
Johdatus tilastotieteeseen

Testit suhdeasteikollisille muuttujille

Testit suhdeasteikollisille muuttujille

Testit normaalijakauman parametreille

Yhden otoksen t -testi

Kahden otoksen t -testi A

Kahden otoksen t -testi B

t -testi parivertailuille

Testi varianssille

Varianssien vertailutesti

Testit suhdeasteikollisille muuttujille: Mitä opimme? – 1/2

- Tarkastelemme tässä luvussa *normaalijakauman parametreja koskevia tilastollisia testejä*.
- Yhden otoksen testit:
 - t*-testi normaalijakauman odotusarvolle**
 - χ^2 -testi normaalijakauman varianssille**
- Kahden otoksen testit:
 - t*-testi A normaalijakaumien odotusarvojen vertailuun erisuurten varianssien tapauksessa**
 - t*-testi B normaalijakaumien odotusarvojen vertailuun yhtä suurten varianssien tapauksessa**
 - t*-testi normaalijakaumien odotusarvojen vertailuun parivertailutilanteessa**
 - F*-testi normaalijakauman varianssien vertailuun**

Testit suhdeasteikollisille muuttujille: Mitä opimme? – 2/2

- Tarkasteltavia testejä *normaalijakauman odotusarvolle* voidaan tavallisesti hyvin soveltaa myös *ei-normaalisille* havaintoaineistoille, *jos havaintojen lukumäärä on riittävän suuri.*

Testit suhdeasteikollisille muuttujille: Esitiedot

- Esitiedot: ks. seuraavia lukuja:
 - Tilastollisten aineistojen kerääminen ja mittaaminen**
 - Tilastollisten aineistojen kuvaaminen**
 - Otos ja otosjakaumat**
 - Estimointi**
 - Estimointimenetelmät**
 - Väliestimointi**
 - Tilastolliset testit**
- Tarvitset esitietoja myös seuraavista kalvokokoelman **Johdatus todennäköisyyslaskentaan** luvuista:
 - Satunnaismuuttujat ja todennäköisyysjakaumat**
 - Jakaumien tunnusluvut**
 - Jatkuvia jakaumia**
 - Normaalijakaumasta johdettuja jakaumia**

Testit suhdeasteikollisille muuttujille: Lisätiedot

- Testejä *järjestysasteikollisille muuttujille* käsitellään luvussa
Testit järjestysasteikollisille muuttujille
- Testejä *laatueroasteikollisille muuttujille* käsitellään luvussa
Testit laatueroasteikollisille muuttujille
- *Jakaumaoletuksien testaamista* käsitellään luvussa
**Yhteensopivuuden, homogeenisuuden ja riippumattomuuden
testaaminen**

Testit suhdeasteikollisille muuttujille

>> Testit normaalijakauman parametreille

Yhden otoksen t -testi

Kahden otoksen t -testi A

Kahden otoksen t -testi B

t -testi parivertailuille

Testi varianssille

Varianssien vertailutesti

Testit normaalijakauman parametreille

Avainsanat

Kahden otoksen testit
Normaalijakauma
Odotusarvo
Otos
Parametri
Riippumattomat otokset
Varianssi
Vertailutesti
Yhden otoksen testit

Normaalijakauman parametrien tilastolliset testit 1/2

- **Normaalijakauma** on *tilastotieteen tärkein jakauma*.
- Oletetaan, että satunnaismuuttuja X noudattaa normaalijakaumaa **parametrein** μ ja σ^2 :

$$X \sim N(\mu, \sigma^2)$$

- Tällöin

$$E(X) = \mu$$

on normaalijakauman *odotusarvo* ja

$$\text{Var}(X) = \sigma^2$$

on normaalijakauman *varianssi*.

- Parametrit μ ja σ^2 *määrittävät täysin normaalijakauman*.

Normaalijakauman parametrien tilastolliset testit 2/2

- *Normaalijakauman parametreja koskevat testit* voidaan jakaa kahteen ryhmään:
 - **Yhden otoksen testit**
 - **Kahden otoksen testit eli vertailutestit**
- *Yhden otoksen testeissä* testataan yksinkertaisia nollahypoteeseja, jotka koskevat normaalijakauman odotusarvo- tai varianssiparametria.
- *Kahden otoksen testit* ovat *vertailutestejä*, joilla verrataan kahden normaalijakauman odotusarvo- tai varianssi-parametreja toisiinsa.

Normaalijakauman parametreille tarkoitettujen testien yleinen soveltuvuus 1/2

- **Testejä normaalijakauman odotusarvolle** *sovelletaan usein myös sellaisissa tilanteissa, joissa havainnot eivät noudata normaalijakaumaa.*
- Tämä perustuu seuraaviin seikkoihin:
 - (i) Esitettävät testit odotusarvolle perustuvat havaintojen *aritmeettisiin keskiarvoihin.*
 - (ii) *Keskeisen raja-arvolauseen* mukaan myös ei-normaalisten havaintojen aritmeettiset keskiarvot noudattavat – tietyin ehdoin – *suurissa otoksissa approksimatiivisesti normaalijakaumaa.*

Testit normaalijakauman parametreille

Normaalijakauman parametreille tarkoitettujen testien yleinen soveltuvuus 2/2

- Sen sijaan **testit normaalijakauman varianssille** *eivät yleensä ole käyttökelpoisia ei-normaalille havainnoille* ja tilanne ei välttämättä parane suurillakaan havaintojen lukumäärillä.

Tavanomaiset testit normaalijakauman parametreille

- Tarkastelemme seuraavia *testejä normaalijakauman parametreille*:
 - **Yhden otoksen t -testi odotusarvolle**
 - **Kahden riippumattoman otoksen t -testi A odotusarvoille erisuurten varianssien tapauksessa**
 - **Kahden riippumattoman otoksen t -testi B odotusarvoille yhtä suurten varianssien tapauksessa**
 - **t -testi parivertailuille**
 - **Yhden otoksen χ^2 -testi varianssille**
 - **Kahden riippumattoman otoksen F -testi variansseille**

Testit suhdeasteikollisille muuttujille

Testit normaalijakauman parametreille

>> Yhden otoksen t -testi

Kahden otoksen t -testi A

Kahden otoksen t -testi B

t -testi parivertailuille

Testi varianssille

Varianssien vertailutesti

Yhden otoksen t -testi

Avainsanat

Aritmeettinen keskiarvo
Nollahypoteesi
Normaalijakauma
Odotusarvo
Otos
Otosvariassi
Parametri
Testisuure
Testisuureen jakauma
Testisuureen normaaliarvo
 t -jakauma
Variassi
Voimakkuus
Yhden otoksen testit
Yleinen hypoteesi

Testausasetelma 1/2

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos perusjoukosta S , joka noudattaa *normaalijakaumaa*

$$N(\mu, \sigma^2)$$

- Jakauma riippuu seuraavista parametreista:

$$\mu = \text{jakauman } \textit{odotusarvo}$$

$$\sigma^2 = \text{jakauman } \textit{varianssi}$$

Testausasetelma 2/2

- Asetetaan normaalijakauman $N(\mu, \sigma^2)$ *odotusarvo-* eli *paikkaparametrille* μ *nollahypoteesi*

$$H_0 : \mu = \mu_0$$

- Testausongelma:

Ovatko havainnot *sopusoinnussa* nollahypoteesin H_0 kanssa?

- Ongelman ratkaisuna on **yhden otoksen t -testi**.

Hypoteesit

- *Yleinen hypoteesi* H :

(i) Havainnot $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$

(ii) Havainnot X_1, X_2, \dots, X_n ovat *riippumattomia*

- *Nollahypoteesi* H_0 :

$$H_0 : \mu = \mu_0$$

- *Vaihtoehtoinen hypoteesi* H_1 :

$$\left. \begin{array}{l} H_1 : \mu > \mu_0 \\ H_1 : \mu < \mu_0 \end{array} \right\} \text{1-suuntaiset vaihtoehtoiset hypoteesit}$$

$$H_1 : \mu \neq \mu_0 \quad \text{2-suuntainen vaihtoehtoinen hypoteesi}$$

Parametrien estimointi

- Olkoot

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

ja

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

tavanomaiset *harhattomat estimaattorit* parametreille

$$E(X_i) = \mu, i = 1, 2, \dots, n$$

ja

$$\text{Var}(X_i) = \sigma^2, i = 1, 2, \dots, n$$

Testisuure ja sen jakauma

- Määritellään t -testisuure

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

- Jos nollahypoteesi

$$H_0 : \mu = \mu_0$$

pätee, niin testisuure t noudattaa *Studentin t -jakaumaa* vapausastein $(n - 1)$:

$$t \sim t(n - 1)$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 1/2

- Oletetaan, että testin yleinen hypoteesi H ja nollahypoteesi H_0 pätevät:

$$X_1, X_2, \dots, X_n \perp$$

$$X_i \sim N(\mu_0, \sigma^2), i = 1, 2, \dots, n$$

- Koska tällöin (ks. lukua **Otos ja otosjakaumat**)

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

niin

$$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$$

- Koska standardipoikkeama σ on *tuntematon*, satunnaismuuttujan z lauseke on testisuureena *epäoperationaalinen*.

