6 Tilastollinen testaus

Tilastolliset testit ovat frekventistisen päättelyn käytetyimpiä (ja huonoiten ymmärrettyjä ja sen takia eniten väärinkäytettyjä) tilastollisen päättelyn menetelmiä.
Niiden avulla pyritään ottamaan kantaa tilastollisia malleja koskeviin väitteisiin, kuten esim.

- Onko tutkittava lantti harhaton?
- Onko tietyllä käsittelyllä vaikutusta? Käsittely voisi olla esimerkiksi uusi hoitomuoto jollekin sairaudelle tai uusi lannoite tai uusi opetusmenetelmä.

Testaus sujuu käytännössä näin

- Lasketaan tilanteeseen sopivan testisuureen arvo.
- Verrataan laskettua arvoa siihen, minkälaisia arvoja hypoteesin mukaisesta populaatiosta satunnaisvaihtelu huomioon ottaen pitäisi tulla.
- Mikäli laskettu testisuureen arvo poikkeaa riittävän paljon hypoteesin mukaisista tyypillisistä arvoista, hypoteesi hylätään. Tällöin meistä ei enää ole uskottavaa, että näin suuri poikkeama aiheutuisi satunnaisvaihtelusta, vaan pidämme uskottavampana selityksenä sitä, että asetettu hypoteesi ei pidä paikkaansa.

6.1 Testauksen peruskäsitteitä

- Tarkastelemme testausta frekventistisessä tilastollisessa mallissa eli jakaumaperheessä

$$\{f(\mathbf{y};\theta),\ \theta \in \Theta\}.$$

- Koska havaintoja vastaavan satunnaisvektorin $\mathbf{Y}$ jakauma tunnetaan täysin, mikäli parametrinarvo θ tunnetaan, niin vektorin $\mathbf{Y}$ jakaumaa koskevat väittämät eli (tilastolliset) hypoteesit voidaan parametrisessa mallissa muotoilla seuraavasti:
  - Väitetään että parametri kuuluu johonkin tiettyyn parametrialueen osajoukkoon.

Yksinkertainen vai yhdistetty hypoteesi?

- Esim. lantinheittoa mallinnetetaan binomikokeena, jossa onnistumistodennäköisyys θ on tuntematon ja toistojen lukumäärä n on tunnettu. Väite "lantti on harhaton" vastaa hypoteesia
$$\theta = \frac{1}{2}, \quad \text{eli} \quad \theta \in \{\frac{1}{2}\}.$$
- Lantin harhattomuutta koskeva hypoteesi on yksinkertainen (engl. *simple*) eli täysin määrätty, sillä hypoteesia vastaava parametriavaruuden osajoukko sisältää vain yhden pisteen θ_0. Tällöin satunnaisvektorin $\mathbf{Y}$ jakaumalla on yptnf/ytf $f(\mathbf{y}; \theta_0)$, mikäli hypoteesi on tosi.
- Paljon tyypillisempää on, että hypoteesi on yhdistetty (engl. *composite*) eli osittain määrätty, mikä tarkoittaa sitä, että hypoteesia vastaava parametriavaruuden osajoukko koostuu useammasta kuin yhdestä pisteestä.

Tarkka hypoteesi saattaakin olla yhdistetty

- Jotkin hypoteesit saattavat ensinäkemältä vaikuttaa yksinkertaisilta, vaikka ne todellisuudessa ovat yhdistettyjä.
- Jos esimerkiksi normaalijakautuneessa populaatiossa $N(\mu, \sigma^2)$ molemmat parametrit ovat tuntemattomia, niin tällöin parametria μ koskeva tarkka hypoteesi

  $$H : \mu = 0$$

  on yhdistetty hypoteesi, sillä se vastaa parametriavaruuden osajoukkoa
  $$\{(\mu, \sigma^2) : \mu = 0 \text{ ja } \sigma^2 > 0\}.$$

Nollahypoteesi H_0

- Nollahypoteesi (engl. *null hypothesis*) H_0 on muotoa

$$H_0 : \theta \in \Theta_0, \tag{1}$$

  missä $\Theta_0 \subset \Theta$ on ei-tyhjä parametriavaruuden osajoukko. Testauksen tavoitteena on arvioida havaintojen $\mathbf{y}$ avulla nollahypoteesin paikkansapitävyyttä.
- Tavallisesti nollahypoteesi vastaa vakiintunutta teoriaa tai sitä pessimististä näkemystä, että käsittelyllä ei ole vaikutusta.
- Tutkija tahtoisi löytää todisteita nollahypoteesin hylkäämiseksi, mutta koska vakiintunutta teoriaa ei voida hylätä löyhin perustein, sitä vastaan pitää saada vakuuttavia todisteita, ennen kuin yhteisö suostuu hylkäämään nollahypoteesin.

Vastahypoteesi H_1

- Monesti nollahypoteesin H_0 lisäksi muotoillaan myös vaihtoehtoinen hypoteesi eli vastahypoteesi (engl. *alternative hypothesis*, myös *study hypothesis*), jota tyypillisesti merkitään symbolilla H_1 (tai H_A).
- Vastahypoteesin mukaan θ kuuluu parametriavaruuden osajoukkoon Θ_1, ts.

$$H_1 : \theta \in \Theta_1. \tag{2}$$

- Vähintäänkin oletetaan, että

$$\Theta_0 \cap \Theta_1 = \emptyset,$$

  ja usein (mutta ei aina) pätee $\Theta = \Theta_0 \cup \Theta_1$.
- Jos vastahypoteesi asetetaan, niin tämä tarkoittaa kannanottoa sen suhteen, mitä parametrin ajatellaan toteuttavan siinä tapauksessa, että nollahypoteesi osoittautuu epäilyksen alaiseksi.

Innocent until proven guilty

- Nollahypoteesia ja vastahypoteesia käsitellään testauksessa täysin epäsymmetrisellä tavalla. Tämän asian ymmärtäminen on edellytys sille, että osaamme tulkita testin lopputuloksen järkevästi.
- Tilanteen voi ajatella olevan analoginen oikeudenkäynnin kanssa, jossa syytettynä on nollahypoteesi H_0.
- H_0 on oikeudenkäynnissä syytön, ellei sitä osoiteta syylliseksi.

Testisuure

Testaus sujuu siten, että aineistosta lasketaan tunnusluku $t(\mathbf{y})$, jota kutsutaan testisuureeksi (engl. *test statistic*). Testisuure asettaa mahdolliset havainnot järjestykseen jollakin seuraavista tavoista sen asian suhteen, miten tyypillisinä tai outoina niitä pidämme, jos nollahypoteesi pitää paikkansa.

1. Pienet tunnusluvun arvot viittaavat siihen, että aineisto on sopusoinnussa H_0:n kanssa, ja suuret viittaavat ristiriitaan H_0:n kanssa.
2. Suuret tunnusluvun arvot viittavat sopusointuun H_0:n kanssa ja pienet arvot ristiriitaan sen kanssa.
3. Suuri poikkeama jostakin vertailuarvosta t_0 ylöspäin tai alaspäin viittaa ristiriitaan ja pieni poikkeama sopusointuun.

Vaatimus testisuureelle

- Edellisen lisäksi testisuureella $t(\mathbf{y})$ täytyy olla se ominaisuus, että hallitsemme vastaavan satunnaismuuttujan $t(\mathbf{Y})$ jakauman ainakin kaikilla nollahypoteesin mukaisilla parametrinarvoilla $\theta \in \Theta_0$.
- Usein testisuureena käytetään saranasuuretta, mikäli sellainen tunnetaan.
- Tällä kurssilla emme voi paneutua tämän syvällisemmin siihen, kuinka testisuure pitäisi valita.

