Tilastollinen testaaminen

- Seuraavaksi tutustumme tilastollisiin testeihin ja niihin liittyviin peruskäsitteisiin
- Esittelemme aluksi hypoteesit sekä testisuureet ja puhumme p-arvosta (eli havaitusta merkitsevyystasosta)

Mistä on kyse?

- Tilastolliset testien avulla otetaan kantaa tutkittavaa ilmiötä koskeviin väitteisiin eli hypoteeseihin
- Esimerkkinä sopivasta kysymyksestä: onko lantti harhaton eli "reilu"?

Esimerkki 1

- Kaverit A ja B heittää lanttia 100 kertaa ja 58 kertaa tuli kruuna
- Lantti on kaverin A ja A sanoo B:lle, että lantti on reilu
- Mitä voimme sanoa A:n väitteestä? Kumpi olisi parempi selitys havainnolle

  H_0: A puhuu totta
  H_1: A puhuu vaihtoehtoista totuutta.
- Jos lanttia heitetään vielä 900 kertaa (yht. 1000 heittoa) ja saadaankin yht. 580 kruunua, niin mitä tällöin voimme sanoa A:n väitteestä?

Esimerkki 2

- Tehtaassa kone annostelee jotain nestettä kanistereihin.
- Oletus: nesteen määrä (litroissa) kanisterissa noudattaa normaalijakaumaa $N(\mu, \sigma^2)$ (onko realistinen oletus? :)
- Tavoite: Keskimäärin kanisterissa on 10 litraa ja keskihajonta olisi korkeintaan $0,2$ litraa
- Testataan 25 kanisterin nesteen määrä, ja havaitaan että $\overline{y} = 9,9$ ja $s = 0,22$
- Ei ihan mennyt tavoitteeseen, joten pitäisikö konetta säätää?

Esimerkki 3

- `https://www.uusisuomi.fi/kotimaa/77312-kimble-ilmio-nyt-todistettu-mullistava-tulos`

Kimble-ilmiö on nyt todistettu: ”Mullistava tulos”

Jaa artikkeli:

Facebook Twitter WhatsApp

Uusi Suomi

US

Uusi Suomi

Luotu: 2.2.2015 08:54

Esimerkki 3

- http://www.iltalehti.fi/uutiset/2015020219132139_uu.shtml

Tutkijat löysivät yllätyksen Kimblestä

Maanantai 2.2.2015 klo 20.54

Tilastotieteellisen testin perusteella Kimblen arpakuution annissa on ennustettavia piirteitä.

Esimerkki 3

- http://www.iltalehti.fi/uutiset/2015020519149136_uu.shtml

Peliyhtiö myöntää: Kimblen noppakohu on totta - kehitteillä uusi versio?

Torstai 5.2.2015 klo 12.12

Valmistaja myöntää sen viimein: Kimblen nopan liikkeissä on ennustettavia piirteitä.

Esimerkki 3

- http://statistition.com/?p=440

Kimblen noppa ei ole täysin satunnainen

Posted by : Statisticko On : 30.1.2015 *29*

Esimerkki 3

- http://blog.jasonknight.us/2013/07/statistical-trouble-with-trouble-board.html?m=1

2013-07-12

(Statistical) Trouble with Trouble (the board game)

Peruskäsitteet

- Lähtökohta
- Hypoteesit (H_0/H_1) (yksinkertainen/yhdistetty)
- Testauksen tavoite
- Testisuureet (t-testi, ...)
- Kriittinen alue ja päätöksen tekeminen
- Merkitsevyystaso
- Havaittu merkitsevyystaso eli p-arvo

Hypoteesit ja kysymyksenasettelu

- *Hypoteesilla* tarkoitetaan *väitettä*, joka koskee mallin parametria ja voidaan kirjoittaa muotoon $\boldsymbol{\theta} \in \Theta'$, jossa Θ' on jokin Θ:n epätyhjä osajoukko
- Hypoteesi on *yksinkertainen*, jos Θ' on yksiö ja *yhdistetty* muuten.