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 2/2

- Jos standardipoikkeama σ korvataan satunnaismuuttujan z lausekkeessa vastaavalla *otossuureella*

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

niin saadaan *t-testisuure*

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

joka nollahypoteesin H_0 pätiessä noudattaa Studentin t -jakaumaa vapausastein $(n - 1)$:

$$t \sim t(n - 1)$$

- Todistus: ks. lukua **Otos ja otosjakaumat**.

t -testisuure mittaa tilastollista etäisyyttä

- Testisuure

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

mittaa havaintoarvojen aritmeettisen keskiarvon ja nollahypoteesin $H_0 : \mu = \mu_0$ kiinnittämän odotusarvoparametrin μ arvon μ_0 tilastollista etäisyyttä.

- *Mittayksikkönä* on erotuksen $\bar{X} - \mu_0$ standardipoikkeaman

$$\sigma / \sqrt{n}$$

estimaattori, jota määrättäessä on oletettu, että nollahypoteesi H_0 pätee.

Yhden otoksen t -testi

Testi

- Testisuureen

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

normaaliarvo = 0, koska *nollahypoteesin* $H_0 : \mu = \mu_0$
pätiessä

$$E(t) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen t arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.

Testin hylkäysalueen valinta 1/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_1 : \mu > \mu_0$$

niin *kriittinen raja* $+t_\alpha$ saadaan ehdosta

$$\Pr(t \geq +t_\alpha) = \alpha$$

jossa

$$t \sim t(n - 1)$$

- Testin **hylkäysalue** on tällöin muotoa
 $(+t_\alpha, +\infty)$

Testin hylkäysalueen valinta 2/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_1 : \mu < \mu_0$$

niin *kriittinen raja* $-t_\alpha$ saadaan ehdosta

$$\Pr(t \leq -t_\alpha) = \alpha$$

jossa

$$t \sim t(n - 1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(-\infty , -t_\alpha)$$

Testin hylkäysalueen valinta 3/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

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

niin *kriittiset rajat* $-t_{\alpha/2}$ ja $+t_{\alpha/2}$ saadaan ehdoista

$$\Pr(t \leq -t_{\alpha/2}) = \alpha/2$$

$$\Pr(t \geq +t_{\alpha/2}) = \alpha/2$$

jossa

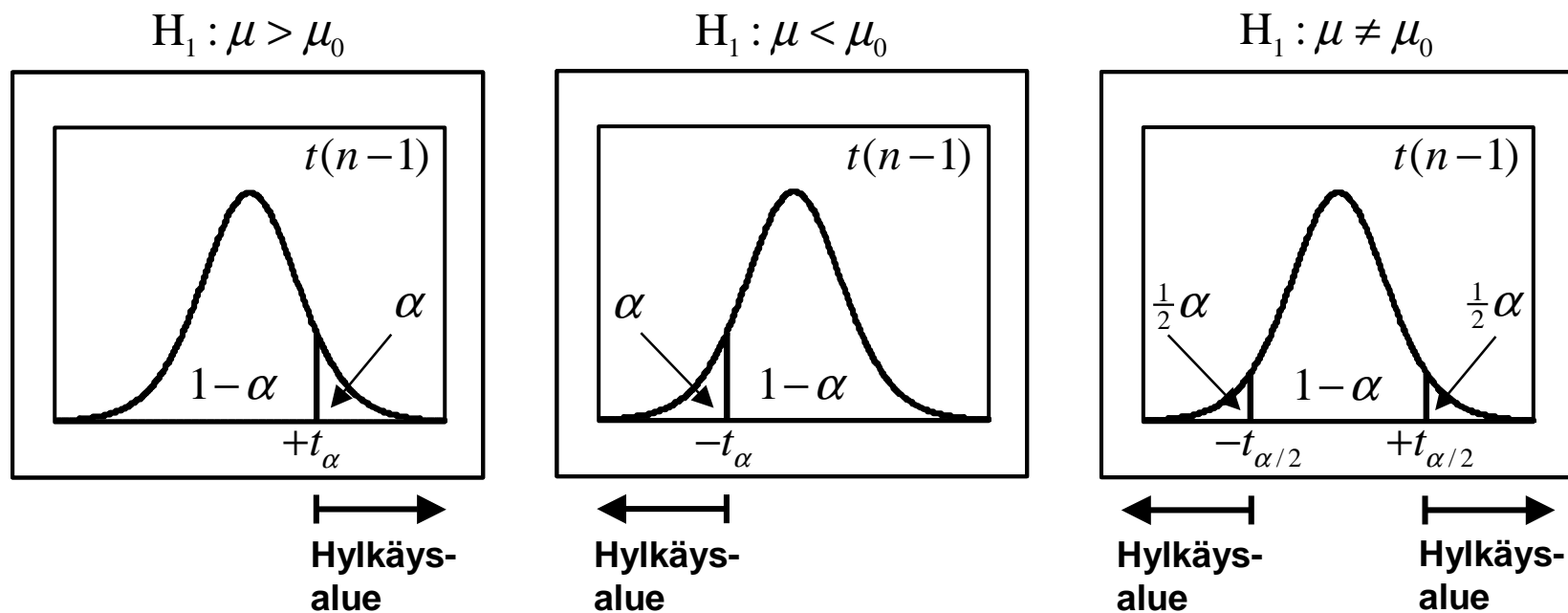
$$t \sim t(n - 1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(-\infty, -t_{\alpha}) \cup (+t_{\alpha}, +\infty)$$

Testin hylkäysalueen valinta 4/4

- Oletetaan, että testin *merkitsevyystasoksi* on valittu α .
- Testin **hylkäysalueen** määrittäminen voidaan havainnollistaa alla olevilla kuvioilla.

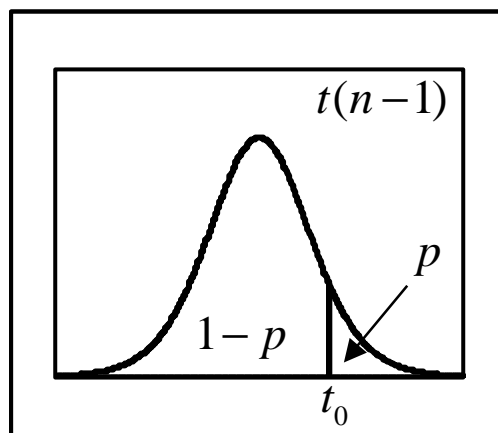


Yhden otoksen t -testi

Testin p -arvo

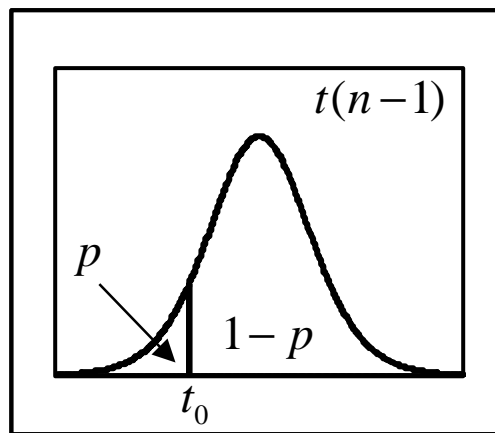
- Olkoon t -testisuureen havaittu arvo t_0 .
- Testin p -arvon määrittämistä voidaan havainnollistaa alla olevilla kuvioilla.

$$H_1 : \mu > \mu_0$$



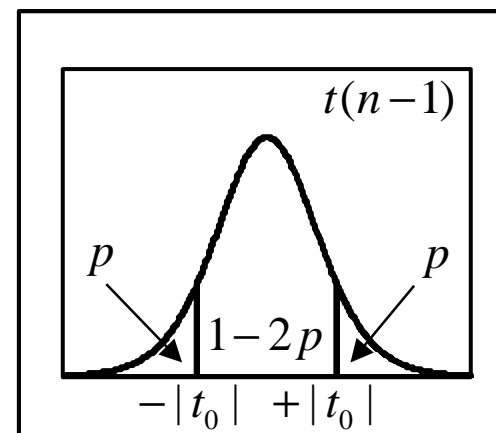
Testin p -arvo = p

$$H_1 : \mu < \mu_0$$



Testin p -arvo = p

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



Testin p -arvo = $2p$

Normaalisuusoletuksen merkitys 1/2

- Yhden otoksen t -testin yleisessä hypoteesissa oletetaan, että havainnot ovat *normaalijakautuneita*.
- t -testi *ei* kuitenkaan *ole herkkä poikkeamille normaalisuudesta*, jos havaintojen lukumäärä n on ”*kyllin suuri*”.

