MS-A0502 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

6A Tilastollisen merkitsevyyden testaaminen

Jukka Kohonen

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2022–2023
Periodi II

Sisältö

Mustekala Paul

Jalkapallon MM-kisoissa 2010 Paul ennusti voittajan oikein jokaiselle Saksan ottelulle.

Opponent	Tournament	Stage	Date	Prediction	Result	Outcome
Poland	Euro 2008	group stage	8 June 2008	Germany	2–0	Correct
Croatia	Euro 2008	group stage	12 June 2008	Germany[3][20]	1–2	Incorrect
Austria	Euro 2008	group stage	16 June 2008	Germany	1–0	Correct
Portugal	Euro 2008	quarter-finals	19 June 2008	Germany	3–2	Correct
Turkey	Euro 2008	semi-finals	25 June 2008	Germany	3–2	Correct
Spain	Euro 2008	final	29 June 2008	Germany[3]	0–1	Incorrect
Australia	World Cup 2010	group stage	13 June 2010	Germany[31]	4–0	Correct
Serbia	World Cup 2010	group stage	18 June 2010	Serbia[31]	0–1	Correct
Ghana	World Cup 2010	group stage	23 June 2010	Germany[31]	1–0	Correct
England	World Cup 2010	round of 16	27 June 2010	Germany[32]	4–1	Correct
Argentina	World Cup 2010	quarter-finals	3 July 2010	Germany[23]	4–0	Correct
Spain	World Cup 2010	semi-finals	7 July 2010	Spain[33]	0–1	Correct
Uruguay	World Cup 2010	3rd place play-off	10 July 2010	Germany	3–2	Correct

Onko poikkeuksellisen hyvä ennustustulos tilastollisesti merkitsevä, vai voidaanko se lukea tavanomaisen satunnaisvaihtelun piiriin?

https://en.wikipedia.org/wiki/Paul_the_Octopus

Merkitsevyystestaus vs. Bayes-päättely

Aiemmilla luennoilla opittiin, miten tuntemattomalle parametrille θ saadaan täysi posteriorijakauma, jos on käytettävissä:

- priori $f(\theta)$ — mitkä θ:n arvot ovat alun perin todennäköisiä
- uskottavuus $f(\vec{x} \mid \theta)$ — stokastinen malli, joka kuvaa miten data syntyy, jos θ:lla on tietty arvo

Entä, jos mitään priorijakaumaa ei osata tai haluta muotoilla? Voidaanko tehdä päätelmiä vain **datasta ja uskottavuusfunktiosta?**

Voidaan tehdä **jotain**. Voidaan tarkastella **θ:n tiettyä arvoa**, ja **hylätä** se, jos se tekisi havaitusta datasta "liian epätodennäköistä".

Näin päädytään klassiseen *hypoteesintestaukseen* jota käsitellään tällä luennolla. (Tämä on siis vaihtoehto Bayes-päättelylle.)

Hypoteesin testaus — ensimmäinen idea (ei hyvä)

Oletetaan esim. stokastinen malli $X \sim \text{Bin}(1000, \theta)$ (tuhat kolikonheittoa), mutta parametria θ ei tiedetä. Pohdimme voisiko olla $\theta = 0.5$, vai tekeekö se havaitun datan liian epätodennäköiseksi.

Esim 1. Havaitaan $x = 510$ kruunaa. Jos $\theta = 0.5$, niin

$$\mathbb{P}(X = 510 \,|\, \theta = 0.5) = \binom{1000}{510} 0.5^{510} \, 0.5^{490} \approx 2.1\%.$$

Onko tämä liian yllättävää, pitäisikö $\theta = 0.5$ hylätä?

Esim 2. Havaitaan $x = 500$ kruunaa. Jos $\theta = 0.5$, niin

$$\mathbb{P}(X = 500 \,|\, \theta = 0.5) = \binom{1000}{500} 0.5^{500} \, 0.5^{500} \approx 2.5\%.$$

Onko tämä liian yllättävää, pitäisikö $\theta = 0.5$ hylätä? Ei kai!