Testin kulku

Havaitusta aineistosta lasketaan testisuureen arvo, ja sitten tarkistetaan kuuluuko se kriittiseen alueeseen (engl. *critical region*) eli hylkäysalueeseen (engl. *rejection region*) C. Testi antaa yhden kahdesta vaihtoehtoisesta päätöksestä, se joko hylkää nollahypoteesin tai sitten ei sen mukaan, kuuluko testisuureen arvo kriittiseen alueeseen vai ei.

- Jos $t(\mathbf{y}) \in C$, niin testi hylkää (engl. *reject*) nollahypoteesin, eli testi on merkitsevä (engl. *significant*).
- Jos $t(\mathbf{y}) \notin C$, niin testi hyväksyy (engl. *accept*) nollahypoteesin (mikä voidaan ilmaista myös sanomalla, että nollahypoteesi jää voimaan), eli testi ei ole merkitsevä (engl. *not significant*).

Huomautus terminologiasta

- Hylkääminen ja sen vastakohta, jota yksinkertaisuuden vuoksi tavallisimmin kutsutaan hyväksymiseksi, ovat testaukseen liittyviä teknisiä termejä. Se mitä käytännön johtopäätöksiä ja käytännön toimia testin lopputuloksen selvittyä tehdään, on eri asia kuin testin antama päätös.
- Varsinkin termi hyväksyä on harhaanjohtava.
- Mikäli H_0 hyväksytään, niin tutkija usein oikeasti edelleen epäilee nollahypoteesin paikkansapitävyyttä, mutta hän ei ole löytänyt aineistosta riittävän vakuttavaa todistetta sitä vastaan.

Kriittinen alue

- Kriittisen alueen muoto riippuu siitä, minkälaiset tunnusluvun arvot ovat nollahypoteesin kanssa yhteensopimattomia.
- Jos suuret tunnusluvun arvot ovat nollahypoteesin kannalta kriittisiä, niin kriittinen alue on muotoa

  $$C = (u, \infty)$$

  ts. testi hylkää nollahypoteesin, jos $t(\mathbf{y}) > u$.
- Tällöin kynnysarvoa u voidaan kutsua kriittiseksi arvoksi (engl. *critical value*).
- Tavallisesti kriittinen alue määräytyy testin merkitsevyystasosta, joka määritellään kohta.

Testin koko eli sen merkitsevyystaso

Määritelmä

Jos testin kriittinen alue on C, niin testin koko (engl. *size*) eli sen merkitsevyystaso (engl. *significance level*) on

$$\alpha = \sup_{\theta \in \Theta_0} P_\theta(t(\mathbf{Y}) \in C) = \sup_{\theta \in \Theta_0} P_\theta(H_0 \text{ hylätään}) \tag{3}$$

Käytämme termejä testin koko ja sen merkitsevyystaso synonyymeinä, mutta jotkut kirjoittajat tekevät näiden käsitteiden välille eron.

Huomautuksia testin koon määritelmästä

- Edellä sup eli *supremum* tarkoittaa pienintä ylärajaa; testin koko α on pienin yläraja hylkäystodennäköisyydelle $P_\theta(t(\mathbf{Y}) \in C)$, kun satunnaisvektorilla $\mathbf{Y}$ on nollahypoteesin mukainen jakauma.
- Usein $P_\theta(t(\mathbf{Y}) \in C)$ pysyy vakiona joukossa Θ_0 — näin käy automaattisesti, jos H_0 on yksinkertainen tai jos testisuure on saranasuure. Tällöin supremumin otosta ei tarvitse huolehtia, vaan

$$\alpha = P_\theta(H_0 \text{ hylätään}), \qquad \text{millä tahansa } \theta \in \Theta_0. \tag{4}$$

Konventionaaliset merkitsevyystasot

- Tyypillisesti testin merkitsevyystaso $0 < \alpha < 1$ asetetaan, ja sitten tämän informaation perusteella määritetään kriittinen alue C siten, että vaatimus (3) toteutuu.
- Ennen vanhaan ei ollut käytössä tilastollisia ohjelmia, ja merkitsevyystasolle α kiinnitettiin tavallisesti jokin seuraavista konventionaalisista arvoista

  $$0.05, \quad 0.01, \quad \text{tai} \quad 0.001$$

  sen takia, että näitä arvoja vastaavat kriittiset arvot löytyivät tilastollisista taulukoista.
- Nämä konventionaaliset tasot ovat mielivaltaisia, ja ne on valittu sillä perusteella, että vastaavat murtoluvut (yksi kahdestakymmenestä, yksi sadasta, yksi tuhannesta) ovat pyöreitä.

I ja II lajin virheet

- Jos H_0 pitää paikkansa, mutta testi hylkää sen, tällöin tapahtuu hylkäämissvirhe eli I lajin virhe (engl. *type I error*).
- Jos H_1 pitää paikkansa, mutta testi hyväksyy H_0:n, tapahtuu hyväksymisvirhe eli II lajin virhe (engl. *type II error*).

Todellisuus	Päätös	
	H_0 hyväksytään	H_0 hylätään
H_0 tosi	oikea päätös	hylkäämisvirhe I lajin virhe
H_1 tosi	hyväksymisvirhe II lajin virhe	oikea päätös

Huomautuksia

- Nollahypoteesia ja vastahypoteesia kohdellaan epäsymmetrisellä tavalla.
- Merkitsevyystaso α on yläraja hylkäämisvirheen todennäköisyydelle. *Jos nollahypoteesi pitää paikkaansa*, niin testisuure saa hylkäämiseen johtavia arvoja niin harvoin, että hylkäämistodennäköisyys on enintään α.
- Tähän asti emme ole miettineet sitä, mitä testissä tapahtuu jos H_1 on tosi.
- Perinteinen tapa raportoida testin tulos on ollut kiinnittää testin koko α sekä kertoa testin päätös, eli hylkäsikö vai hyväksyikö testi nollahypoteesin.
- Nykyään ei välttämättä toimita näin yksioikoisesti.

6.2 Normaalijakautuneen populaation odotusarvon testaus, kun varianssi on tunnettu

- Tarkastelemme satunnaisotosta $Y_1, \ldots, Y_n$ normaalijakaumasta $N(\mu, \sigma^2)$. Ts. satunnaismuuttujat Y_i ovat riippumattomia, ja niillä on kaikilla normaalijakauma $N(\mu, \sigma^2)$.
- Tarkastelemme odotusarvon testausta, kun varianssi on tunnettu.
- Tällä tilanteella ei ole suurta käytännön arvoa.
- Tilannetta käsitellään sen vuoksi, että teoria on tässä tapauksessa helpointa ymmärtää.

Yksisuuntainen testi

- Jos populaation varianssi σ^2 on tunnettu luku, niin

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

  on saranasuure.
- Tarkastelemme nollahypoteesia

$$H_0 : \mu = \mu_0,$$

  jossa μ_0 on tunnettu luku (esim. $\mu_0 = 0$). Tämä on yksinkertainen hypoteesi.
- Otamme vastahypoteesiksi yksisuuntaisen hypoteesin

$$H_1 : \mu > \mu_0,$$

  joka on yhdistetty hypoteesi.

Testisuure z-testissä

- Käytämme testisuureena tunnuslukua
$$z = t(\mathbf{y}) = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}. \tag{5}$$
- Huomaa, että testisuure on tunnusluku (toisin kuin sitä vastaava saranasuure), sillä testisuureessa tuntemattoman parametrin μ tilalla on tunnettu arvo μ_0.
- Testisuuretta vastaavalla satunnaismuuttujalla
$$t(\mathbf{Y}) = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}}$$
on standardinormaalijakauma $N(0, 1)$, kun nollahypoteesi pitää paikkansa.
- Suuret testisuureen arvot ovat nollahypoteesin kannalta kriittisiä, sillä $\bar{Y}$ estimoi populaatioparametria μ, ja testisuure on kasvava funktio tästä estimaattorista.