Nollahypoteesi ja vastahypoteesi

- *Nollahypoteesi* H_0 on väite

$$H_0: \ \boldsymbol{\theta} \in \Theta_0$$

  jonka paikkaansapitävyyden arviointi on testin tavoitteena
- Joskus halutaan päättää, että pitäisikö H_0 hyväksyä vai hylätä
- nollahypoteesi kuvaa *neutraalia* tai *oletusarvoista* tilannetta, joten erityisesti sen hylkäämistä väärin perustein tulee välttää

Nollahypoteesi ja vastahypoteesi

- *Vaihtoehtoinen hypoteesi* eli *vastahypoteesi* H_1 on väite

  $$H_1: \ \boldsymbol{\theta} \in \Theta_1$$

  jonka paikkaansapitävyyden arviointi on testin tavoitteena
- Aina $\Theta_0 \cap \Theta_1 = \emptyset$ ja usein (muttei aina) $\Theta = \Theta_0 \cup \Theta_1$.
- Jos vastahypoteesi asetetaan, arvioidaan pitäisikö H_1 hyväksyä nollahypoteesin H_0 sijaan, eli jos H_0 päätetään hylätä ollaan hyväksymässä H_1

Testin vaiheet yleisesti

- Jotta kysymystä voitaisiin tarkastella tilastollisesti, formuloidaan koeasetelmaa kuvaava malli
- Testin vaiheet ovat yleisesti:
  1. Asetetaan nollahypoteesi H_0 ja mahdollisesti vastahypoteesi H_1.
  2. Valitaan käyttävä testisuure
  3. Lasketaan havaittua aineistoa vastaava testisuureen arvo sekä havaittu merkitsevyystaso eli p-arvo
  4. Tehdään johtopäätökset

Havaittu merkitsevyystaso

- Oletus: käytämme testisuuretta t, jonka suuret arvot ovat H_0:lle kriittisiä
- Olkoon $T = t(\mathbf{Y})$ sitä vastaava sm ja $\mathbf{y}$ havaittu aineisto.

Määritelmä

Testin *p-arvo* eli *havaittu merkitsevyystaso* p on

$$p = p(\mathbf{y}) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T \geq t(\mathbf{y})).$$

Havaittu merkitsevyystaso

- Oletus: käytämme testisuuretta t, jonka pienet arvot ovat H_0:lle kriittisiä
- Olkoon $T = t(\mathbf{Y})$ sitä vastaava sm ja $\mathbf{y}$ havaittu aineisto.

Määritelmä

Testin *p-arvo* eli *havaittu merkitsevyystaso* p on

$$p = p(\mathbf{y}) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T \leq t(\mathbf{y})).$$

Havaittu merkitsevyystaso

- Oletus: käytämme testisuuretta t, jonka suuret poikkeamat arvosta $t_0 \in \mathbb{R}$ ovat H_0:lle kriittisiä
- Olkoon $T = t(\mathbf{Y})$ sitä vastaava sm ja $\mathbf{y}$ havaittu aineisto.

Määritelmä

Testin *p-arvo* eli *havaittu merkitsevyystaso* p on

$$p = p(\mathbf{y}) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(|\, T - t_0 \,| \geq |\, t(\mathbf{y}) - t_0 \,|).$$

Tavallisia merkitsevyystasoja

- p-arvon avulla, kun testin merkitsevyystaso $\alpha \in (0, 1)$, on valittu, voidaan tehdä päätöksiä:

$$\mathbf{y} \mapsto \begin{cases} H_0 \text{ hylätään,} & \text{kun} \quad p(\mathbf{y}) < \alpha \\ H_0 \text{ jää voimaan,} & \text{kun} \quad p(\mathbf{y}) > \alpha \end{cases}$$