Hypoteesin testaus — klassinen menetelmä

Askel	Esimerkki
Muotoile hypoteesi H_0 siitä, miten data syntyy.	$\vec{X} = 30$ kahvikupillista $\mathrm{Nor}(10, 3^2)$
Muotoile a tunnusluku $t = t(\vec{X})$, joka lasketaan datasta	Otoksen keskiarvo $m(\vec{X})$
Päättele t:n jakauma (jos H_0 on totta).	$m(\vec{X}) \sim \mathrm{Nor}(\ldots)$
Hylkää H_0 jos *havaittu* arvo $t(\vec{x})$ osuu jakauman häntään; määritellään hännät niin että niissä on yhteensä tn α	$\alpha = 0.05$

Idea: Jos H_0 on totta, niin on (hiukan) *yllättävää* jos t osuu jakauman 5% kokoiseen häntään. Joten päätetään silloin "hylätä" H_0. Häntiä kutsutaan testin hylkäysalueeksi.

Huom. Jos H_0 on totta, tämä menettely johtaa silti H_0:n hylkäämiseen (noin) 5%:ssa tapauksista. Arvoa 5% kutsutaan testin merkitsevyystasoksi.

Hypoteesin testaus — toinen tapa p-arvolla

Askel	Esimerkki
Muotoile hypoteesi H_0 siitä, miten data syntyy.	$\vec{X} = 30$ kahvikupillista $\mathrm{Nor}(10, 3^2)$
Muotoile a tunnusluku $t = t(\vec{X})$, joka lasketaan datasta	Otoksen keskiarvo $m(\vec{X})$
Päättele t:n jakauma (jos H_0 on totta).	$m(\vec{X}) \sim \mathrm{Nor}(\ldots)$
Laske arvoa $t(\vec{x})$ vastaavat häntä-tn:n molemmilla puolilla yhteensä. Tämä on p-arvo	$p = 0.018$
Hylkää H_0 jos $p < \alpha$	hylätään koska $0.018 < 0.05$

Tulos on sama kuin edellä, mutta nyt p-arvo kertoo *millä* merkitsevyystasolla testitulos vielä juuri ja juuri hyväksyttäisiin.

Sisältö

Kahviautomaatti, normaalimalli

Kahviautomaatin on tarkoitus laskea jokaiseen kuppiin keskimäärin 10.0 cl kahvia. Oletamme kahvimäärät normaalijakautuneiksi tuntemattomalla odotusarvolla. Testausta varten valutettiin 30 kupillista ja mitattiin kahvimäärät (cl):

11.05 9.65 10.93 9.46 10.27 10.02 10.07 10.74 11.15 10.40 10.12 11.20 10.07 10.27 9.99 9.80 10.83 10.21 11.26 10.11 10.49 10.10 10.15 11.02 10.00 11.68 10.51 11.20 11.29 10.15

Onko kahviautomaatti oikein kalibroitu?

Mittausdatan $\vec{x}$ keskiarvo on $m(\vec{x}) = 10.473$, joka poikkeaa tavoitearvosta $\mu_0 = 10.0$.

Toisaalta on odotettavissakin, että keskiarvo poikkeaa odotusarvosta ainakin vähän. Poikkeaako se liikaa?

Kahviautomaatti, normaalimalli

11.05 9.65 10.93 9.46 10.27 10.02 10.07 10.74 11.15 10.40 10.12 11.20 10.07 10.27 9.99 9.80 10.83 10.21 11.26 10.11 10.49 10.10 10.15 11.02 10.00 11.68 10.51 11.20 11.29 10.15

Datajoukon keskiarvo $m(\vec{x}) = 10.473$, keskihajonta $\mathrm{sd}(\vec{x}) = 0.563$

H_0: $\mu = 10.0$
H_1: $\mu \neq 10.0$

Havaitun datajoukon testisuure:

$$t(\vec{x}) \;=\; \frac{m(\vec{x}) - \mu_0}{\mathrm{sd}(\vec{x})/\sqrt{n}} \;=\; \frac{10.473 - 10.0}{0.563/\sqrt{30}} \;=\; 4.60$$

