

# MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

## 6A Tilastollisen merkitsevyyden testaaminen

Lasse Leskelä

Matematiikan ja systeemianalyysin laitos  
Perustieteiden korkeakoulu  
Aalto-yliopisto

Lukuvuosi 2018–2019  
Periodi IV

# Sisältö

Nollahypoteesi ja p-arvo

Yhdistetty nollahypoteesi

Testausvirheet

Odotusarvon testaaminen suurille datajoukoille

# 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 <sup>[3][20]</sup>	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 <sup>[3]</sup>	0–1	Incorrect
Australia	World Cup 2010	group stage	13 June 2010	Germany <sup>[31]</sup>	4–0	Correct
Serbia	World Cup 2010	group stage	18 June 2010	Serbia <sup>[31]</sup>	0–1	Correct
Ghana	World Cup 2010	group stage	23 June 2010	Germany <sup>[31]</sup>	1–0	Correct
England	World Cup 2010	round of 16	27 June 2010	Germany <sup>[32]</sup>	4–1	Correct
Argentina	World Cup 2010	quarter-finals	3 July 2010	Germany <sup>[23]</sup>	4–0	Correct
Spain	World Cup 2010	semi-finals	7 July 2010	Spain <sup>[33]</sup>	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](https://en.wikipedia.org/wiki/Paul_the_Octopus)

# Nollahypoteesi $H_0$

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

## Esim

$H_0$ : Selvännäkijän ennustukset eivät ole arvauksia parempia

$H_0$ : Uusi lääke ei ole lumelääkettä tehokkaampi

$H_0$ : Salkunhoitajan rahaston tuotto ei ole pörssi-indeksiä parempi

Merkitsevyydestin **vastahypoteesi  $H_1$**  on yleensä nollahypoteesin vastakohta.

## Nollahypoteesi vs. data

Kuuluuko havaittu data tyypillisen satunnaisvaihtelun piiriin vai onko syytä epäillä nollahypoteesia?

### 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 4 todettiin huonoiksi.

$H_0$ : Huonolaatuisten osuus  $\theta \leq 0.05$

$H_1$ : Huonolaatuisten osuus  $\theta > 0.05$

## 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 reaalityyppiseksi.

Testisuureen **p-arvo** on todennäköisyys, jolla nollahypoteesin mukaisen datalähteen ennakoitaan tuottavan poikkeavampia tai yhtä poikkeavia testisuureen arvoja kuin  $t(\vec{x})$ .

<b>p-arvo</b>	<b>Tulkinta</b>
$> 0.10$	Havainto ei ole ristiriidassa $H_0$ :n kanssa
$\approx 0.05$	Havainto todistaa jonkun verran $H_0$ :aa vastaan
$< 0.01$	Havainto todistaa vahvasti $H_0$ :aa vastaan

## Esim. Kolikko

Tasaisesti 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:  $z(\vec{x}) = \frac{x_1 - 0}{3} = 1.6$

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

Havainto on selitettävissä tavanomaisella satunnaisvaihtelulla.

Havainto ei puhu  $H_0$ :aa vastaan.



# Sisältö

Nollahypoteesi ja p-arvo

Yhdistetty nollahypoteesi

Testausvirheet

Odotusarvon testaaminen suurille datajoukoille

## Esim. Laadunvalvonta

Tukkukauppiaan väitteen mukaan sen toimittamista tomaateista enintään 5% on huonolaatuisia. Suuresta erästä poimittiin satunnaisesti 50 tomaattia ja niistä 4 todettiin huonolaatuisiksi.

$H_0$ : Huonojen osuus  $\theta \leq 0.05$

$H_1$ : Huonojen osuus  $\theta > 0.05$

Testisuure: Huonojen lkm:  $t(\vec{x}) = 4$

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

$$\mathbb{P}_\theta\left(T - \mathbb{E}_\theta(T) \geq t(\vec{x}) - \mathbb{E}_\theta(T)\right) = \mathbb{P}_\theta(T \geq t(\vec{x})) = \sum_{t=4}^{50} f_\theta(t).$$

Ongelma:  $t_n$  riippuu  $\theta$ :sta. Valitaan suurin  $t_n$  (miksi?).

$$p\text{-arvo} = \max_{\theta \leq 0.05} \mathbb{P}_\theta(T \geq t(\vec{x})) = \mathbb{P}_{0.05}(T \geq t(\vec{x})) = \sum_{t=4}^{50} f_{0.05}(t) \approx 24\%$$

# Tilastollisen merkitsevyydestin vaiheet

Joskus asetetaan myös vastahypoteesi  $H_1$ , jolloin tarkoitus on tutkia, puoltaako havaittu datajoukko enemmän  $H_0$ :n vai  $H_1$ :n hyväksymistä.