Tason α testi saadaan aikaan käyttämällä kriittistä arvoa z_α, sillä

$$P_{\mu_0}(t(\mathbf{Y}) > z_\alpha) = P(Z > z_\alpha) = \alpha,$$

jossa $Z \sim N(0,1)$. Tästä nähdään, että luottamustason α testi hypoteesiparille

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu > \mu_0,$$

tekee päätöksen seuraavasti. Ensin lasketaan testisuureen arvo

$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$$

ja sitten testi toimii seuraavasti

$$\begin{cases} \text{jos } z > z_\alpha, & H_0 \text{ hylätään} \\ \text{jos } z \leq z_\alpha & H_0 \text{ hyväksytään.} \end{cases}$$

Yksisuuntainen z-testi

Kun testataan hypoteesia

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu > \mu_0$$

testi hylkää nollahypoteesin silloin (ja vain silloin), kun

$$z > z_\alpha. \tag{6}$$

Tämä on ns. yksisuuntainen eli yksitahoinen z-testi (engl. *one-sided* tai *one-tailed z-test*).

Onko yksisuuntainen testi järkevä?

- Äskeisellä tavalla

  $$H_0 : \mu = \mu_0, \qquad H_1 : \mu > \mu_0$$

  muotoiltuna yksisuuntainen testi on omituinen, sillä

  $$\Theta_0 \cup \Theta_1 = \{\mu_0\} \cup (\mu_0, \infty) = [\mu_0, \infty) \neq \Theta = \mathbb{R},$$

  vaan parametriavaruudesta jätetään kokonaan huomioimatta ne μ, joille $\mu < \mu_0$.
- On vaikea sanoa esim., tehdäänkö virhe vai toimitaanko oikein, jos todellisuudessa $\mu < \mu_0$, mutta nollahypoteesi hylätään.

Järkevämpi muotoilu yksisuuntaisen testin nollahypoteesille

- Näemme myöhemmin, että sama yksisuuntainen testi on koon α testi myös hypoteesiparille

$$H_0 : \mu \leq \mu_0, \qquad H_1 : \mu > \mu_0.$$

- Tälle hypoteesiparille

$$\Theta_0 \cup \Theta_1 = (-\infty, \mu_0] \cup (\mu_0, \infty) = \mathbb{R} = \Theta.$$

- Voidaan ajatella, että yksisuuntaisella z-testillä selvitetään aina yhdistetyn nollahypoteesin $\mu \leq \mu_0$ paikkansapitävyyttä. Testin kriittinen alue on paljon helpompi johtaa, jos nollahypoteesina käytetään yksinkertaista hypoteesia $\mu = \mu_0$. Tämä lienee se ainoa syy, miksi tätä tarkkaa nollahypoteesin muotoilua lainkaan käytetään yksisuuntaiselle z-testille.

Se toinen yksisuuntainen z-testi

Vastaavilla laskuilla nähdään, että sekä hypoteesiparille

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu < \mu_0$$

että hypoteesiparille

$$H_0 : \mu \geq \mu_0, \qquad H_1 : \mu < \mu_0$$

luottamustason α testi tekee päätöksen seuraavasti. Ensin lasketaan testisuureen arvo

$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}.$$

Tällä kertaa pienet arvot (ts. itseisarvoltaan suureet negatiiviset arvot) ovat nollahypoteesille kriittisiä. Testi hylkää nollahypoteesin silloin (ja vain silloin), kun

$$z < -z_\alpha \tag{7}$$

Kaksisuuntainen testi

- Nollahypoteesi ja vastahypoteesi ovat

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0.$$

- Testisuure on edelleen

$$z = t(\mathbf{y}) = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}},$$

  ja sitä vastaavalla satunnaismuuttujalla $t(\mathbf{Y})$ on $N(0,1)$-jakauma, kun H_0 pitää paikkansa.
- Nyt sekä suuret että pienet testisuureen arvot ovat nollahypoteesin kannalta kriittisiä.

Kaksisuuntainen z-testi

Kaksisuuntainen (engl. *two-sided*, *two-tailed*) z-testi hypoteeseille

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0.$$

luottamustasolla α hylkää nollahypoteesin (täsmälleen) silloin, kun

$$|z| > z_{\alpha/2}. \tag{8}$$

Tämä perustuu siihen, että

$$P(|Z| > z_{\alpha/2}) = \alpha,$$

kun $Z \sim N(0, 1)$.

Fiktiivinen numeerinen esimerkki

- Planeetalla Z seurataan tiiviisti JTP-kurssin aineistoja, koska paikalliset tutkijat ovat huomanneet kosmisen yhteyden planeetan Z ilmaston tilan ja JTP-kurssin simuloitujen aineistojen parametrien, erityisesti kuvan 4.3 aineiston parametrien välillä.
- Valitettavasti planeetalla Z ei osata suomea, vaan koko väestö puhuu englantia (hassusti murtaen). Tämän takia kukaan tutkija ei ole saanut selville, että oikeasti $\mu = 0.2012$. Sen sijaan ollaan saatu selville, että kyseessä on satunnaisotos normaalijakaumasta $N(\mu, \sigma^2)$, jossa $\sigma^2 = 1$.
- Planeetalla Z suositaan z-testejä kiinteällä merkitsevyystasolla 0.05.

$H_0 : \mu \leq 0$

- Pessimistit ovat sitä mieltä, että $\mu \leq 0$, mikä tarkoittaa planeetan Z pikaista tuhoa.
- Teemme z-testin, jossa hypoteesit ovat

$$H_0 : \mu \leq 0, \qquad H_1 : \mu > 0.$$

- Aineistossa

$$\bar{y} = 0.726, \quad n = 10,$$

  Aineistosta laskettu z-arvo on

$$z = \frac{\bar{y} - 0}{\sigma/\sqrt{n}} = 2.296$$

- Koska $z > z_{0.05} = 1.645$, *nollahypoteesi* $\mu \leq 0$ *hylätään* merkitsevyystasolla 5 %.
- Planeetan sanomalehdet kirjoittavat etusivuillaan, että tutkijat ovat todistaneet, että maailmanloppua ei tule.

$H_0 : \mu = \frac{1}{2}$

- Suurin osa tutkijoista ei kuitenkaan hyväksy hypoteesin $\mu \leq 0$ tieteellistä relevanssia, vaan uskoo teoriaan, jonka mukaan $\mu = \frac{1}{2}$, joka tarkoittaa sitä että planeetan ilmasto säilyy ikuisesti yhtä suotuisana kuin nykyään.
- Teemme z-testin, jossa hypoteesit ovat

$$H_0 : \mu = \frac{1}{2}, \qquad H_1 : \mu \neq \frac{1}{2}.$$

- Aineistosta laskettu z-arvo on

$$z = \frac{\bar{y} - \frac{1}{2}}{\sigma/\sqrt{n}} = 0.715,$$

  Koska $|z| \leq z_{\alpha/2} = 1.960$, niin *nollahypoteesi* $\mu = \frac{1}{2}$ *hyväksytään* merkitsevyystasolla 5 %.
- Planeetan sanomalehdet kirjoittavat etusivullaan, että tutkijat ovat todistaneet, että ilmasto säilyy ikuisesti suotuisana.