Koska datajoukon koko $n = 30$ on melko suuri, käytetään suuren datajoukon testiä, ts. pidetään tunnuslukua $t(\vec{X})$ standardinormaalijakautuneena.

$$\text{p-arvo} \;\approx\; \mathbb{P}(|t(\vec{X})| \geq |t(\vec{x})| \mid H_0) \;\approx\; \mathbb{P}(|Z| \geq 4.60) \;\approx\; 4.2 \times 10^{-6}$$

Johtopäätös: Hyvin pieni p-arvo puoltaa vahvasti H_0:n hylkäämistä.

Nollahypoteesi H_0

Tilastollisen merkitsevyystestin lähtökohdaksi muotoillaan nollahypoteesi H_0, joka vastaa tilannetta, jossa mitään uutta tai yllättävää ei tarvita havaintojen selittämiseen.

Useimmiten se on muotoa "parametri = arvo" (ja useimmiten tarkastellaan *odotusarvo*parametria).

Esim

H_0: Paulin ennusteet osuvat oikeaan tn:llä 0.5

H_0: Kahvimäärän odotusarvo on $\mu = 10.0$ cl kuten on tarkoitus

H_0: Uusi lääke ei ole lumelääkettä tehokkaampi

H_0: Salkunhoitajan rahaston tuotto ei ole pörssi-indeksiä parempi

Vastahypoteesi H_1 on yleensä nollahypoteesin *komplementti* eli negaatio. Jos H_0 sanoo $\mu = 10$, niin H_1 sanoo $\mu \neq 10$. Huom. vastahypoteesi ei tällöin ilmaise tiettyä tn-jakaumaa.

Testisuureen p-arvo

Havaitun datajoukon $\vec{x} = (x_1, \dots, x_n)$ poikkeuksellisuutta analysoidaan laskemalla testisuure

$$t(\vec{x}) = t(x_1, \dots, x_n),$$

joka tiivistää havaitut datapisteet yhdeksi reaaliluvuksi.

Testisuureen p-arvo on todennäköisyys, jolla nollahypoteesin mukaisen datalähteen ennakoidaan tuottavan ainakin niin poikkeavia testisuureen arvoja kuin $t(\vec{x})$.

p-arvo	**Tulkinta**
> 0.10	Havainto ei ole ristiriidassa H_0:n kanssa
≈ 0.05	Havainto todistaa jonkun verran H_0:aa vastaan
< 0.01	Havainto todistaa vahvasti H_0:aa vastaan

Hypoteesitestaus vs. luottamusväli

Hypoteesitestaus merkitsevyystasolla α voidaan (usein) muotoilla myös luottamusvälin avulla:

Jos lasketaan $1-\alpha$ luottamustason luottamusväli parametrille θ, sisältääkö luottamusväli nollahypoteesin väittämän arvon θ_0?

Jos luottamusväli sisältää θ_0:n, niin havainnot ovat (riittävän) sopusoinnussa kyseisen arvon kanssa ja nollahypoteesi hyväksytään.

Jos luottamusväli on kokonaan θ_0 on luottamusvälin ylä- tai alapuolella, havainnot sopivat huonosti hypoteesiin $\theta = \theta_0$ ja se hylätään.

Kahviautomaatti — testi vs. luottamusväli

11.05 9.65 10.93 9.46 10.27 10.02 10.07 10.74 11.15 10.40 10.12 11.20
10.07 10.27 9.99 9.80 10.83 10.21 11.26 10.11 10.49 10.10 10.15 11.02
10.00 11.68 10.51 11.20 11.29 10.15

Datajoukon keskiarvo $m(\vec{x}) = 10.473$, keskihajonta $\mathrm{sd}(\vec{x}) = 0.563$

H_0: $\mu = \mu_0 = 10.0$
H_1: $\mu \neq \mu_0 = 10.0$

Oletetaan taaskin $\sigma = \mathrm{sd}(\vec{x}) = 0.563$ ("tunnettu varianssi").