- Tavallisesti käytetään seuraavia merkitsevyystasoja:

$$\alpha = \begin{cases} 0.05 & \textit{melkein merkitsevä} \\ 0.01 & \textit{merkitsevä} \\ 0.001 & \textit{erittäin merkitsevä} \end{cases}$$

- Higgsin bosonin kohdalla merkitsevyystasona α käytettiin lukua $\alpha = 0.0000003$ (yksisuuntaiselle testille)

Päätökset ja niiden seuraukset

- Hypoteesin testauksen päätöspeli voidaan esittää taulukkona

	Hyväksytään H_0	Hylätään H_0
H_0 totta	Päätös ok :)	Hylkäämisvirhe!!
H_0 ei totta	Hyväksymisvirhe	Päätös ok :)

- Hylkäämisvirhettä nimitään usein *I lajin* ja hyväksymisvirhettä *II lajin virheeksi*, koska hylkäämisvirhettä pidetään vakavampana
- Hylkäämisvirhettä voidaan ajatella analogisesti "syytön tuomitaan" -päätöksen ja hyväksymisvirhettä "syyllinen vapautetaan" -päätöksen kanssa.

Merkitsevyystaso ja hylkäämisvirheen tn

- Hylkäysvirhettä on syytä välttää, joten kuinka todennäköistä sen tekeminen on?
- Havaitsemme, että merkitsevyystasolla α hylkäämme H_0:n virheellisesti todennäköisyydellä

$$\mathbb{P}_{H_0}(\text{hylätään } H_0) = \mathbb{P}_{H_0}(p(\mathbf{Y}) < \alpha) \leq \alpha$$

- Eli: merkitsevyystaso α on hylkäämisvirheen todennäköisyyden yläraja (katso Määritelmä 6.1)

Havaitun merkitsevyystason tulkinnasta

- p-arvo on siis yläraja H_0:n pätiessä sille tn:lle, että satunnaismuuttuja T saa arvon, joka on yhtä suuri tai suurempi kuin nyt havaittu arvo $t(\mathbf{y})$.
- hyvin pieni p-arvo todistaa H_0:aa vastaan (ja voi jopa johtaa hylkäykseen), ei ihan pieni p-arvo taas on sopusoinnussa H_0:n kanssa
- Huom. p-arvon laskeminen edellyttää sitä, että testisuuretta vastaavan satunnaismuuttujan T jakauma hallitaan ainakin kaikilla nollahypoteesiarvoilla $\theta \in \Theta_0$

Havaitun merkitsevyystason tulkinnasta

- on siis eduksi, että testisuureen jakauma on hyvin tunnettu ainakin nollahypoteesin H_0 ollessa voimassa
- Käytännössä p-arvon laskennassa joudutaan usein turvautumaan approksimatiivisiin jakaumatuloksiin ja niistä saataviin likiarvoihin (siksi normaalijakauman testit ovat keskeisiä)

P-arvo ei puhu H_0:n todennäköisyydestä

- Huom. joskus saattaa törmätä ajatukseen, että p-arvo on nollahypoteesin todennäköisyys, ja lyhyesti: tästä ei ole kyse
- Frekventistisesti väitteeseen $\{ \theta \in \Theta_0 \}$ emme luonnolisestikaan liitä mitään tn-tulkintaa
- Bayesiläisittäin ajateltunakaan p-arvo ei puhu väitteen $\{ \theta \in \Theta_0 \}$ todennäköisyydestä vaan testisuureen häntätodennäköisyydestä, kun H_0 oletetaan.

Väärinkäsityksistä

- tilastotieteen soveltajille on syntynyt sellainen mielikuva, että kokeellisen tutkimuksen päämääränä on laskea p-arvo jollekin testille
- julkaisuharha (engl. *publication bias*): julkaistaan yleensä vain pieniä p-arvoja saaneita tutkimuksia, jotka hylkäävät nollahypoteesin riippumatta asian todellisesta tilasta
- Hölmöt nollahypoteesit (engl. *silly null*) joista jo etukäteen tiedetään, että ne eivät voi paikkaansa. Esim. liian tarkat nollahypoteesit hylätään taatusti, kunhan otoskoko on riittävän iso.