1. Asetetaan nollahypoteesi  $H_0$  ja mahdollisesti vastahypoteesi  $H_1$ .
2. Valitaan käytettävä testisuure ja lasketaan havaittua dataa  $\vec{x}$  vastaava testisuureen arvo  $t(\vec{x})$ .
3. Määritetään datalähteen tuottamia testisuureen arvoja mallintavan satunnaismuuttujan  $T$  jakauma nollahypoteesin vallitessa.
4. Määritetään mitkä testisuureen arvot tulkitaan havaittua testisuureen arvoa poikkeuksellisemmiksi (puoltavat  $H_1$ :n hyväksymistä  $H_0$ :n sijaan) ja lasketaan testin p-arvo.
5. Tehdään johtopäätökset:
  - Pieni p-arvo  $\implies H_0$  hylätään
  - Suuri p-arvo  $\implies H_0$  jää voimaan

# Testaaminen valitulla merkitsevyystasolla

Miten pieni p-arvo on riittävän pieni?

Tietyissä tilanteissa vaaditaan yksiselitteistä johtopäätöstä: testin pohjalta  $H_0$  joko hyväksytään tai hylätään. Tällaisen testin pohjaksi valitaan **merkitsevyystaso**  $\alpha \in (0, 1)$  ja johtopäätös muodostetaan seuraavasti:

- Jos p-arvo  $\geq \alpha$ , nollahypoteesi hyväksytään (jää voimaan),
- Jos p-arvo  $< \alpha$ , nollahypoteesi hylätään.

Usein on tapana valita merkitsevyystasoksi  $\alpha = 1\%$  tai  $\alpha = 5\%$

# Sisältö

Nollahypoteesi ja p-arvo

Yhdistetty nollahypoteesi

Testausvirheet

Odotusarvon testaaminen suurille datajoukoille

# Testausvirheet

Mikään ei takaa, että tehty johtopäätös olisi oikea.

	Johtopäätös	
Totuus	$H_0$ hyväksytään	$H_0$ hylätään
$H_0$ tosi	Oikea päätös	Hylkäysvirhe
$H_0$ epätosi	Hyväksymisvirhe	Oikea päätös

Testaajan johtopäätös on aina arvaus.

Hyvä arvaus on todennäköisesti oikein.

## Testausvirheiden todennäköisyydet

$p(\vec{x})$  = testisuureen p-arvo datajoukolle  $\vec{x}$

$\vec{X} = (X_1, \dots, X_n)$  mallintaa datalähteen tuottamia arvoja ennen niiden havaitsemista  $\implies p(\vec{X})$  on satunnaisluku

Hylkäysvirheen todennäköisyys on

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

Hyväksymisvirheen todennäköisyys on

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

$\alpha$	Hylkäysvirheen tn	Hyväksymisvirheen tn
Lähellä nollaa	Pieni	Suuri
Lähellä ykköstä	Suuri	Pieni

### Fakta

*Hylkäysvirheen  $tn \leq \alpha$ .*

# 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. Rikosoikeus

$H_0$ : Epäilty on syytön

$H_1$ : Epäilty on syyllinen

Havaittu data: Saatavilla oleva todistusaineisto

Hylkäysvirhe: Syytön tuomitaan

Hyväksymisvirhe: Syyllistä ei tuomita

## Anni Aktiivi

- Käyttää merkitsevyystasoa  
 $\alpha = 5\%$
- Langettaa useammin tuomioita
- Tuomitsee pitkällä tähtäyksellä  
 $\leq 5\%$  syyttömiä
- Jättää Villeä harvemmin  
syyllisiä tuomitsematta

## Ville Varovainen

- Käyttää merkitsevyystasoa  
 $\alpha = 1\%$
- Langettaa harvemmin tuomioita
- Tuomitsee pitkällä tähtäyksellä  
 $\leq 1\%$  syyttömiä
- Jättää Annia useammin syyllisiä  
tuomitsematta

## 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:  $t(\vec{x}) = \text{kruunien lkm}$

Testisuureen stokastinen malli:  $T = t(\vec{X})$

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

Havainnon  $\vec{x}$  p-arvo

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

Johtopäätös: Nollahypoteesi hylätään (5% merkitsevyystasolla).