Lassketaan esim. 99% luottamusväli:

$$10.473 \pm 2.58 \cdot \frac{0.563}{\sqrt{30}} \approx 10.473 \pm 0.265,$$

joka on kokonaan arvon 10.0 yläpuolella, joten nollahypoteesi $\mu = 10$ hylätään 1% merkitsevyystasolla.

Huom: Normaalimallissa tämä on näin yksinkertaista, mutta tietyissä tapauksissa (esim. binäärimallissa) voi tulla pieniä eroja. Perusidea on kuitenkin sama.

Lisää esimerkkejä

Esimerkki (Kolikko)

Tasaiseksi väitettyä kolikkoa 50 kertaa heitettäessä saadaan 42 kruunaa.
H_0: Kruunan tn $\theta = 1/2$
H_1: Kruunan tn $\theta \neq 1/2$

Esimerkki (Kohinainen kanava)

Tiedonsiirtovirheiden väitetään olevan normaalijakautuneita parametreina $\mu = 0$ ja $\sigma = 3$. Kanavaa kerran testaamalla mitattiin virheeksi $x_1 = 4.8$.
H_0: $\mu = 0$
H_1: $\mu \neq 0$

Esimerkki (Laadunvalvonta)

Tukkukauppias väittää, että sen tomaateista enintään 5% on huonoja.
Suuresta erästä poimittiin 50 tomaattia ja niistä 7 todettiin huonoiksi.
H_0: Huonolaatuisten osuus $\theta \leq 0.05$
H_1: Huonolaatuisten osuus $\theta > 0.05$
Tässä nollahypoteesi on *yhdistetty* (sallii monta eri θ-arvoa)

Esim. Kolikko

Tasaiseksi väitetty kolikko tuottaa 42 kruunaa 50 heitolla.
H_0: Kruunan tn $\theta = 1/2$
H_1: Kruunan tn $\theta \neq 1/2$
Testisuure = kruunien lukumäärä: $t(x) = 42$
$T = t(X) =$ "kruunien lkm stokastisessa mallissa"

$$f(x) \;=\; \mathbb{P}(T = x \,|\, H_0) \;=\; \binom{50}{x} \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{50-x}$$

Testisuureen odotusarvo $t_0 = \mathbb{E}(T \,|\, H_0) = 25$.

$$\begin{aligned} \text{p-arvo} \;&=\; \mathbb{P}(|T - t_0| \geq |t(x) - t_0| \,|\, H_0) \\ &=\; \mathbb{P}(|T - 25| \geq 17 \,|\, H_0) \\ &=\; \sum_{x=0}^{8} f(x) + \sum_{x=42}^{50} f(x) \;\approx\; 1.2 \times 10^{-6}. \end{aligned}$$

Havainto todistaa vahvasti H_0:aa vastaan.

Esim. Kohinainen kanava

Tiedonsiirtovirheiden väitetään olevan normaalijakautuneita odotusarvona $\mu = 0$ ja keskihajontana 3. Havainto: $x_1 = 4.8$.
H_0: Odotusarvo $\mu = 0$
H_1: Odotusarvo $\mu \neq 0$
Testisuure = normitettu poikkeama nollahypoteesin mukaisesta odotusarvosta: $z(\vec{x}) = \frac{x_1 - 0}{3} = 1.6$

$$\text{p-arvo} \;=\; \mathbb{P}(|Z| \geq 1.6 \,|\, H_0) \;=\; 2\mathbb{P}(Z \geq 1.6 \,|\, H_0) \;\approx\; 11\%,$$

Havainto on selitettävissä tavanomaisella satunnaisvaihtelulla.
Havainto ei puhu H_0:aa vastaan.

Sisältö

Muunnelma: Testataan μ:ta, suuri epänormaali data

Oletus: Datalähde tuottaa riippumattomia, samoin jakautuneita lukuja $X_1, X_2, \ldots, X_n$ eräästä jakaumasta, jolla odotusarvo μ. Haluamme testata onko $\mu = \mu_0$.