Pohdintaa

- Mikä tulosten uutisoinnissa oli vikana?
- Nollahypoteesin hylkääminen tarkoittaa sitä, että ollaan löydetty todisteita sitä vastaan.
- Nollahypoteesin hyväksyminen ei tarkoita sitä, että oltaisiin löydetty todisteita nollahypoteesin puolesta. Se tarkoittaa sitä, että ei olla löydetty painavia todisteita nollahypoteesia vastaan.
- Mitä varten tässä esimerkissä hölmö ja epätosi nollahypoteesi $H_0 : \mu = \frac{1}{2}$ hyväksyttiin?
- Tähän asiaan saamme lisävalaistusta sen jälkeen, kun olemme nähneet, kuinka z-testin voima saadaan laskettua.

6.3 Testin voima

Määritelmä (Testin voima)

Jos C on tarkasteltavan testin kriittinen alue, niin parametriavaruudella määriteltyä funktio

$$\pi(\theta) = P_\theta(t(\mathbf{Y}) \in C) = P_\theta(H_0 \text{ hylätään}) \tag{9}$$

on nimeltään testin voima (engl. *power*) tai sen voimafunktio (engl. *power function*).

Voimafunktio on testin hylkäystodennäköisyys parametrin funktiona — se mittaa hylkäysvoimaa.

Huomioita testin voimasta

- Testin voiman voi ilmaista monimutkaisemmalla tavalla ensimmäisen ja toisen lajin virheiden todennäköisyyden avulla:

$$\pi(\theta) = \begin{cases} P_\theta(\text{I lajin virhe}), & \text{kun } \theta \in \Theta_0, \\ 1 - P_\theta(\text{II lajin virhe}), & \text{kun } \theta \in \Theta_1. \end{cases}$$

- Jos testin koko on α, niin testin koon määritelmän (3) ja sen voiman määritelmän (9) mukaan

$$\alpha = \sup_{\theta \in \Theta_0} \pi(\theta).$$

- Merkitsevyystaso asettaa ylärajan testin voimalle nollahypoteesin mukaisilla parametrinarvoilla. Tahtoisimme, että testin voima olisi mahdollisimman suuri vaihtoehtohypoteesin mukaisilla parametrinarvoilla.

Testin voiman käyttö tutkimuksen suunnitteluvaiheessa

- Testin voimaa on syytä tarkastella tutkimuksen suunnitteluvaiheessa.
- Yleisesti ottaen, mitä suurempi on otoskoko, sitä suurempi on testin voima (vaihtoehtohypoteesin mukaisilla parametrinarvoilla).
- Tavallisesti tutkijalla on käsitys siitä, miten suuret poikkeamat nollahypoteesin mukaisista parametrinarvoista ovat käytännössä merkittäviä.
- Otoskoko voidaan yrittää valita siten, että saavutetaan vähintään jokin annettu voima (esim. vähintään 80 %) aina, kun poikkeama nollahypoteesista on käytännön kannalta merkittävä.

6.4 Testin *p*-arvo

- Vanhanaikainen tapa raportoida testin tulos on kertoa testin koko α sekä kertoa testin päätös, eli hylkäsikö vai hyväksyikö testi nollahypoteesin.
- Nykyään on tapana kertoa tämän lisäksi (tai sijasta) testin *p*-arvo (engl. *p-value*) eli havaittu merkitsevyystaso (engl. *observed significance level*).
- Testin *p*-arvo mittaa tietyllä tavalla nollahypoteesin ja aineiston yhteensopivuutta.

p-arvon määrittely

- Testin p-arvon määrittely on hieman erilainen sen mukaan, mitkä testisuureen arvot ovat nollahypoteesille kriittisiä.
- Mikäli testisuureen $t(\mathbf{y})$ suuret arvot ovat nollahypoteesille kriittisiä, niin p-arvo määritellään kaavalla

$$p = p(\mathbf{y}) = \sup_{\theta \in \Theta_0} P_\theta[t(\mathbf{Y}) \geq t(\mathbf{y})] \tag{10}$$

Tässä $t(\mathbf{y})$ on havaitusta aineistosta $\mathbf{y}$ laskettu testisuureen arvo, ja $t(\mathbf{Y})$ on satunnaisvektorista $\mathbf{Y}$ laskettu testisuureen arvo, kun sillä on jakaumana mallin mukainen yptnf/ytf $f(\mathbf{y}; \theta)$.

p-arvon sanallinen määritelmä

Useissa tilanteissa p-arvon määritelmässä esiintyvä todennäköisyys ei riipu siitä, mitä nollahypoteesin mukaista parametrinarvoa $\theta \in \Theta_0$ tarkastellaan, jolloin p-arvo voidaan määritellä sanallisesti seuraavasti.

p-arvo on se todennäköisyys, jolla nollahypoteesin mukaisesta populaatiosta saadaan testisuureen arvo, joka on vähintään yhtä kummallinen kuin aineistosta laskettu testisuureen arvo.

Kummallisuutta mitataan testisuureen arvolla: kummallisempia ovat ne arvot jotka poikkeavat vielä enemmän nollahypoteesille kriittiseen suuntaan kuin havaittu arvo. Usein kummallisuuden sijasta sanotaan "vähintään yhtä äärevä" (engl. *at least as extreme as*).

Pieni p-arvo viittaa ristiriitaan H_0:n kanssa

- Jos p-arvo on pieni (esim. 1 %), niin tällöin hyvin pienellä todennäköisyydellä nollahypoteesin mukaisesta populaatiosta saadaan vähintään yhtä kummallisia havaintoja kuin mitä todellisuudessa saatiin.
- Jos taas p-arvo on suuri (esim. 20 %), niin tällöin kohtuullisen usein nollahypoteesin mukaisesta populaatiosta saadaan havaintoja, jotka ovat vähintään yhtä kummallisia kuin mitä todellisuudessa havaittiin.

Pieni p-arvo viittaa ristiriitaan aineiston ja nollahypoteesin välillä.

Tarkennus

Jos p-arvon määritelmässä (10) oleva todennäköisyys kuitenkin riippuu parametrinarvosta $\theta \in \Theta_0$, niin oikea sanallinen määritelmä on

- p-arvo on pienin yläraja sille todennäköisyydelle, jolla nollahypoteesin mukaisesta populaatiosta saadaan testisuureen arvo, joka on vähintään yhtä kummallinen kuin aineistosta laskettu testisuureen arvo.

Testin päätöksen lukeminen p-arvosta

- Testi saadaan suoritettua myös laskemalla aineistosta p-arvo, ja toimimalla seuraavasti:
- Testi hylkää H_0:n, jos $p < \alpha$. Muussa tapauksessa H_0 jää voimaan.
- Osoitamme esimerkeissä, että näin menetellen testin kooksi tulee α.
- Nykyään on tapana ilmoittaa testin p-arvo (parilla desimaalilla), ja lisäksi varmuuden vuoksi kommentoida (asiantuntematonta lukijaa ajatellen), tulisko nollahypoteesi hylättyä vai hyväksyttyä konventionaalisilla merkitsevyystasoilla. Tämä on paljon informatiivisempaa kuin pelkästään kertoa testin päätös jollakin kiinteällä merkitsevyystasolla.

6.5 z-testin p-arvo ja voima

Laskemme seuraavaksi jaksossa 6.2 käsitellyn z-testin p-arvon ja voimafunktion sekä yksi- että kaksisuuntaisessa tapauksessa.

Yksisuuntainen z-testi

- Hypoteesit

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu > \mu_0.$$

- Testisuure lasketaan kaavalla

$$z = t(\mathbf{y}) = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}.$$

- Suuret testisuureen arvot ovat H_0:lle kriittisiä.
- Vertailujakauma: $t(\mathbf{Y}) \sim N(0, 1)$, kun nollahypoteesi pitää paikkansa.