Mitä osataan sanoa virhetodennäköisyyksistä?

## Esim. Kolikko — Hylkäysvirheen tn

Testin p-arvot testisuureen funktiona:

# kruunat	0	1	2	3	4	5	6	7	8	9	10
p-arvo (%)	0.2	2.1	10.9	34.4	75.4	100	75.4	34.4	10.9	2.1	0.2

5% merkitsevyytasolla testin hylkäysalue on  $\{0, 1, 9, 10\}$ .

Hylkäysvirheen tn on

$$\mathbb{P}(t(\vec{X}) \in \{0, 1, 9, 10\} \mid H_0) = \sum_{t=0}^1 f_{H_0}(t) + \sum_{t=9}^{10} f_{H_0}(t) \approx 2.1\%.$$

## Esim. Kolikko — Hyväksymisvirheen tn

Testin p-arvot testisuureen funktiona:

# kruunat	0	1	2	3	4	5	6	7	8	9	10
p-arvo (%)	0.2	2.1	10.9	34.4	75.4	100	75.4	34.4	10.9	2.1	0.2

5% merkitsevyytasolla testin hylkäysalue on  $\{0, 1, 9, 10\}$ .

Hyväksymisvirheen tn on

$$\mathbb{P}(t(\vec{X}) \in \{2, 3, \dots, 8\} | H_1) = ?$$

Ongelma: vastahypoteesi ( $H_1 : \theta \neq 0.5$ ) ei määrää  $\theta$ :n arvoa.

Ääritapaus  $\theta \approx 0.5$ , jolloin

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

## 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.

$$H_0: \text{Kruunan tn } \theta = 0.5,$$

$$H_1: \text{Kruunan tn } \theta = 0.9.$$

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

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

Hyväksymisvirheen tn:

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

# Sisältö

Nollahypoteesi ja p-arvo

Yhdistetty nollahypoteesi

Testausvirheet

Odotusarvon testaaminen suurille datajoukoille

## Odotusarvon testisuure suurille datajoukoille

Datalähde tuottaa toisistaan riippumattomia ja samoin jakautuneita satunnaislukuja  $X_1, X_2, \dots, X_n$  odotusarvona (tuntematon)  $\mu$ . Datalähteen väitetty odotusarvo on  $\mu_0$ .

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Jakauma tuntematon  $\implies$  mahdoton testata?

Approksimatiivinen testi mahdollinen, jos paljon dataa.

Testisuure:

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

### Fakta

*Suurille datajoukoille testisuureen stokastinen malli  $t(\vec{X})$  noudattaa likimain normitettua normaalijakaumaa, jolloin*

$$\text{p-arvo} \approx \mathbb{P}(|t(\vec{X})| \geq |t(\vec{x})| \mid H_0) \approx \mathbb{P}(|Z| \geq |t(\vec{x})|).$$

## Esim. Kahviautomaatti

Kahviautomaatin on tarkoitus laskea jokaiseen kuppiin keskimäärin 10.0 cl kahvia. Kahviautomaatin toimintaa testattiin valuttamalla automaatista 30 kupillista ja mittamalla kahvin määrät kupeissa.

Mittauksessa havaittiin arvot (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  $\mu_0 = 10.0$ .

Onko poikkeama **tilastollisesti merkitsevä**?



## Esim. Kahviautomaatti

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  $sd(\vec{x}) = 0.563$

$H_0: \mu = 10.0$

$H_1: \mu \neq 10.0$

Havaitun datajoukon testisuure:

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

Koska datajoukon koko  $n = 30$  on melko suuri, käytetään suuren datajoukon testiä:

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

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

Perjantaina kertaustuento. Kertaustuennolle saa ja kannattaa tuoda mukanaan kysymyksiä kurssilla käsitellyistä aiheista.