutta lisää tutkimuksen kustannuksia.

Yhdistetty hypoteesi

Kaksisuuntainen testi, normaalijakauma, σ tunnettu

Oletetaan $\mu_0 = 3$, $\sigma = 1$, $n = 10$ ja $\alpha = 0.05$.

Hypoteesit: $H_0 : \mu = 3$, $H_1 : \mu \neq 3$.



Voimalaskelmia R-ohjelmistolla

stats-paketista löytyvät funktiot `power.t.test`, `power.prop.test`, `power.anova.test` t-testille, binomikokeelle ja varianssianalyysille, esim.

```
power.t.test(n = NULL, delta = NULL, sd = 1,
  sig.level = 0.05, power = NULL,
  type = c("two.sample", "one.sample", "paired"),
  alternative = c("two.sided", "one.sided"),
  strict = FALSE)
```

Yksisuuntainen testi esim.

```
power.t.test(power = .90, delta = 1, alternative =
  "one.sided")
```

Muita menetelmiä löytyy CRAN / Task Views -kohdasta *Clinical Trial Design, Monitoring, and Analysis*

Voimalaskelmia R-ohjelmistolla

pwr-paketilla tehdä voimalaskelmat esim. binomi- ja normaalijakaumakokeille yhdellä tai kahdella otoksella:

<code>wr.p.test</code>	test for one proportion (ES=h)
<code>pwr.2p.test</code>	test for two proportions (ES=h)
<code>pwr.2p2n.test</code>	test for two proportions (ES=h, unequal sample sizes)
<code>pwr.t.test</code>	one sample and two samples (equal sizes) t tests for means (ES=d)
<code>pwr.t2n.test</code>	two samples (different sizes) t test for means (ES=d)
<code>pwr.anova.test</code>	test for one-way balanced anova (ES=f)
<code>pwr.r.test</code>	correlation test (ES=r)
<code>pwr.chisq.test</code>	chi-squared test (ES=w)
<code>pwr.f2.test</code>	test for the general linear model (ES=f2)

Käyttö esim.

```
library(pwr)
## t-jakauma, kahden ryhmän ero delta=1, keskihajonta sd=2:
effect.size <- 1 / 2
pwr.t.test(d=effect.size, power=.90, sig.level=.05, type="two.sample")
```

Tällä kurssilla esiteltäviä testejä

- ▶ Luokitellut vs. jatkuvat muuttujat
- ▶ Parametriset vs. parametrittomat testit
- ▶ Yksi, kaksi tai useampia otoksia
- ▶ Riippumattomat vs. riippuvat otokset