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$: p-arvo

- Testisuure on $z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$ ja hylkäysalue on $z > z_\alpha$.
- p-arvo on
$$p = P_{\mu_0}(Z \geq z) = 1 - \Phi(z),$$
jossa Φ on $N(0,1)$-jakauman kertymäfunktio, ja $Z \sim N(0,1)$.
- Tämä testi hylkää täsmälleen silloin, kun
$$\begin{aligned} & z > z_\alpha \\ \Leftrightarrow \quad & \Phi(z) > 1 - \alpha \\ \Leftrightarrow \quad & p = 1 - \Phi(z) < \alpha. \end{aligned}$$
- Olemme tarkistaneet tälle testille, että testin päätös voidaan lukea sen p-arvosta.

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu > \mu_0$: voimafunktio

- Voimafunktio (eli hylkäystodennäköisyys parametrin funktiona) on

$$\pi(\mu) = P_\mu[t(\mathbf{Y}) > z_\alpha] = P_\mu\left[\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha\right]$$

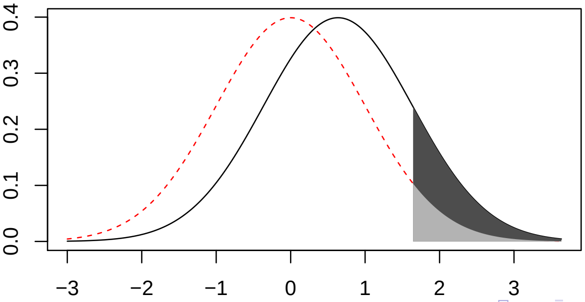
- Tässä

$$\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \sim N\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}}, 1\right),$$

kun todellinen parametrinarvo on μ.

Havainnollistus yksisuuntaisen z-testin voimasta

- Tarkastellaan hypoteesin $H_0 : \mu = 0$ voimaa, kun todellisuudessa $\mu = 0.2012$. Normaalijakauman varianssi $\sigma^2 = 1$ on oletettu tunnetuksi.
- Normaalijakauman $N(0, 1)$ häntäalueen pinta-ala on $\alpha = 0.05$, ja testin voima on normaalijakauman $N((\mu - 0)/(\sigma/\sqrt{n}), 1)$ kuvaan merkityn häntäalueen pinta-ala.



Voiman laskeminen

Voima (eli hylkäystodennäköisyys parametrin funktiona) on

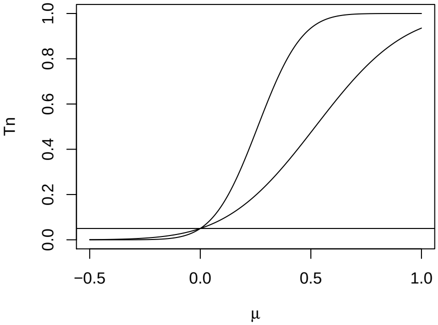
$$
\begin{aligned}
\pi(\mu) &= P_\mu[t(\mathbf{Y}) > z_\alpha] = P_\mu\left[\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha\right] \\
&= P_\mu\left[\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} > z_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right] \\
&= P\left[Z > z_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right] \\
&= 1 - \Phi\left(z_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}} - z_\alpha\right).
\end{aligned}
$$

Viimeinen yhtäsuuruus perustuu $N(0,1)$-jakauman symmetrisyyteen, minkä takia sen kertymäfunktio toteuttaa identiteetin

$$1 - \Phi(a) = \Phi(-a), \qquad \text{kaikilla } a.$$

Kuva yksisuuntaisen z-testin voimafunktiosta

Yksisuuntaisen z-testin voimafunktio, kun $\alpha = 0.05$, $\mu_0 = 0$, $\sigma = 1$, ja otoskoko on $n = 10$ tai $n = 40$. Suuremmalla otoskoolla saavutetaan suurempi voima vastahypoteesin $\mu > \mu_0$ mukaisilla parametrinarvoilla. Testin koko on osoitettu vaakaviivalla.



Kaksi muotoilua yksisuuntaiselle testille

- Äsken johdetusta voimafunktion kaavasta (sekä kuvasta) nähdään, että se on aidosti kasvava funktio, minkä takia

$$\pi(\mu) \leq \pi(\mu_0) = \alpha, \quad \text{kaikilla } \mu \leq \mu_0.$$

- Johtopäätös: yksisuuntainen z-testi hypoteesiparille

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu > \mu_0$$

on tason α testi myös hypoteesiparille

$$H_0 : \mu \leq \mu_0, \qquad H_1 : \mu > \mu_0.$$

- Jälkimmäiselle hypoteesiparille

$$\Theta_0 \cup \Theta_1 = (-\infty, \mu_0] \cup (\mu_0, \infty) = \mathbb{R} = \Theta.$$

z-testi $H_0 : \mu = \mu_0$ vs. $H_1 : \mu < \mu_0$

Toisen mahdollisen yksisuuntaisen testin

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu < \mu_0$$

p-arvo on

$$p = P_{\mu_0}(Z \leq z) = \Phi(z),$$

jossa Φ on $N(0, 1)$-jakauman kertymäfunktio, $Z \sim N(0, 1)$, ja z on aineistosta laskettu testisuureen arvo, eli

$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}},$$

jonka pienet arvot ovat nollahypoteesille kriittisiä.

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu < \mu_0$: voimafunktio

- Tämän testin voimafunktio on

$$\begin{aligned} \pi(\mu) &= P_\mu[t(\mathbf{Y}) \le -z_\alpha] \\ &= P\left[Z \le -z_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right] \\ &= \Phi\left(-z_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) \end{aligned}$$

- Nyt voimafunktio on aidosti vähenevä, joten

$$\pi(\mu) \le \pi(\mu_0) = \alpha, \quad \text{kaikilla } \mu \ge \mu_0.$$

- Voimafunktion kuvaaja on peilikuva ensimmäiseksi käsitellyn yksisuuntaisen z-testin voimafunktion kuvaajasta. Tämä yksisuuntainen z-testi on tason α testi myös hypoteesiparille

$$H_0 : \mu \ge \mu_0, \qquad H_1 : \mu < \mu_0$$

Kaksisuuntainen z-testi

Kaksisuuntaisen testin

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

testaamisessa testisuureen

$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}},$$

sekä suuret että pienet (negatiiviset) arvot ovat nollahypoteesille kriittisiä.

Kaksisuuntainen z-testi: p-arvo

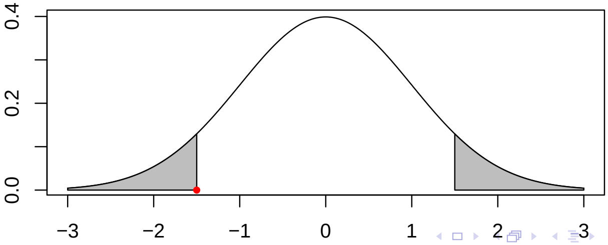
P-arvo määritellään summaamalla häntätodennäköisyys molemmilta viitejakauman hänniltä, ts. kummallisia tai vielä kummallisempi arvoja ovat ne, joille

$$|Z| \geq |z|,$$

Tällä perusteella

$$p = P[|Z| \geq |z|] = \Phi(-|z|) + 1 - \Phi(|z|) = 2(1 - \Phi(|z|))$$

Kuvan tilanteessa havainnoista on laskettu arvo $z = -1.5$.