Normaalisuusoletuksen merkitys 2/2

- *Testiä on melko turvallista käyttää, kun havaintojen lukumäärä*

$$n > 15$$

ellei havaintojen jakauma ole kovin vino ja havaintojen joukossa ole poikkeavia havaintoja.

- *Jos havaintojen lukumäärä*

$$n > 40$$

testiä voidaan melko turvallisesti käyttää jopa selvästi vinoille havaintojen jakaumille.

Testin hyväksymisvirheen todennäköisyys ja voimakkuus 1/6

- Tarkastellaan t -testin *hyväksymisvirheen todennäköisyyttä* ja *voimakkuutta* tilanteessa, jossa normaalijakauman $N(\mu, \sigma^2)$ **varianssi σ^2 oletetaan tunnetuksi**.
- Olkoon *nollahypoteesi* muotoa

$$H_0 : \mu = \mu_0$$

ja *vaihtoehtoinen hypoteesi* muotoa

$$H_1 : \mu < \mu_0$$

- Huomautus:

Jos normaalijakauman varianssia σ^2 *ei oleteta tunnetuksi*, vaatii t -testin voimakkuuden määrittäminen ns. *epäkeskisen t -jakauman* määrittelemistä.

Tämän yleisen tapauksen käsittely sivuutetaan tässä esityksessä.

Testin hyväksymisvirheen todennäköisyys ja voimakkuus 2/6

- t -testisuure

$$t = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

noudattaa nollahypoteesin

$$H_0 : \mu = \mu_0$$

pätiessä standardoitua normaalijakaumaa (ks. lukua **Otos ja otosjakaumat**):

$$t \sim N(0, 1)$$

Testin hyväksymisvirheen todennäköisyys ja voimakkuus 3/6

- *Vaihtoehtoisen hypoteesin*

$$H_1 : \mu < \mu_0$$

tapauksessa t -testin päätössääntö on muotoa:

Hylkää nollahypoteesi

$$H_0 : \mu = \mu_0$$

jos

$$t = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} < -z_\alpha$$

- *Kriittinen raja $-z_\alpha$ saadaan ehdosta*

$$\Pr(z \leq -z_\alpha) = \alpha$$

jossa $z \sim N(0, 1)$.

Testin hyväksymisvirheen todennäköisyys ja voimakkuus 4/6

- *Vaihtoehtoisen hypoteesin*

$$H_1 : \mu < \mu_0$$

tapauksessa t -testin päätössääntö voidaan kirjoittaa myös seuraavaan muotoon:

Hylkää nollahypoteesi

$$H_0 : \mu = \mu_0$$

jos

$$\bar{X} < \mu_0 - z_\alpha \sigma / \sqrt{n} = \bar{X}_c$$

- *Kriittinen raja $-z_\alpha$ saadaan ehdosta*

$$\Pr(z \leq -z_\alpha) = \alpha$$

jossa $z \sim N(0, 1)$.

Testin hyväksymisvirheen todennäköisyys ja voimakkuus 5/6

- *Vaihtoehdoisen hypoteesin*

$$H_1 : \mu < \mu_0$$

tapauksessa t -testin hyväksymisvirheen todennäköisyys β on ehdollinen todennäköisyys

$$\beta = \Pr(H_0 \text{ jätetään voimaan} \mid H_0 \text{ ei ole tosi})$$

$$= \Pr(\bar{X} \geq \bar{X}_c \mid \mu = \mu^*)$$

$$= \Pr\left(z \geq \frac{\bar{X}_c - \mu^*}{\sigma / \sqrt{n}}\right)$$

jossa

$$\mu^* \neq \mu_0$$

Testin hyväksymisvirheen todennäköisyys ja voimakkuus 6/6

- *Vaihtoehdoisen hypoteesin*

$$H_1 : \mu < \mu_0$$

tapauksessa t -testin **voimakkuus** $1 - \beta$ on ehdollinen todennäköisyys

$$1 - \beta = \Pr(H_0 \text{ hylätään} \mid H_0 \text{ ei ole tosi})$$

$$= \Pr(\bar{X} < \bar{X}_c \mid \mu = \mu^*)$$

$$= \Pr\left(z < \frac{\bar{X}_c - \mu^*}{\sigma / \sqrt{n}}\right)$$

jossa

$$\mu^* \neq \mu_0$$

Testin hyväksymisvirheen todennäköisyys ja voimakkuus: Havainnollistus 1/3

- Kuvio oikealla havainnollistaa t -testin hyväksymisvirheen todennäköisyyttä β ja voimakkuutta $1 - \beta$.

- Yleinen hypoteesi H :

$$X_1, X_2, \dots, X_n \perp$$

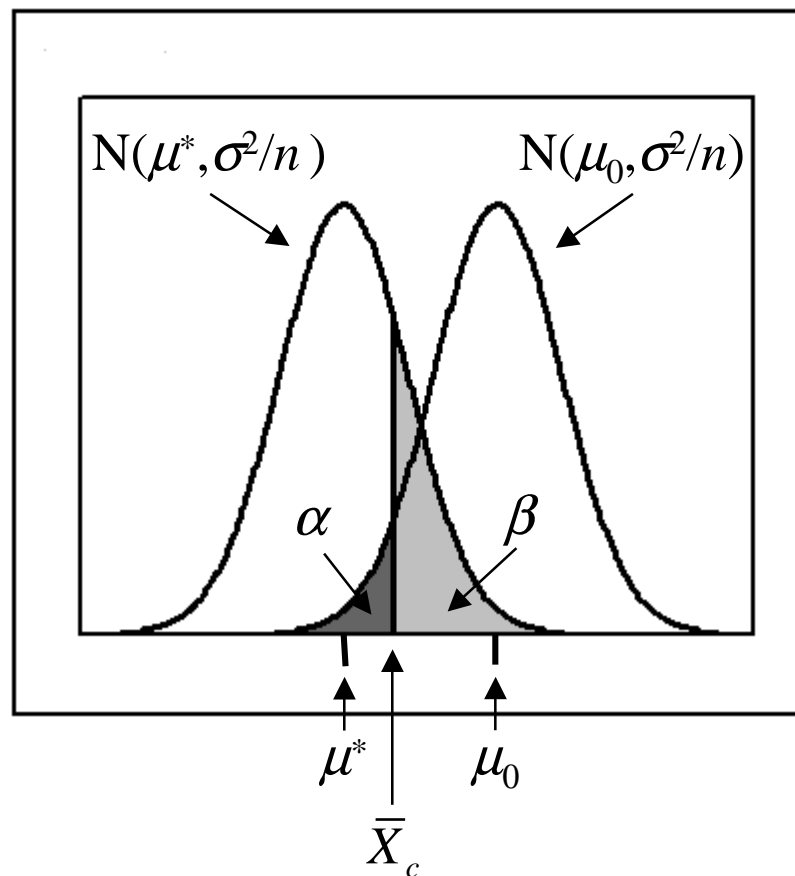
$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$$

- Nollahypoteesi H_0 :

$$\mu = \mu_0$$

- Vaihtoehtoinen hypoteesi H_1 :

$$\mu < \mu_0$$



Testin hyväksymisvirheen todennäköisyys ja voimakkuus: Havainnollistus 2/3

- Valitaan *merkitsevyystasoksi* α .