H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$

Mitä tehdä?
Keskeisen raja-arvolauseen nojalla otoksen keskiarvo on likimain normaalijakautunut, vaikka yksittäiset havainnot eivät olisikaan.

Testisuure aivan kuten normaalimallissa:

$$t(\vec{x}) \;=\; \frac{m(\vec{x}) - \mu_0}{\mathrm{sd}(\vec{x})/\sqrt{n}}.$$

Muunnelma: Tuntematon varianssi

Usein **datalähteen** keskihajonta σ ei ole etukäteen tiedossa, vaan estimoidaan otoksen keskihajonnasta $\mathrm{sd}(\vec{x})$.

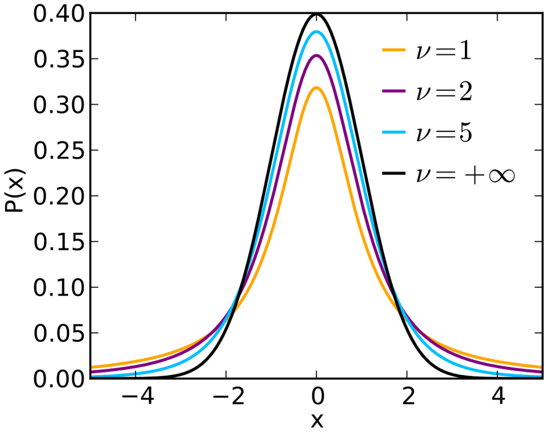
Jos otos on suuri (esim. $n > 30$), estimaatti on melko tarkka, mutta ...

Jos otos on pieni, testisuure

$$t(\vec{X}) = \frac{m(\vec{X}) - \mu_0}{\mathrm{sd}(\vec{X})\ \sqrt{n}}$$

on kahden satunnaismuuttujan osamäärä, eikä se olekaan normaalijakautunut, vaan t-jakautunut parametrilla $n - 1$. Menetelmä sama kuin edellä, mutta normaalijakauman sijasta tarkastellaan t-jakauman häntiä (taulukoista tai tietokoneella). R:ssä `pt` on kertymäfunktio ja `qt` on kvantiilifunktio. (Vrt. `pnorm` ja `qnorm`.)

Studentin t-jakauma



Välihuomio: Eri jakaumia R:ssä

jakauma	tiheys	kertymä	kvantiili	satunnaisluku
tasa	`dunif`	`punif`	`qunif`	`runif`
beeta	`dbeta`	`pbeta`	`qbeta`	`rbeta`
normaali	`dnorm`	`pnorm`	`qnorm`	`rnorm`
t-jakauma	`dt`	`pt`	`qt`	`rt`
eksponentti	`dexp`	`pexp`	`qexp`	`rexp`
...	`d...`	`p...`	`q...`	`r...`

Vrt. 0.975-kvantiileja: standardinormaali vs. t-jakauma, kun $n = 50$ tai $n = 10$.

```
> qnorm(.975)
[1] 1.959964
> qt(.975, 49)
[1] 2.009575
> qt(.975, 9)
[1] 2.262157
```

Välihuomio: Eri jakaumia Matlabissa

jakauma	tiheys	kertymä	kvantiili	satunnaisluku
tasa	`unifpdf`	`unifcdf`	`unifinv`	`unifrnd`
beeta	`betapdf`	`betacdf`	`betainv`	`betarnd`
normaali	`normpdf`	`normcdf`	`norminv`	`normrnd`
t-jakauma	`tpdf`	`tcdf`	`tinv`	`trnd`
eksponentti	`exppdf`	`expcdf`	`expinv`	`exprnd`
...	`...pdf`	`...cdf`	`...inv`	`...rnd`

Vrt. 0.975-kvantiileja: standardinormaali vs. t-jakauma, kun $n = 50$ tai $n = 10$.

```
>> norminv(.975)
ans =
   1.959963984540054
>> tinv(.975, 49)
ans =
   2.009575237129235
>> tinv(.975, 9)
ans =
   2.262157162798204
```