Kaksisuuntaisen testin päätöksen lukeminen p-arvosta

- Kaksisuuntainen z-testi hylkää täsmälleen silloin, kun

$$
\begin{aligned}
& |z| > z_{\alpha/2} \\
\Leftrightarrow \quad & \Phi(|z|) > 1 - \frac{\alpha}{2} \\
\Leftrightarrow \quad & p = 2(1 - \Phi(|z|)) < \alpha.
\end{aligned}
$$

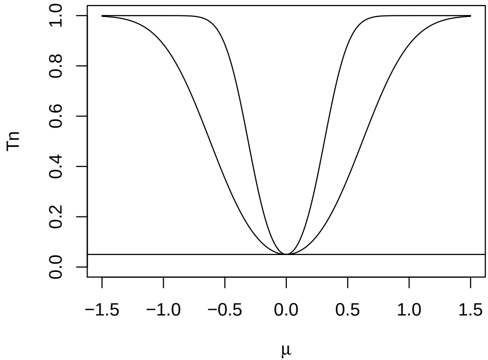
- Nyt olemme tarkistaneet myös kaksisuuntaiselle z-testille, että testin päätös voidaan lukea sen p-arvosta.

Lasketaan kaksisuuntaisen testin voimafunktio $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

$$
\begin{aligned}
\pi(\mu) &= P_\mu[|t(\mathbf{Y})| \geq z_{\alpha/2}] \\
&= P_\mu\left[\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \geq z_{\alpha/2}\right] + P_\mu\left[\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \leq -z_{\alpha/2}\right] \\
&= P\left[Z \geq z_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right] + P\left[Z \leq -z_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right] \\
&= 1 - \Phi\left(z_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right).
\end{aligned}
$$

Kuva kaksisuuntaisen z-testin voimafunktiosta

Kaksisuuntaisen z-testin voimafunktio, kun $\alpha = 0.05$, $\mu_0 = 0$, $\sigma = 1$, ja otoskoko on $n = 10$ tai $n = 40$. Suuremmalla otoskoolla saavutetaan suurempi voima. Testin koko on osoitettu vaakaviivalla.



p-arvo ja voima planeetan Z esimerkissä

- Aineiston yhteenveto oli

$$\bar{y} = 0.726, \quad n = 10, \quad \sigma^2 = 1.$$

- Aluksi testattiin yksisuuntaista hypoteesia

$$H_0 : \mu \leq 0, \qquad H_1 : \mu > 0,$$

  joka hylättiin merkitsevyystasolla $\alpha = 0.05$.

- Tämän testin p-arvo ja testin voima todelliselle parametrinarvolle $\mu = 0.2012$ saadaan laskettua R:llä seuraavasti. (Perus-R:stä ei löydy funktiota z-testiä varten). Tässä funktio `pnorm` laskee normaalijakauman kertymäfunktion arvon.

R-koodia: *p*-arvo ja voima yksisuuntaiselle *z*-testille

```
mean.y <- 0.726
n <- 10
sigma <- 1
mu0 <- 0
alpha <- 0.05
mu.true <- 0.2012
sem <- sigma/sqrt(n)
z <- (mean.y - mu0)/sem
p <- 1 - pnorm(z)
zcrit <- qnorm(lower = FALSE, alpha)
pwr <- pnorm((mu.true - mu0)/sem - zcrit)
c(z, zcrit, p, pwr)

## [1] 2.29581 1.64485 0.01084 0.15658
```

Tulosten käsittelyä

- Testin p-arvo $p = 0.011$. Tämä on yläraja sille todennäköisyydelle, että nollahypoteesin mukaisesta populaatiosta saadaan z-tunnusluvun arvo, joka on suurempi tai yhtä suuri kuin aineistosta laskettu z-tunnusluvun arvo.
- Testin voima todelliselle parametrinarvolle on vain 0.16. Ennen aineiston simulointia todennäköisyys, että testi tulee hylkäämään nollahypoteesin $\mu \leq 0$ oli (ainoastaan) 0.16.
- Olimme aika onnekkaita, koska pystyimme (todellisen tilanteen mukaisesti) hylkäämään tämän epätoden nollahypoteesin.

Kaksisuuntainen testi

Toisen tutkijayhteisön mielipiteen mukaisesti tehtiin seuraavaksi testi

$$H_0 : \mu = \frac{1}{2}, \qquad H_1 : \mu \neq \frac{1}{2},$$

joka hyväksyttiin merkitsevyystasolla $\alpha = 0.05$. Tässä tapauksessa p-arvo ja testin voima voidaan laskea seuraavasti.

```
mu0 <- 0.5
z <- (mean.y - mu0)/sem
zcrit <- qnorm(alpha/2, lower = FALSE)
p <- 2 * (1 - pnorm(abs(z)))
pwr <- 1 - pnorm(zcrit - (mu.true - mu0)/sem) + pnorm(-zcrit -
    (mu.true - mu0)/sem)
c(z, zcrit, p, pwr)

## [1] 0.7147 1.9600 0.4748 0.1569
```

Tulosten käsittelyä

- Yksisuuntaisen testin p-arvo on $p = 0.47$, joka on se todennäköisyys, että nollahypoteesin mukaisesta populaatiosta saadaan z-tunnusluvun arvo, joka on itseisarvoltaan vähintään yhtä suuri kuin aineistosta laskettu z-tunnusluvun itseisarvo.
- Testin voima todelliselle parametrinarvolle on (taas) noin 0.16. Ennen aineiston simulointia todennäköisyys, että testi tulee hylkäämään nollahypoteesin $\mu = \frac{1}{2}$ oli 0.16.
- Tässä tapauksessa testin voima on niin pieni, että hyväksymispäätöstä ei pitäisi tulkita todisteena H_0:n puolesta, mikäli $\mu = 0.2012$ on käytännön eli planeetan Z ilmaston kannalta merkittävästi erilainen kuin arvo $\mu_0 = \frac{1}{2}$.

6.6 Testien ja luottamusvälien duaalisuus

- Tarkastelemme kaksisuuntaisen z-testin

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

hyväksymisaluetta, eli niitä μ_0 joita tason α testi (8) ei hylkää.

- Testi ei hylkää, mikäli $|z| \leq z_{\alpha/2}$ eli mikäli

$$\begin{aligned} & -z_{\alpha/2} \leq \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \\ \Leftrightarrow \quad & \bar{y} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \mu_0 \leq \bar{y} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \end{aligned}$$

- Tämä on z-luottamusväli luottamustasolla $1 - \alpha$.
- Kaksisuuntainen z-luottamusväli saadaan kääntämällä kaksisuuntaisen z-testin hyväksymisalue: ts. ratkaisemalla kaikkien muotoa $H_0 : \mu = \mu_0$ olevien kaksisuuntaisten testien hyväksymisalueet.

z-testeille duaaliset luottamusvälit

Voimme samaan tapaan kääntää myös yksisuuntaisten z-testien hyväksymisalueet, ja näin menetellen saadaan seuraavat tapaukset. Taulukossa $z = (\bar{y} - \mu_0)/(\sigma/\sqrt{n})$, σ^2 on tunnettu luku, ja z_u on $N(0,1)$-jakauman u-yläkvantiili.

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

Uusi tulkinta testin p-arvolle

- Mikäli $p > \alpha$, niin tällöin μ_0 kuuluu testiä vastaavaan duaalisen luottamustason $1 - \alpha$ luottamusväliin.
- Jos taas $p < \alpha$, niin μ_0 ei kuulu ko. luottamusväliin.

Huomioita testeistä ja luottamusväleistä

- Merkitsevyystason α testit ja luottamustason $1-\alpha$ luottamusvälit yrittävät antaa vastauksen samantapaiseen kysymykseen, mutta erilaisista näkökulmista.
- Testissä kiinnitetään yksi parametrinarvo, ja tutkitaan ovatko havainnot sopusoinnussa tämän parametrinarvon kanssa.
- Luottamusväli yrittää kertoa suoraan, mitkä parametrinarvot ovat sopusoinnussa havaintojen kanssa.
- Tähän testien ja luottamusvälien vastaavuuteen voidaan viitata sanomalla, että ne ovat duaalisia käsitteitä.