- *Kriittinen raja* z_α :

$$\Pr(z \leq -z_\alpha) = \alpha$$

$$z \sim N(0, 1)$$

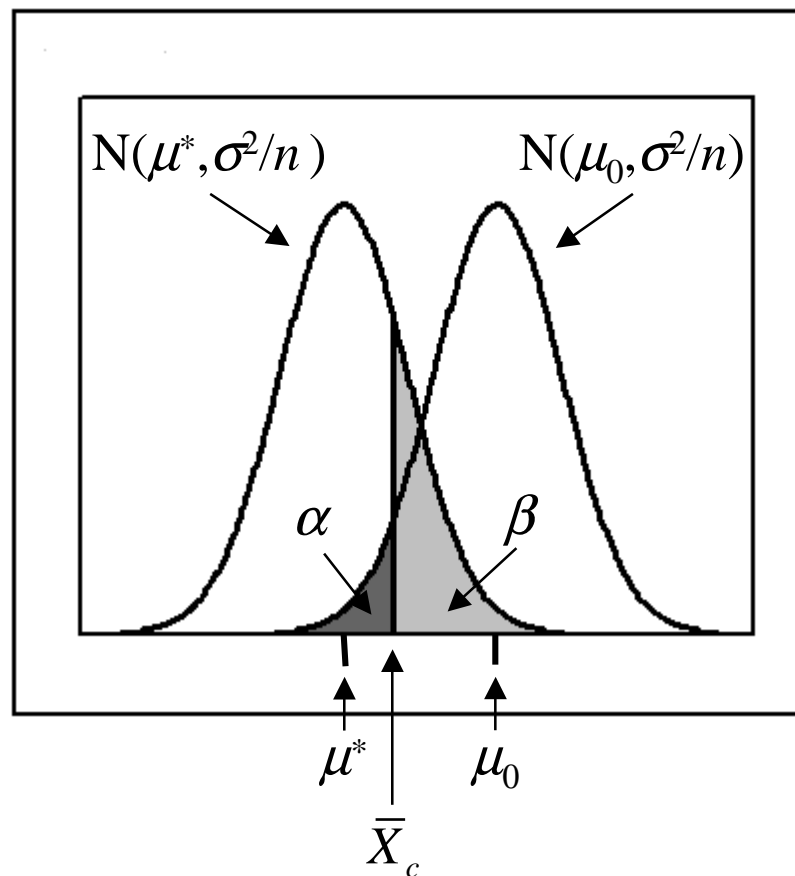
- *Kriittinen raja* \bar{X}_c :

$$\bar{X}_c = \mu_0 - z_\alpha \sigma / \sqrt{n}$$

- *Päätössääntö*:

Hylkää nollahypoteesi H_0 , jos

$$\bar{X} < \bar{X}_c$$



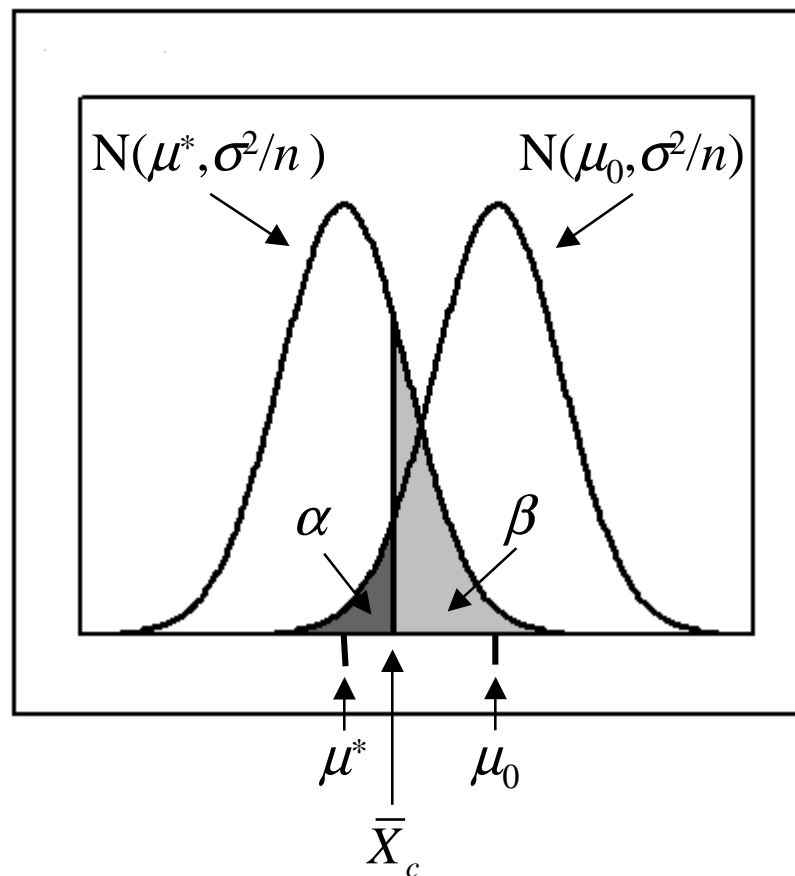
Testin hyväksymisvirheen todennäköisyys ja voimakkuus: Havainnollistus 3/3

- *Hyväksymisvirheen todennäköisyys* β :

$$\beta = \Pr(\bar{X} \geq \bar{X}_c \mid \mu = \mu^*)$$

- *Voimakkuus* $1 - \beta$:

$$1 - \beta = \Pr(\bar{X} < \bar{X}_c \mid \mu = \mu^*)$$



Testit suhdeasteikollisille muuttujille

Testit normaalijakauman parametreille

Yhden otoksen t -testi

>> Kahden otoksen t -testi A

Kahden otoksen t -testi B

t -testi parivertailuille

Testi varianssille

Varianssien vertailutesti

Kahden otoksen t -testi A

Avainsanat

Aritmeettinen keskiarvo
Kahden otoksen testit
Nollahypoteesi
Normaalijakauma
Odotusarvo
Otosvariassi
Parametri
Riippumattomat otokset
Testisuure
Testisuureen jakauma
Testisuureen normaaliarvo
 t -jakauma
Variassi
Vertailutesti
Yleinen hypoteesi

Testausasetelma 1/4

- Olkoon

$$X_{11}, X_{21}, \dots, X_{n_11}$$

yksinkertainen satunnaisotos perusjoukosta S_1 , joka noudattaa *normaalijakaumaa*

$$N(\mu_1, \sigma_1^2)$$

- Jakauma riippuu seuraavista parametreista:

$$\mu_1 = \text{jakauman } \textit{odotusarvo}$$

$$\sigma_1^2 = \text{jakauman } \textit{varianssi}$$

Testausasetelma 2/4

- Olkoon

$$X_{12}, X_{22}, \dots, X_{n_22}$$

yksinkertainen satunnaisotos perusjoukosta S_2 , joka noudattaa *normaalijakaumaa*

$$N(\mu_2, \sigma_2^2)$$

- Jakauma riippuu seuraavista parametreista:

$$\mu_2 = \text{jakauman } \textit{odotusarvo}$$

$$\sigma_2^2 = \text{jakauman } \textit{varianssi}$$

Testausasetelma 3/4

- Oletetaan lisäksi, että perusjoukosta S_1 poimittu otos

$$X_{11}, X_{21}, \dots, X_{n_11}$$

ja perusjoukosta S_2 poimittu otos

$$X_{12}, X_{22}, \dots, X_{n_22}$$

ovat toisistaan *riippumattomia*.

- Otosten riippumattomuus merkitsee sitä, että se mikä alkio poimitaan perusjoukosta S_1 *ei vaikuta* siihen mikä alkioista poimitaan perusjoukosta S_2 ja kääntäen.

Testausasetelma 4/4

- Asetetaan normaalijakaumien $N(\mu_1, \sigma_1^2)$ ja $N(\mu_2, \sigma_2^2)$ odotusarvo- eli paikkaparametreille μ_1 ja μ_2 nollahypoteesi

$$H_0 : \mu_1 = \mu_2 = \mu$$

- Testausongelma:

Ovatko havainnot *sopusoinnussa* nollahypoteesin H_0 kanssa?

- Ongelman ratkaisuna on **kahden riippumattoman otoksen t -testi** erisuurten varianssien tapauksessa.

- Huomautus:

Jos voidaan olettaa, että $\sigma_1^2 = \sigma_2^2$, testauksessa *kannattaa käyttää* kahden riippumattoman otoksen t -testiä B.

Yleinen hypoteesi

- *Yleinen hypoteesi* H :
 - (1) Havainnot $X_{i1} \sim N(\mu_1, \sigma_1^2)$, $i = 1, 2, \dots, n_1$
 - (2) Havainnot $X_{j2} \sim N(\mu_2, \sigma_2^2)$, $j = 1, 2, \dots, n_2$
 - (3) Havainnot X_{i1} ja X_{j2} ovat *riippumattomia* kaikille i ja j .
- Huomautuksia:
 - Oletus (3) sisältää *kolme riippumattomuusoletusta*:
 - Havainnot ovat riippumattomia otoksien 1 ja 2 *sisällä*.
 - Havainnot ovat riippumattomia otoksien 1 ja 2 *välillä*.
 - Jakaumien variansseja ja *ei ole oletettu yhtä suuriksi*;
vrt. kahden otoksen t -testi B.