Muunntelma: Yhdistetty nollahypoteesi

Tukkukauppiaan väitteen mukaan sen toimittamista tomaateista enintään 5% on huonolaatuisia. Suuresta erästä poimittiin satunnaisesti 50 tomaattia ja niistä 7 todettiin huonolaatuisiksi.
H_0: Huonojen osuus $\theta \leq 0.05$
H_1: Huonojen osuus $\theta > 0.05$
Testisuure: Huonojen lkm: $t(\vec{x}) = 7$
Datalähteen tuottamien testisuureen arvojen stokastinen malli:

$$\mathbb{P}_\theta(T = t) \;=\; f_\theta(t) \;=\; \binom{50}{t} \theta^t (1-\theta)^{50-t}$$

Poikkeavuus: testisuureen arvo poikkeaa *ylöspäin* odotusarvosta

$$\mathbb{P}_\theta\Big(T - \mathbb{E}_\theta(T) \geq t(\vec{x}) - \mathbb{E}_\theta(T) \Big) \;=\; \mathbb{P}_\theta(T \geq t(\vec{x})) \;=\; \sum_{t=7}^{50} f_\theta(t). \;=\; 1 - \sum_{t=0}^{6} f_\theta(t).$$

Ongelma: tn riippuu θ:sta. Valitaan suurin tn (miksi?).

$$\text{p-arvo} \;=\; \max_{\theta \leq 0.05} \mathbb{P}_\theta(T \geq t(\vec{x})) \;=\; \mathbb{P}_{0.05}(T \geq t(\vec{x})) \;=\; \sum_{t=7}^{50} f_{0.05}(t) \;\approx\; 1.2\%$$

Nollahypoteesi (tomaattikauppiaan väite) voidaan hylätä.

Sisältö

Hyväksyminen ja hylkääminen

Laskettuasi p-arvon voisit pelkästään raportoida sen ja olla tekemättä päätöksiä, onko se "liian" pieni tai suuri.

Usein kuitenkin päätöksiä on tehtävä: pidetäänkö testin perusteella H_0:aa epätodennäköisenä (hylättävänä)? Tämä voi vaikuttaa esim. siihen, jatketaanko tutkimuksia, otetaanko lääke käyttöön ...

Päätöksentekoa varten valitaan merkitsevyystaso α (α $(0 < \alpha < 1)$.

- Jos p-value $\geq \alpha$, nollahypoteesi *hyväksytään*
- Jos p-value $< \alpha$, nollahypoteesi *hylätään*

Tyypillisesti käytetään esim. merkitsevyystasoa $\alpha = 1\%$ tai $\alpha = 5\%$.

(Huom. Tämä on hyvin karkea tapa tehdä päätöksiä. Edistyneempää olisi eksplisiittisesti mallintaa päätösten *vaikutuksia* ja niiden hyötyjä tai haittoja $\longrightarrow$ päätösteoria, kurssin ulkopuolella.)

Tyypin I ja II virheet

Teimmepä kumman tahansa päätöksen, se voi olla väärä.

		Tehty johtopäätös	
		H_0 hyväksytään	H_0 hylätään
Totuus	H_0 tosi	Oikea päätös	Hylkäysvirhe
	H_0 epätosi	Hyväksymisvirhe	Oikea päätös

Testaajan johtopäätös on aina arvaus. Hyvä arvaus on todennäköisesti oikein.

Tn-laskennan keinoin voidaan yrittää laskea, millä *todennäköisyydellä* sattuu hyväksymis- ja hylkäysvirheitä.

Tilastotieteessä käytetään myös nimiä “tyypin I virhe” (väärä hylkäys) ja “tyypin II virhe” (väärä hyväksyminen).