Yksisuuntainen z-testi planeetalla Z

- Planeetan Z esimerkissä yksisuuntaista hypoteesia

$$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0$$

  vastaa parametrin μ yksisuuntainen 95 %:n luottamusväli $[0.20, \infty)$, joka ei sisällä esimerkissä kiinnostavaa arvoa 0, joten nollahypoteesi $H_0 : \mu \leq 0$ hylätään merkitsevyystasolla 0.05.
- Testin p-arvo on $p = 0.011 < 0.05$, joten myös tästä nähdään, että $\mu_0 = 0$ ei kuuluu ko. 95 %:n yksisuuntaiseen luottamusväliin ilman että luottamusväliä edes tarvitsee laskea.

Kaksisuuntainen z-testi planeetalla Z

- Kaksisuuntaista hypoteesia

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

  vastaa parametrin μ kaksisuuntainen 95 %:n luottamusväli $[0.10, 1.35]$, joka sisältää esimerkissä kiinnostavan arvon $\frac{1}{2}$, joten nollahypoteesi $\mu = \frac{1}{2}$ hyväksytään merkitsevyystasolla 0.05.
- Testin p-arvo on $p = 0.47$, mistä myöskin nähdään, että $\mu_0 = \frac{1}{2}$ kuuluu kaksisuuntaiseen 95 %:n luottamusväliin ilman, että luottamusväliä tarvitsee edes laskea.

Pohdintaa: testi vai luottamusväli

- Testaaminen johtaa käyttäjän helposti mustavalkoiseen ajatteluun: nollahypoteesi joko hyväksytään tai hylätään. Huomio kiinnittyy pois siitä, kuinka epävarmaa aineiston antama informaatio parametrista on.
- Luottamusväli kvantifioi epävarmuuden selkeällä tavalla. Kun piste-estimaatti sekä luottamusväli lasketaan käytännön kannalta kiinnostavalle parametrille, niin saadaan tuloksia, jotka voidaan tulkita suoraan.
- Asiantuntemattomat lukijat tulkitsevat testien tulokset toisinaan aivan nurinkurisella tavalla.
- Testauksessa tutkijalla pitäisi olla selkeä käsitys testin voimafunktiosta sellaisilla parametrinarvoilla, jotka ovat käytännön kannalta merkityksellisiä.

Johtopäätös

- Vaikka testit ja luottamusvälit ovat duaalisia käsitteitä, niin
- sellaisissa yksinkertaisissa tilanteissa, joissa molemmat lähestymistavat ovat mahdollisia, luottamusvälien laskeminen on parempi tapa analysoida aineistoa kuin testaaminen.

6.7 Normaalijakautuneen populaation odotusarvon testaus, kun myös varianssi on tuntematon

- Jos sekä odotusarvo että varianssi ovat tuntemattomia, niin testit perustetaan saranasuureelle

$$t(\mathbf{Y}) = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}},$$

jolla on t-jakauma vapausasteluvulla $n-1$, mikäli $\mu = \mu_0$. Tästä tiedosta saadaan johdettua t-testit eri tapauksille matkimalla z-testien johtoa.

- Käytännössä t-testi suoritetaan aina jollakin tarkoitukseen sopivalla tietokoneohjelmalla. Esim. R-ohjelmistosa kaikki yksi- sekä kaksisuuntaiset t-testit saadaan laskettua vaivattomasti funktiolla `t.test`.
- Voimafunktion laskeminen t-testille on monimutkaisempaa kuin z-testille, mutta tämä kuitenkin onnistuu ns. epäkeskisen t-jakauman avulla. Tuloksena saadaan samantapaisia käyriä kuin z-testin voimafunktiolle.

t-testit ja t-luottamusvälit

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

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

Kaksisuuntaisen t-testin johto esimerkkinä

- Olkoon vastahypoteesi on kaksisuuntainen

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

  jossa μ_0 on kiinteä luku.
- Testisuure lasketaan kaavalla $t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$, jossa s on otoskeskihajonta.
- Vastaava satunnaismuuttuja $t(\mathbf{Y})$ noudattaa t-jakaumaa $n-1$ vapausasteella, ja testisuureen sekä pienet että suuret arvot ovat kriittisiä.
- Testi hylkää nollahypoteesin, mikäli $|t| > t_{n-1}(\frac{\alpha}{2})$.
- Sen p-arvo on

$$p = P(|T| \geq |t|) = 2(1 - F_{n-1}(|t|)),$$

  jossa satunnaismuuttuja $T \sim t_{n-1}$, ja F_ν tarkoittaa t_ν-jakauman kertymäfunktiota.

Testille duaalisen t-luottamusvälin ratkaiseminen

Ratkaistaan kaikki ne μ_0-arvot, joilla t-testi hyväksyisi nollahypoteesin $H_0 : \mu = \mu_0$ kaksisuuntaiselle vastahypoteesille:

$$\begin{aligned} & \left| \frac{\bar{y} - \mu_0}{s/\sqrt{n}} \right| \le t_{n-1}\left(\frac{\alpha}{2}\right) \\ \Leftrightarrow \quad & -t_{n-1}\left(\frac{\alpha}{2}\right) \le \frac{\mu_0 - \bar{y}}{s/\sqrt{n}} \le t_{n-1}\left(\frac{\alpha}{2}\right) \\ \Leftrightarrow \quad & \bar{y} - \frac{s}{\sqrt{n}}\, t_{n-1}\left(\frac{\alpha}{2}\right) \le \mu_0 \le \bar{y} + \frac{s}{\sqrt{n}}\, t_{n-1}\left(\frac{\alpha}{2}\right) \end{aligned}$$

Tulos on tuttu kaksisuuntaisen t-luottamusvälin kaava.

6.8 Binomijakauman parametrin testaus

- Jos tahdotaan testata lantin harhattomuutta, niin testi voidaan perustaa suoraan onnistumisten lukumäärälle $X \sim \mathrm{Bin}(n, p)$.
- Tässä tapauksessa hypoteesit ovat

$$H_0 : p = \frac{1}{2}, \qquad H_1 : p \neq \frac{1}{2}.$$

- Testi voidaan perustaa sille idealle, että tutkitaan havainnon poikkeamaa nollahypoteesin mukaisen jakauman odotusarvosta $EX = n/2$.
- Kaksisuuntaisen testin p-arvo voidaan laskea kaavalla

$$p = P_{1/2}\left(\left|X - \frac{n}{2}\right| \geq \left|x - \frac{n}{2}\right|\right).$$

Tässä alaindeksi $1/2$ tarkoittaa sitä, että oletamme nollahypoteesin mukaisesti, että $X \sim \mathrm{Bin}(n, 1/2)$, ja x on havaittu onnistumisten lukumäärä.

p-arvo normaaliapproksimaatiolla

- Tämän testin p-arvo saadaan laskettua karkeasti käyttämällä normaaliapproksimaatiota, jonka mukaan nollahypoteesin vallitessa satunnaismuuttujalla
$$\frac{X - EX}{\sqrt{\operatorname{var} X}} = \frac{X - \frac{n}{2}}{\sqrt{n \, \frac{1}{2} \, \frac{1}{2}}}$$
on osapuilleen standardinormaalijakauma $N(0, 1)$.
- Tämän takia p-arvo saadaan osapuilleen kaavalla
$$2\left(1 - \Phi\left(\frac{2}{\sqrt{n}}\left|x - \frac{n}{2}\right|\right)\right),$$
tai kaavalla
$$2\left(1 - \Phi\left(\frac{2}{\sqrt{n}}\left(\left|x - \frac{n}{2}\right| - \frac{1}{2}\right)\right)\right),$$
Jälkimmäisessä kaavassa tehtiin jatkuvuuskorjaus, ja Φ on $N(0, 1)$-jakauman kertymäfunktio.