Nollahypoteesi ja vaihtoehtoiset hypoteesit

- *Nollahypoteesi* H_0 :

$$H_0 : \mu_1 = \mu_2 = \mu$$

- *Vaihtoehtoinen hypoteesi* H_1 :

$$\left. \begin{array}{l} H_1 : \mu_1 > \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{array} \right\} \text{1-suuntaiset vaihtoehtoiset hypoteesit}$$

$$H_1 : \mu_1 \neq \mu_2 \quad \text{2-suuntainen vaihtoehtoinen hypoteesi}$$

Parametrien estimointi

- Olkoot

$$\bar{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ik}, k = 1, 2$$

ja

$$s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)^2, k = 1, 2$$

tavanomaiset *harhattomat estimaattorit* parametreille

$$E(X_{ik}) = \mu_k, i = 1, 2, \dots, n_k, k = 1, 2$$

ja

$$\text{Var}(X_{ik}) = \sigma_k^2, i = 1, 2, \dots, n_k, k = 1, 2$$

Testisuure ja sen asymptoottinen jakauma

- Määritellään t -testisuure

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Jos nollahypoteesi

$$H_0 : \mu_1 = \mu_2 = \mu$$

pätee, niin testisuure t noudattaa suurissa otoksissa approksimatiivisesti standardoitua normaalijakaumaa $N(0,1)$:

$$t \sim_a N(0,1)$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 1/3

- Oletetaan, että testin yleinen hypoteesi H ja nollahypoteesi H_0 pätevät:

$$X_{11}, X_{21}, \dots, X_{n_1 1}, X_{12}, X_{22}, \dots, X_{n_2 2} \perp$$

$$X_{i1} \sim N(\mu, \sigma_1^2), i = 1, 2, \dots, n_1$$

$$X_{j2} \sim N(\mu, \sigma_2^2), j = 1, 2, \dots, n_2$$

- Tällöin (ks. lukua **Otos ja otosjakaumat**)

$$\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{i1} \sim N\left(\mu, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{X}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} X_{j2} \sim N\left(\mu, \frac{\sigma_2^2}{n_2}\right)$$

- Koska $\bar{X}_1 \perp \bar{X}_2$, niin

$$\bar{X}_1 - \bar{X}_2 \sim N\left(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 2/3

- Edellä esitetystä seuraa, että

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

- Koska varianssit σ_1^2 ja σ_2^2 ovat *tuntemattomia*, satunnaismuuttujan z lauseke on testisuureena *epäoperationaalinen*.

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 3/3

- Jos satunnaismuuttujan z lausekkeessa varianssit σ_1^2 ja σ_2^2 korvataan vastaavilla *otossuureilla*

$$s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)^2, \quad k = 1, 2$$

niin saadaan *t-testisuure*

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

joka nollahypoteesin H_0 pätiessä noudattaa suurissa otoksissa *approksimatiivisesti standardoitua normaalijakaumaa* $N(0, 1)$:

$$t \sim_a N(0, 1)$$

- Todistus sivuutetaan.

Testisuureen jakauman approksimointi

- *Pienissä otoksissa* saadaan testisuureen t jakaumalle *parempi approksimaatio* käyttämällä approksimaationa *Studentin t -jakaumaa* vapausastein (ns. Satterthwaiten approksimaatio)

$$v = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right]^2}{\frac{1}{n_1 - 1} \left[\frac{s_1^2}{n_1} \right]^2 + \frac{1}{n_2 - 1} \left[\frac{s_2^2}{n_2} \right]^2}$$

t -testisuure mittaa tilastollista etäisyyttä

- Testisuure

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

mittaa otoksien 1 ja 2 aritmeettisten keskiarvojen *tilastollista etäisyyttä*.

- *Mittayksikkönä* on erotuksen $\bar{X}_1 - \bar{X}_2$ standardipoikkeaman

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

estimaattori, jota määrättäessä on oletettu, että nollahypoteesi H_0 pätee.

Kahden otoksen t -testi A

Testi

- Testisuureen

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

normaaliarvo = 0, koska *nollahypoteesin* $H_0 : \mu_1 = \mu_2 = \mu$ *pätiessä*

$$E(t) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen t arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.

Testin hylkäysalueen valinta 1/5

- Käsittelemme seuraavassa *kahden otoksen t -testin A hylkäysalueen valintaa*, jos testisuureen approksimoidaan *standardoidulla normaalijakaumalla* $N(0, 1)$.
- Jos kahden otoksen t -testin A testisuuretta approksimoidaan *Studentin t -jakaumalla*, jossa vapausasteiden lukumäärä ν lasketaan Satterthwaiten kaavan mukaan, testin hylkäysalue määrätään samalla tavalla kuin yhden otoksen t -testin tapauksessa.

Testin hylkäysalueen valinta 2/5

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_1 : \mu > \mu_0$$

niin *kriittinen raja* $+t_\alpha$ saadaan ehdosta

$$\Pr(t \geq +t_\alpha) = \alpha$$

jossa

$$t \sim_a N(0, 1)$$

- Testin **hylkäysalue** on tällöin muotoa
 $(+t_\alpha, +\infty)$

Testin hylkäysalueen valinta 3/5

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_1 : \mu < \mu_0$$

niin *kriittinen raja* $-t_\alpha$ saadaan ehdosta

$$\Pr(t \leq -t_\alpha) = \alpha$$

jossa

$$t \sim_a N(0, 1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(-\infty, -t_\alpha)$$

Testin hylkäysalueen valinta 4/5

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

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

niin *kriittiset rajat* $-t_{\alpha/2}$ ja $+t_{\alpha/2}$ saadaan ehdoista

$$\Pr(t \leq -t_{\alpha/2}) = \alpha/2$$

$$\Pr(t \geq +t_{\alpha/2}) = \alpha/2$$

jossa

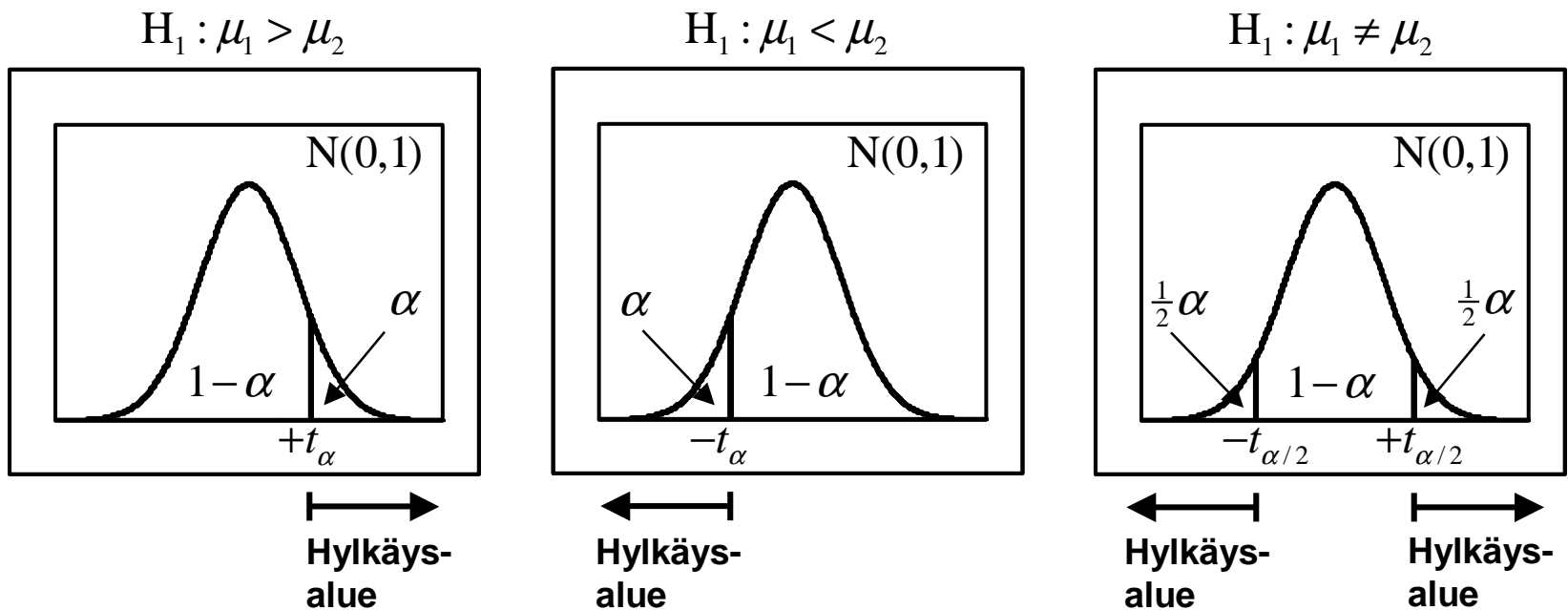
$$t \sim_a N(0,1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(-\infty, -t_{\alpha}) \cup (+t_{\alpha}, +\infty)$$

Testin hylkäysalueen valinta 5/5

- Oletetaan, että testin *merkitsevyystasoksi* on valittu α .
- Testin **hylkäysalueen** määrittäminen voidaan havainnollistaa alla olevilla kuvioilla.