Väärinkäsityksistä

- Kysymys: jos $H_0: \mu = 0$ ja riittävän suurella otoskoolla se saadaan hylättyä, mutta su-estimaatti on 0.1, niin onko tällä käytännössä merkitystä?
- Tämä vaatii alan asiantuntijatietoa, tilastotiede ei pysty tähän vastaamaan (mitä mitattiin? mitä yksiköt ovat? jne.)

Yhteenveto

- Nollahypoteesin hyväksyminen testissä ei tarkoita, että oltaisiin löydetty todisteita nollahypoteesin puolesta vaan ettei ole löydetty riittävän painavia todisteita nollahypoteesia vastaan
- Testin tekemän päätöksen voi lukea p-arvosta, joka mittaa kuinka hyvin aineisto on sopusoinnussa nollahypoteesin kanssa
- Testin p-arvo ei ole todennäköisyys sille, että nollahypoteesi pitää paikkansa
- Testaamisen sijasta kannattaa (jos mahdollista) laskea piste-estimaatteja ja luottamusvälejä (kvantifioi epävarmuuden selkeällä tavalla)
- Testaaminen on mainio työkalu, mutta vain yksi useista tilastollisen päättelyn työkaluista

Z-testi

Taulukko 6.1 Merkitsevyystason $0 < \alpha < 1$ testejä ja luottamusvälejä normaalijakautuneelle populaatiolle $N(\mu, \sigma^2)$, kun varianssi σ^2 on tunnettu. Tässä $z = (\bar{y} - \mu_0)/(\sigma/\sqrt{n})$, ja $Z \sim N(0,1)$ ja z_u on $N(0,1)$-jakauman u-yläkvantiili.

H_0	H_1	Hylkäysalue	p-arvo	Luottamusväli
$\mu \leq \mu_0$	$\mu > \mu_0$	$z > z_\alpha$	$P(Z \geq z)$	$[\bar{y} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$z < -z_\alpha$	$P(Z \leq z)$	$(-\infty, \bar{y} + z_\alpha \frac{\sigma}{\sqrt{n}}]$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert z\rvert > z_{\alpha/2}$	$P(\lvert Z\rvert \geq \lvert z\rvert)$	$[\bar{y} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}]$

T-testi

Taulukko 6.2 Merkitsevyystason $0 < \alpha < 1$ testejä ja luottamusvälejä normaalijakautuneen populaation $N(\mu, \sigma^2)$ odotusarvolle μ, kun myös varianssi σ^2 on tuntematon. Tässä $t = (\bar{y} - \mu_0)/(s/\sqrt{n})$, s on otoskeskihajonta, $T \sim t_{n-1}$ ja $t_{n-1}(u)$ on t_{n-1}-jakauman u-yläkvantiili.

H_0	H_1	Hylkäysalue	p-arvo	Luottamusväli
$\mu \leq \mu_0$	$\mu > \mu_0$	$t > t_{n-1}(\alpha)$	$P(T \geq t)$	$[\bar{y} - t_{n-1}(\alpha)\frac{s}{\sqrt{n}}, \infty)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$t < -t_{n-1}(\alpha)$	$P(T \leq t)$	$(-\infty, \bar{y} + t_{n-1}(\alpha)\,\frac{s}{\sqrt{n}}]$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert t\rvert > t_{n-1}(\frac{\alpha}{2})$	$P(\lvert T\rvert \geq \lvert t\rvert)$	$[\bar{y} - t_{n-1}(\frac{\alpha}{2})\,\frac{s}{\sqrt{n}}, \bar{y} + t_{n-1}(\frac{\alpha}{2})\,\frac{s}{\sqrt{n}}]$