Testausvirheiden todennäköisyydet

$p(\vec{x})$ = testisuureen p-arvo datajoukolle $\vec{x}$
$\vec{X} = (X_1, \ldots, X_n)$ mallintaa datalähteen tuottamia arvoja ennen niiden havaitsemista $\implies$ $p(\vec{X})$ on satunnaisluku
Hylkäysvirheen todennäköisyys on

$$\mathbb{P}(H_0 \text{ hylätään} \,|\, H_0) \;=\; \mathbb{P}(p(\vec{X}) < \alpha \,|\, H_0)$$

Hyväksymisvirheen todennäköisyys on

$$\mathbb{P}(H_0 \text{ hyväksytään} \,|\, H_1) \;=\; \mathbb{P}(p(\vec{X}) \geq \alpha \,|\, H_1),$$

α	Hylkäysvirheen tn	Hyväksymisvirheen tn
Lähellä nollaa	Pieni	Suuri
Lähellä ykköstä	Suuri	Pieni

Fakta

Hylkäysvirheen tn $\leq \alpha$. *(Miksi?)*

Testausvirheiden tulkinta

Anni Aktiivi

- Käyttää merkitsevyystasoa $\alpha = 5\%$
- Hylkää useammin nollahypoteeseja
- On henkisesti varautunut siihen, että tietty osuus testien johtopäätöksistä on virheellisiä
- Tietää, että pitkällä tähtäyksellä hänen hylkäämistään nollahypoteeseista enintään 5% on virheellisesti hylätty (Hän ei kuitenkaan tiedä mitkä niistä.)
- Hyväksyy Villeä harvemmin virheellisesti nollahypoteeseja (ei tiedä miten usein)

Ville Varovainen

- Käyttää merkitsevyystasoa $\alpha = 1\%$
- Hylkää harvemmin nollahypoteeseja
- On henkisesti varautunut siihen, että tietty osuus testien johtopäätöksistä on virheellisiä
- Tietää, että pitkällä tähtäyksellä hänen hylkäämistään nollahypoteeseista enintään 1% on virheellisesti hylätty (Hän ei kuitenkaan tiedä mitkä niistä.)
- Hyväksyy Annia useammin virheellisesti nollahypoteeseja (ei tiedä miten usein)

Esim. Kolikko

Tasaiseksi väitettyä kolikkoa 10 kertaa heitettäessä havaitaan data $\vec{x} = (0,0,1,0,0,0,0,0,0,0)$. Testaa väitettä 5% merkitsevyystasolla.

> *H_0: Kruunan tn $\theta = 0.5$,*
> *H_1: Kruunan tn $\theta \neq 0.5$.*

Testisuure: $t(\vec{x})$=kruunien lkm
Testisuureen stokastinen malli: $T = t(\vec{X})$

$$f_{H_0}(t) \;=\; \mathbb{P}(T = t \mid H_0) \;=\; \binom{10}{t}\left(\frac{1}{2}\right)^{10}$$

Havainnon $\vec{x}$ p-arvo

$$p(\vec{x}) \;=\; \mathbb{P}(\,|t(\vec{X}) - 5| \geq 4 \mid H_0) \;=\; \sum_{t=0}^{1} f_{H_0}(t) + \sum_{t=9}^{10} f_{H_0}(t) \;\approx\; 2.1\%.$$

Johtopäätös: Nollahypoteesi hylätään (5% merkitsevyystasolla).
Mitä osataan sanoa virhetodennäköisyyksistä?

Esim. Kolikko — Hylkäysvirheen tn (kun H_0 on TOTTA)

Testin p-arvot testisuureen funktiona:

# kruunat	0	1	2	3	4	5	6	7	8	9	10
tn(%) jos H_0	0.10	0.98	4.39	11.72	20.51	24.61	20.51	11.72	4.39	0.98	0.10
p-arvo (%)	0.20	2.15	10.94	34.38	75.39	100.00	75.39	34.38	10.94	2.15	0.20

5% merkitsevyystasolla testin hylkäysalue on $\{0, 1, 9, 10\}$.
Hylkäysvirheen tn on

$$\mathbb{P}(t(\vec{X}) \in \{0, 1, 9, 10\} \mid H_0) \;=\; \sum_{t=0}^{1} f_{H_0}(t) + \sum_{t=9}^{10} f_{H_0}(t) \;\approx\; 2.15\%.$$

Esim. Kolikko — Hyväksymisvirheen tn (kun H_0 EPÄTOSI)

Testin p-arvot testisuureen funktiona:

# kruunat	0	1	2	3	4	5	6	7	8	9	10
tn(%) jos H_0	0.10	0.98	4.39	11.72	20.51	24.61	20.51	11.72	4.39	0.98	0.10
tn(%) jos H_1	?	?	?	?	?	?	?	?	?	?	?
p-arvo (%)	0.20	2.15	10.94	34.38	75.39	100.00	75.39	34.38	10.94	2.15	0.20

5% merkitsevyystasolla testin hyväksymisalue on $\{2, 3, \ldots, 7, 8\}$.
Hyväksymisvirheen tn on

$$\mathbb{P}(t(\vec{X}) \in \{2, 3, \ldots, 8\} \mid H_1) \; = \; ?$$

Ongelma: vastahypoteesi ($H_1 : \theta \neq 0.5$) ei määrää θ:n arvoa.
Ääritapaus $\theta \approx 0.5$ (esim. $\theta = 0.5001$), jolloin

$$\begin{aligned}\mathbb{P}(t(\vec{X}) \in \{2, 3, \ldots, 8\} \mid H_1) \; &\approx \; \mathbb{P}(t(\vec{X}) \in \{2, 3, \ldots, 8\} \mid H_0) \\ &= \; \sum_{t=2}^{8} f_{H_0}(t) \; \approx \; 97.9\%.\end{aligned}$$

Esim. Kahden tyyppisiä kolikoita

Tasaiseksi väitettyä kolikkoa 10 kertaa heitettäessä havaitaan data $\vec{x} = (0, 0, 1, 0, 0, 0, 0, 0, 0, 0)$. Tiedetään, että joko $\theta = 0.5$ tai $\theta = 0.9$. Testaa väitettä 5% merkitsevyystasolla.

> *H_0: Kruunan tn $\theta = 0.5$,*
> *H_1: Kruunan tn $\theta = 0.9$.*

Sama testisuure, sama p-arvo, sama johtopäätös (H_0 hylätään).

$$f_{H_1}(t) \;=\; \mathbb{P}(T = t \mid H_1) \;=\; \binom{10}{t} 0.9^t (1 - 0.9)^{10-t}$$

Hyväksymisvirheen tn:

$$\mathbb{P}(\, t(\vec{X}) \in \{2, 3, \ldots, 8\} \mid H_1) \;=\; \sum_{t=2}^{8} f_{H_1}(t) \;\approx\; 26\%.$$

Sisältö

Hypoteesin testaus laajemmin

Edellä testattiin *odotusarvon* hypoteesin (esim. $\theta = 1/2$ tai $\mu = 10.0$) testausta *vahvoilla yksinkertaistavilla oletuksilla* (esim. "paljon dataa, testisuure normaali").

Yleisemmin tilastotieteessä tutkitaan muitakin kysymyksiä, esim.

- Muut tunnusluvut. Esim. odotusarvo tunnetaan, mutta kysytään onko datalähteen keskihajonta $\sigma = \sigma_0$ vai jotain muuta? $\longrightarrow \chi^2$-testi
- Heikommat oletukset. Esim. vähän dataa eikä normaalia, testisuure ei olekaan normaalijakautunut. $\longrightarrow$ epäparametriset testit, tietylle jakaumalle räätälöidyt testit
- Jakauman muodon testaaminen. Haluamme testata onko data normaalijakautunutta. $\longrightarrow$ lisää testejä ...

Hypoteesin testaus laajemmin

Moniin tuntematonta datalähdettä koskeviin kyllä/ei-kysymyksiin voidaan soveltaa samaa yleistä menetelmää:

1. Muotoillaan nollahypoteesi H_0 (oletus datalähteen jakaumasta).
2. Muotoillaan testisuure, päätellään sen jakauma (jos H_0 totta).
3. Tutkitaan, sopiiko testisuureen havaittu arvo ko. jakaumaan hyvin, vai osuuko sen häntään.

Yksityiskohdat vaihtelevat mutta perusidea on sama. Näistä lisää esim. kurssilla Statistical inference MS-C1620.