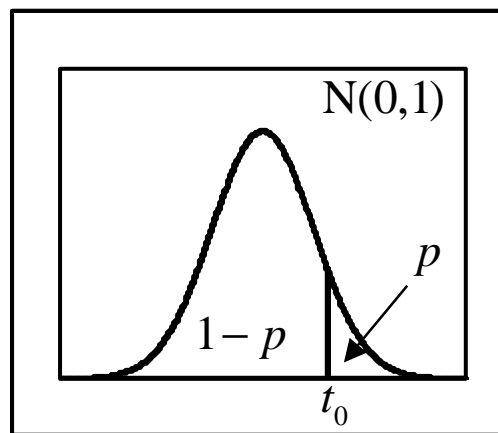


Kahden otoksen t -testi A

Testin p -arvo

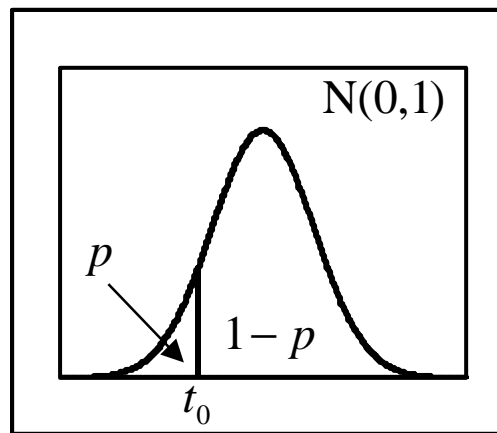
- Olkoon t -testisuureen havaittu arvo t_0 .
- Testin p -arvon määrittämistä voidaan havainnollistaa alla olevilla kuvioilla.

$$H_1 : \mu_1 > \mu_2$$



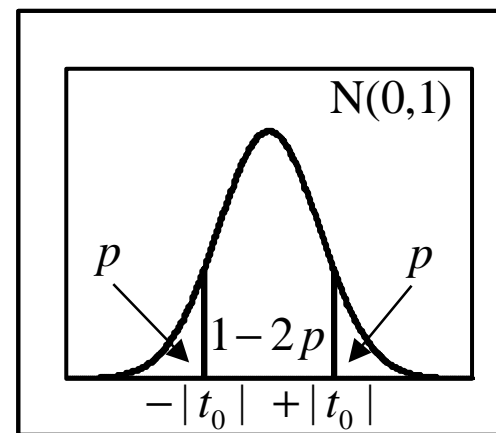
Testin p -arvo = p

$$H_1 : \mu_1 < \mu_2$$



Testin p -arvo = p

$$H_1 : \mu_1 \neq \mu_2$$



Testin p -arvo = $2p$

Normaalisuusoletuksen merkitys 1/2

- Kahden otoksen t -testin A yleisen hypoteesin mukaan havainnot ovat molemmissa otoksissa *normaalijakautuneita*.
- Testi *ei* kuitenkaan *ole herkkä poikkeamille normaalisuudesta*, jos molempien otosten otoskoot ovat ”*kyllin suuria*”.

Normaalisuusoletuksen merkitys 2/2

- *Testiä on melko turvallista käyttää, kun*

$$n_1 > 15 \text{ ja } n_2 > 15$$

ja n_1 ja n_2 eivät eroa toisistaan kovin paljon, *elleivät* havaintojen jakaumat *ole kovin vinoja* ja *ellei* havaintojen joukossa *ole poikkeavia havaintoja*.

- Jos

$$n_1 > 40 \text{ ja } n_2 > 40$$

testiä voidaan melko turvallisesti käyttää jopa selvästi vinoille havaintojen jakaumille.

Testit suhdeasteikollisille muuttujille

Testit normaalijakauman parametreille

Yhden otoksen t -testi

Kahden otoksen t -testi A

>> Kahden otoksen t -testi B

t -testi parivertailuille

Testi varianssille

Varianssien vertailutesti

Kahden otoksen t -testi B

Avainsanat

Aritmeettinen keskiarvo
Kahden otoksen testit
Nollahypoteesi
Normaalijakauma
Odotusarvo
Otosvariassi
Parametri
Riippumattomat otokset
Testisuure
Testisuureen jakauma
Testisuureen normaaliarvo
 t -jakauma
Variassi
Vertailutesti
Yleinen hypoteesi

Testausasetelma 1/4

- Olkoon

$$X_{11}, X_{21}, \dots, X_{n_11}$$

yksinkertainen satunnaisotos perusjoukosta S_1 , joka noudattaa *normaalijakaumaa*

$$N(\mu_1, \sigma^2)$$

- Jakauma riippuu seuraavista parametreista:

$$\mu_1 = \text{jakauman } \textit{odotusarvo}$$

$$\sigma^2 = \text{jakauman } \textit{varianssi}$$

Testausasetelma 2/4

- Olkoon

$$X_{12}, X_{22}, \dots, X_{n_22}$$

yksinkertainen satunnaisotos perusjoukosta S_2 , joka noudattaa *normaalijakaumaa*

$$N(\mu_2, \sigma^2)$$

- Jakauma riippuu seuraavista parametreista:

$$\mu_2 = \text{jakauman } \textit{odotusarvo}$$

$$\sigma^2 = \text{jakauman } \textit{varianssi}$$

Testausasetelma 3/4

- Oletetaan lisäksi, että perusjoukosta S_1 poimittu otos

$$X_{11}, X_{21}, \dots, X_{n_11}$$

ja perusjoukosta S_2 poimittu otos

$$X_{12}, X_{22}, \dots, X_{n_22}$$

ovat toisistaan *riippumattomia*.

- Otosten riippumattomuus merkitsee sitä, että se mikä alkio poimitaan perusjoukosta S_1 *ei vaikuta* siihen mikä alkioista poimitaan perusjoukosta S_2 ja kääntäen.

Testausasetelma 4/4

- Asetetaan normaalijakaumien $N(\mu_1, \sigma^2)$ ja $N(\mu_2, \sigma^2)$ odotusarvo- eli paikkaparametreille μ_1 ja μ_2 nollahypoteesi

$$H_0 : \mu_1 = \mu_2 = \mu$$

- Testausongelma:

Ovatko havainnot *sopusoinnussa* nollahypoteesin H_0 kanssa?

- Ongelman ratkaisuna on **kahden riippumattoman otoksen t -testi** yhtä suurten varianssien tapauksessa.
- Huomautus:

Jos jakaumien varianssit eivät ole yhtä suuret, testauksessa on käytettävä kahden riippumattoman otoksen t -testiä A.

Yleinen hypoteesi

- *Yleinen hypoteesi* H :
 - (1) $X_{i1} \sim N(\mu_1, \sigma^2)$, $i = 1, 2, \dots, n_1$
 - (2) $X_{j2} \sim N(\mu_2, \sigma^2)$, $j = 1, 2, \dots, n_2$
 - (3) Havainnot X_{i1} ja X_{j2} ovat *riippumattomia* kaikille i ja j
- Huomautuksia:
 - Oletus (3) sisältää *kolme riippumattomuusoletusta*:
 - Havainnot ovat riippumattomia otoksien 1 ja 2 *sisällä*.
 - Havainnot ovat riippumattomia otoksien 1 ja 2 *välillä*.
 - Jakaumien varianssit *on* tässä oletettu yhtä suuriksi;
vrt. kahden otoksen t -testi A.