Lantin harhattomuuden testaus normaaliapproksimaation avulla

Jos $n = 1000$ ja onnistumisia on tullut $x = 460$, niin nämä p-arvon normaaliapproksimaatiot ovat seuraavat.

```
n <- 1000
x <- 460
2 * pnorm(lower = FALSE, 2/sqrt(n) * abs(x - n/2))

## [1] 0.01141

2 * pnorm(lower = FALSE, 2/sqrt(n) * (abs(x - n/2) - 1/2))

## [1] 0.01248
```

Jälkimmäisessä komennossa tehtiin jatkuvuuskorjaus.

Lantin harhattomuuden testaus ilman approksimaatiota

Testissä tarvittavat binomijakauman häntätodennäköisyydet saadaan laskettua suoraan tietokoneohjelmilla. Esim., jos $n = 1000$ ja onnistumisia on $x = 460$, niin testi saadaan tehtyä komennolla

```
binom.test(460, 1000, p = 0.5)
```

```
##
##  Exact binomial test
##
## data:  460 and 1000
## number of successes = 460, number of trials = 1000, p-value =
## 0.01244
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4288 0.4915
## sample estimates:
## probability of success
##                   0.46
```

Testin yhteys Clopperin ja Pearsonin luottamusväliin

- Voimme samaan tapaan käsitellä muutkin kaksisuuntaiset testit
$$H_0 : p = p_0, \qquad H_1 : p \neq p_0,$$
- Kääntämällä näiden testien hyväksymisalueet saadaan jaksossa 5.8 mainittu Clopperin ja Pearsonin tarkka luottamusväli parametrille p.
- Yksisuuntaiset testit saadaan käsiteltyä samaan tapaan.

6.9 *p*-arvo ei ole todennäköisyys sille, että nollahypoteesi pitää paikkansa

- *P*-arvoa voidaan ajatella mittana sille, miten hyvin havainto on sopusoinnussa nollahypoteesin kanssa.
- Koska tämä käsite on vaikeatajuinen, tilastotieteen soveltajilla on kaikenlaisia harhakäsityksiä siitä, mitä *p*-arvo tarkoittaa.
- Eräs yleinen harhakäsitys on se, että *p*-arvo on todennäköisyys sille, että nollahypoteesi pitää paikkansa. Jotta tällaiseen täysin virheelliseen käsitykseen voisi päätyä, pitää tehdä monta räikeää käsitteellistä virhettä.
- Testin *p*-arvo ei ole todennäköisyys sille, että nollahypoteesi pitää paikkansa, vaikka näin lukijalle kerrotaan lukuisissa tilastotieteen soveltajille tarkoitetussa oppikirjoissa.

5.10 Tilastollisten testien väärinkäyttöä

- Monilla tilastotieteen soveltajilla on sellainen mielikuva, että kokeellisen tutkimuksen päämääränä on laskea p-arvo jollekin testille.
- Jos p-arvo on riittävän pieni, niin tuloksen saa julkaistua jossakin alan lehdessä. Jos p-arvo ei ole riittävän pieni, tutkimusta ei kannata lähettää arvioitavaksi, koska sitä ei kuitenkaan tulla julkaisemaan.
- Tämä harha ei koske yksinomaan yksittäisiä tutkijoita, vaan tällainen käsitys on ollut yleinen myös vaikutusvaltaisten lehtien arvioijien ja toimittajien parissa.

Julkaisuharha

- Tällainen käytäntö johtaa julkaisuharhaan (engl. *publication bias*): kirjallisuudessa julkaistaan enimmäkseen nollahypoteesin hylkääviä tutkimuksia riippumatta siitä, mikä todellisuudessa on asian laita.
- Tällainen käytäntö perustuu väärinkäsityksiin, rituaaleihin ja taikauskoon eikä sillä ole mitään tekemistä kunnollisen tieteellisen tutkimuksen kanssa eikä kunnollisen tilastotieteen soveltamisen kanssa.

Hölmöt nollahypoteesit

- Usein julkaisuissa testataan nollahypoteeseja, joista jo ennen tutkimuksen tekoa tiedetään, että ne eivät voi pitää paikkaansa.
- Nämä ovat ns. hölmöjä nollahypoteeseja (engl. *silly null*).
- Käsittelyllä on todellisuudessa kuitenkin aina jokin vaikutus, joten nollahypoteesi $\mu = 0$ (ei vaikutusta) ei voi pitää kirjaimellisesti paikkaansa, eikä kukaan oikeasti usko tätä tarkkaa, pistemäistä (engl. *sharp null*, *point null*) nollahypoteesia.
- Jos paikkansa pitämätöntä pistemäistä nollahypoteesia ei saada testillä hylättyä, niin syynä on se, että testillä ei ollut riittävästi voimaa, eli otoskoko oli liian pieni.

Tilastollinen merkitsevyys vs. käytännön merkittävyys

- Jos otoskokoa kasvatetaan, ja vihdoin pystytään nollahypoteesi hylkäämään, niin voi olla, että aineiston nojalla arvioitu vaikutuksen suuruus on niin pieni, että sillä ei ole mitään käytännön merkitystä.
- Kysymys: Jos nollahypoteesin mukaan $\mu = 0$, ja paras estimaattimme vaikutuksen suuruudelle on $\hat{\mu} = 0.1$, niin onko tällä erolla käytännössä merkitystä?
- Tähän kysymykseen tilastotiede ei pysty antamaan vastausta. Vastauksen pitää tulla substanssialan asiantuntijalta. (Mitä asiaa mitattiin? Mitä yksikköjä käytettiin? jne.)
- Tilastollinen merkitsevyys (engl. *statistical significance*) ja käytännön merkittävyys (engl. *practical significance*) ovat aivan eri asioita.

NHST

- On paljon hedelmällisempää ja informatiivisempaa yrittää estimoida vaikutuksen suuruutta ja yrittää kvantifioida estimaattiin liittyvää epävarmuutta (keskivirhe, luottamusväli!) kuin yrittää testata, onko vaikutus nolla.
- Tilastotieteen soveltajien intoon testata kaikkea mahdollista ja intoon tulkita virheellisesti testien tuloksia viitataan usein lyhenteellä NHST (*null hypothesis significance testing*).
- Hakusanan NHST avulla on helppo löytää tämän käytännön kritiikkiä.

Uudistusliike: uusi tilastotiede

Joillakin tilastotieteeseen vahvasti nojaavilla aloilla (esim. lääketiede, terveystieteet ja psykologia) on käynnissä uudistusliike, jossa pyritään pois epätarkoituksenmukaisesta hypoteesien testauksesta. Tämän sijasta

- lasketaan piste-estimaatteja ja väliestimaatteja vaikutuksen suuruudelle,
- yhdistetään aikaisempien tutkimusten tuloksia, eli harrastetaan meta-analyysia.

Jotkut kirjoittajat käyttävät tästä uudistusliikkeestä nimitystä uusi tilastotiede (engl. *new statistics*) — tilastotieteen näkökulmasta uudessa tilastotieteessä ei ole paljoa uutta.

Yhteenveto

- Nollahypoteesin hyväksyminen testissä ei tarkoita sitä, että oltaisiin löydetty todisteita nollahypoteesin puolesta. Se tarkoittaa sitä, että ei olla löydetty riittävän painavia todisteita nollahypoteesia vastaan.
- Testin tekemän päätöksen voi lukea p-arvosta, joka mittaa sitä kuinka hyvin aineisto on sopusoinnussa nollahypoteesin kanssa.
- Testin p-arvo ei ole todennäköisyys sille, että nollahypoteesi pitää paikkansa.
- Testaamisen sijasta kannattaa laskea piste-estimaatteja ja luottamusvälejä, mikäli tämä on mahdollista.