Nollahypoteesi ja vaihtoehtoiset hypoteesit

- *Nollahypoteesi* H_0 :

$$H_0 : \mu_1 = \mu_2 = \mu$$

- *Vaihtoehtoinen hypoteesi* H_1 :

$$\left. \begin{array}{l} H_1 : \mu_1 > \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{array} \right\} \text{1-suuntaiset vaihtoehtoiset hypoteesit}$$

$$H_1 : \mu_1 \neq \mu_2 \quad \text{2-suuntainen vaihtoehtoinen hypoteesi}$$

Parametrien estimointi

- Olkoot

$$\bar{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ik}, k = 1, 2$$

ja

$$s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)^2, k = 1, 2$$

tavanomaiset *harhattomat estimaattorit* parametreille

$$E(X_{ik}) = \mu_k, i = 1, 2, \dots, n_k, k = 1, 2$$

ja

$$\text{Var}(X_{ik}) = \sigma^2, i = 1, 2, \dots, n_k, k = 1, 2$$

Yhdistetty varianssiestimaattori

- Määritellään ns. **yhdistetty varianssiestimaattori**

$$s_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Yhdistetty varianssiestimaattori s_P^2 on *harhaton estimaattori* varianssiparametrille σ^2 , jos nollahypoteesi $H_0 : \mu_1 = \mu_2 = \mu$ pätee.

- Huomautus:

Yhdistetty varianssiestimaattori s_P^2 ei ole sama kuin yhdistetyn otoksen varianssi, koska otoskeskiarvot \bar{X}_1 ja \bar{X}_2 eivät (yleensä) ole yhtä suuria.

Testisuure ja sen jakauma

- Määritellään t -testisuure

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Jos nollahypoteesi

$$H_0 : \mu_1 = \mu_2 = \mu$$

pätee, niin testisuure t noudattaa *Studentin t -jakaumaa* vapausastein $(n_1 + n_2 - 2)$:

$$t \sim t(n_1 + n_2 - 2)$$

Testisuureen jakauma nollassa hypoteesin H_0 pätiessä: Perustelu 1/3

- Oletetaan, että testin yleinen hypoteesi H ja nollassa hypoteesi H_0 pätevät:

$$X_{11}, X_{21}, \dots, X_{n_1 1}, X_{12}, X_{22}, \dots, X_{n_2 2} \perp$$

$$X_{i1} \sim N(\mu, \sigma^2), i = 1, 2, \dots, n_1$$

$$X_{j2} \sim N(\mu, \sigma^2), j = 1, 2, \dots, n_2$$

- Tällöin (ks. lukua **Otos ja otosjakaumat**)

$$\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{i1} \sim N\left(\mu, \frac{\sigma^2}{n_1}\right)$$

$$\bar{X}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} X_{j2} \sim N\left(\mu, \frac{\sigma^2}{n_2}\right)$$

- Koska $\bar{X}_1 \perp \bar{X}_2$, niin

$$\bar{X}_1 - \bar{X}_2 \sim N\left(0, \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 2/3

- Edellä esitetystä seuraa, että

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

- Koska standardipoikkeama σ on *tuntematon*, satunnaismuuttujan z lauseke on testisuureena *epäoperationaalinen*.
- Määritellään otosvariانسit

$$s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)^2, \quad k = 1, 2$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 3/3

- Jos satunnaismuuttujan z lausekkeessa standardipoikkeama σ korvataan *otossuureella*

$$s_P = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

niin saadaan *t-testisuure*

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

joka *nollahypoteesin H_0 pätiessä noudattaa Studentin t -jakaumaa* vapausastein $(n_1 + n_2 - 2)$:

$$t \sim t(n_1 + n_2 - 2)$$

- Todistus sivuutetaan; ks. kuitenkin vastaavaa todistusta *yhden otoksen t -testin* tapauksessa ja siellä esitettyjä viittauksia.

t -testisuure mittaa tilastollista etäisyyttä

- Testisuure

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

mittaa otoksien 1 ja 2 aritmeettisten keskiarvojen *tilastollista etäisyyttä*.

- *Mittayksikkönä* on erotuksen $\bar{X}_1 - \bar{X}_2$ standardipoikkeaman

$$\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

estimaattori, jota määrättäessä on oletettu, että nollahypoteesi H_0 pätee.

Kahden otoksen t -testi B

Testi

- Testisuureen

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

normaaliarvo = 0, koska *nollahypoteesin* $H_0 : \mu_1 = \mu_2 = \mu$ *pätiessä*

$$E(t) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen t arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.

Testin hylkäysalueen valinta sekä testin p -arvo

- Kahden otoksen t -testin B **hylkäysalueen** valinta tapahtuu kuten yhden otoksen t -testin tapauksessa paitsi, että t -testisuure noudattaa tässä Studentin t -jakaumaa vapausastein $(n_1 + n_2 - 2)$.
- Kahden otoksen t -testin B testisuureen arvoa vastaavan **p -arvon** määrittäminen tapahtuu kuten yhden otoksen t -testin tapauksessa paitsi, että t -testisuure noudattaa tässä Studentin t -jakaumaa vapausastein $(n_1 + n_2 - 2)$.

Normaalisuusoletuksen merkitys 1/2

- Kahden otoksen t -testin B yleisen hypoteesin mukaan havainnot ovat molemmissa otoksissa *normaalijakautuneita*.
- Testi *ei* kuitenkaan *ole herkkä poikkeamille normaalisuudesta*, jos molempien otosten otoskoot ovat ”*kyllin suuria*”.

Normaalisuusoletuksen merkitys 2/2

- *Testiä on melko turvallista käyttää, kun*

$$n_1 > 15 \text{ ja } n_2 > 15$$

ja n_1 ja n_2 eivät eroa toisistaan kovin paljon, *elleivät* havaintojen jakaumat *ole kovin vinoja* ja *ellei* havaintojen joukossa *ole poikkeavia havaintoja*.

- Jos

$$n_1 > 40 \text{ ja } n_2 > 40$$

testiä voidaan melko turvallisesti käyttää jopa selvästi vinoille havaintojen jakaumille.

Testit suhdeasteikollisille muuttujille

Testit normaalijakauman parametreille

Yhden otoksen t -testi

Kahden otoksen t -testi A

Kahden otoksen t -testi B

>> t -testi parivertailuille

Testi varianssille

Varianssien vertailutesti

***t*-testi parivertailuille**

Avainsanat

Aritmeettinen keskiarvo

Nollahypoteesi

Normaalijakauma

Odotusarvo

Otos

Otosvariassi

Parametri

Parivertailu

Testisuure

Testisuureen jakauma

Testisuureen normaaliarvo

***t*-jakauma**

Variassi

Yhden otoksen testit

Yleinen hypoteesi

Parivertailuasetelma

- **Parivertailuasetelma** syntyy tilastollisessa tutkimuksessa esimerkiksi seuraavissa tilanteissa:
 - (i) Päämääränä on *verrata kahta mittaria* mittaamalla molemmilla mittareilla samoja kohteita *samoissa olosuhteissa*.
 - (ii) Päämääränä on *tutkia jonkin käsittelyn vaikutusta* mittaamalla samoja kohteita *ennen* käsittelyä ja käsittelyn *jälkeen*.
 - (iii) Päämääränä on *vertailla kahta perusjoukkoa* mittaamalla saman muuttujan arvoja perusjoukkojen alkioden *sovitetuissa pareissa*.

Testausasetelma 1/2

- Oletetaan, että havainnot muodostuvat muuttujaa X koskevista mittaustuloksien *pareista*

$$(X_{i1}, X_{i2}), i = 1, 2, \dots, n$$

jotka ovat toisistaan *riippumattomia*.

- Päämääränä on *verrata mittauksia* toisiinsa:
Antavatko mittaukset *keskimäärin saman tuloksen?*
- *Tällaisissa parivertailuasetelmissa ei saa käyttää riippumattomien otoksien t-testejä A tai B, koska parivertailuasetelmissa mittaustulokset X_{i1} ja X_{i2} eivät ole riippumattomia.*

Testausasetelma 2/2

- Muodostetaan mittaustuloksien X_{i1} ja X_{i2} erotukset

$$D_i = X_{i1} - X_{i2}, i = 1, 2, \dots, n$$

- Mittaukset 1 ja 2 antavat keskimäärin saman tuloksen, jos erotukset D_i saavat keskimäärin arvon nolla.
- Testausongelman ratkaisuna on tavanomainen **yhden otoksen *t*-testi** mittaustuloksien X_{i1} ja X_{i2} erotuksien D_i odotusarvolle.

Hypoteesit

- *Yleinen hypoteesi* H :

(1) Erotukset $D_i \sim N(\mu_D, \sigma_D^2)$, $i = 1, 2, \dots, n$

(2) Erotukset D_1, D_2, \dots, D_n ovat riippumattomia

- *Nollahypoteesi* H_0 :

$$H_0 : \mu_D = 0$$

- *Vaihtoehtoinen hypoteesi* H_1 :

$$\left. \begin{array}{l} H_1 : \mu_D > 0 \\ H_1 : \mu_D < 0 \end{array} \right\} \text{1-suuntaiset vaihtoehtoiset hypoteesit}$$

$$H_1 : \mu_D \neq 0 \quad \text{2-suuntainen vaihtoehtoinen hypoteesi}$$

Parametrien estimointi

- Olkoot

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$$

ja

$$s_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$$

tavanomaiset *harhattomat estimaattorit* parametreille

$$E(D_i) = \mu_D, i = 1, 2, \dots, n$$

ja

$$\text{Var}(D_i) = \sigma_D^2, i = 1, 2, \dots, n$$

Testisuure ja sen jakauma

- Määritellään *t*-testisuure

$$t = \frac{\bar{D}}{s_D / \sqrt{n}}$$

- Jos nollahypoteesi

$$H_0 : \mu_D = 0$$

pätee, niin testisuure *t* noudattaa *Studentin t-jakaumaa* vapausastein ($n - 1$):

$$t \sim t(n - 1)$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 1/2

- Oletetaan, että testin yleinen hypoteesi H ja nollahypoteesi H_0 pätevät:

$$D_1, D_2, \dots, D_n \perp$$
$$D_i \sim N(0, \sigma_D^2), i = 1, 2, \dots, n$$

- Koska tällöin (ks. lukua **Otos ja otosjakaumat**)

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i \sim N\left(0, \frac{\sigma_D^2}{n}\right)$$

niin

$$z = \frac{\bar{D}}{\sigma_D / \sqrt{n}} \sim N(0, 1)$$

- Koska standardipoikkeama σ_D on *tuntematon*, satunnaismuuttujan z lauseke on testisuureena *epäoperationaalinen*.

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 2/2

- Jos satunnaismuuttujan z lausekkeessa standardipoikkeama σ_D korvataan vastaavalla *otossuureella*

$$s_D = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2}$$

niin saadaan *t*-testisuure

$$t = \frac{\bar{D}}{s_D / \sqrt{n}}$$

joka nollahypoteesin H_0 pätiessä noudattaa Studentin *t*-jakaumaa vapausastein $(n - 1)$:

$$t \sim t(n - 1)$$

- Todistus: ks. kappaletta **Yhden otoksen *t*-testi**.

t-testisuure mittaa tilastollista etäisyyttä

- Testisuure

$$t = \frac{\bar{D}}{s_D / \sqrt{n}}$$

mittaa havaintoarvojen erotuksien aritmeettisen keskiarvon *tilastollista etäisyyttä* nolasta.

- *Mittayksikkönä* on erotuksien D_i aritmeettisen keskiarvon \bar{D} standardipoikkeaman

$$\frac{\sigma_D}{\sqrt{n}}$$

estimaattori, jota määrättäessä on oletettu, että nollahypoteesi H_0 pätee.

t-testi parivertailuille

Testi

- Testisuureen

$$t = \frac{\bar{D}}{s_D / \sqrt{n}}$$

normaaliarvo = 0, koska *nollahypoteesin* $H_0 : \mu_D = 0$
pätiessä

$$E(t) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen *t* arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin *p*-arvo on *kyllin pieni*.

Testin hylkäysalueen valinta sekä testin *p*-arvo

- Parivertailutestin **hylkäysalueen** *valinta* tapahtuu kuten yhden otoksen *t*-testin tapauksessa.
- Parivertailutestin testisuureen arvoa vastaavan ***p*-arvon** *määrittäminen* tapahtuu kuten yhden otoksen *t*-testin tapauksessa.

Normaalisuusoletuksen merkitys 1/2

- Parivertailuasetelman *t*-testin yleisessä hypoteesissa oletetaan, että havaintoarvojen erotukset ovat *normaalijakautuneita*.
- Testi *ei* kuitenkaan *ole herkkä poikkeamille normaalisuudesta*, jos havaintojen lukumäärä *n* on ”*kyllin suuri*”.

Normaalisuusoletuksen merkitys 2/2

- *Testiä on melko turvallista käyttää, kun*

$$n > 15$$

ellei erotusten jakauma ole kovin vino ja erotuksien joukossa ole poikkeavia erotuksia.

- Jos havaintojen lukumäärä

$$n > 40$$

testiä voidaan melko turvallisesti käyttää jopa selvästi vinoille erotuksien jakaumille.

Testit suhdeasteikollisille muuttujille

Testit normaalijakauman parametreille

Yhden otoksen t -testi

Kahden otoksen t -testi A

Kahden otoksen t -testi B

t -testi parivertailuille

>> Testi varianssille

Varianssien vertailutesti

Testi varianssille

Avainsanat

Aritmeettinen keskiarvo

χ^2 -jakauma

Nollahypoteesi

Normaalijakauma

Odotusarvo

Otos

Otosvariassi

Parametri

Testisuure

Testisuureen jakauma

Testisuureen normaaliarvo

Variassi

Yhden otoksen testit

Yleinen hypoteesi

Testausasetelma 1/2

- Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos perusjoukosta S , joka noudattaa *normaalijakaumaa*

$$N(\mu, \sigma^2)$$

- Jakauma riippuu seuraavista parametreista:

$$\mu = \text{jakauman } \textit{odotusarvo}$$

$$\sigma^2 = \text{jakauman } \textit{varianssi}$$

Testausasetelma 2/2

- Asetetaan normaalijakauman $N(\mu, \sigma^2)$ varianssi-parametrille σ^2 nollahypoteesi

$$H_0 : \sigma^2 = \sigma_0^2$$

- Testausongelma:

Ovatko havainnot *sopusoinnussa* nollahypoteesin H_0 kanssa?

- Ongelman ratkaisuna on **yhden otoksen χ^2 -testi varianssille.**

Testi varianssille

Hypoteesit

- *Yleinen hypoteesi* H :

(1) Havainnot $X_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$

(2) Havainnot X_1, X_2, \dots, X_n ovat *riippumattomia*.

- *Nollahypoteesi* H_0 :

$$H_0 : \sigma^2 = \sigma_0^2$$

- *Vaihtoehtoinen hypoteesi* H_1 :

$$\left. \begin{array}{l} H_1 : \sigma^2 > \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{array} \right\} \text{1-suuntaiset vaihtoehtoiset hypoteesit}$$

$$H_1 : \sigma^2 \neq \sigma_0^2 \quad \text{2-suuntainen vaihtoehtoinen hypoteesi}$$

Parametrien estimointi

- Olkoot

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

ja

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

tavanomaiset *harhattomat estimaattorit* parametreille

$$E(X_i) = \mu, i = 1, 2, \dots, n$$

ja

$$\text{Var}(X_i) = \sigma^2, i = 1, 2, \dots, n$$

Testisuure ja sen jakauma

- Määritellään χ^2 -testisuure

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

- Jos nollahypoteesi

$$H_0 : \sigma^2 = \sigma_0^2$$

pätee, niin testisuure χ^2 noudattaa χ^2 -jakaumaa vapausastein $(n - 1)$:

$$\chi^2 \sim \chi^2(n-1)$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 1/3

- Oletetaan, että testin yleinen hypoteesi H ja nollahypoteesi H_0 pätevät:

$$X_1, X_2, \dots, X_n \perp$$

$$X_i \sim N(\mu, \sigma_0^2), i = 1, 2, \dots, n$$

- Tällöin (ks. lukua **Otos ja otosjakaumat**)

$$Y = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma_0} \right)^2 \sim \chi^2(n)$$

- Koska odotusarvo μ on *tuntematon*, satunnaismuuttujan Y lauseke on testisuurena *epäoperationaalinen*.

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 2/3

- Jos satunnaismuuttujan z lausekkeessa odotusarvo μ korvataan vastaavalla *otossuureella*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

niin saadaan χ^2 -testisuure

$$\chi^2 = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma_0} \right)^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

jossa

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Testi varianssille

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 3/3

- Jos *nollahypoteesi* H_0 pätee, testisuure χ^2 noudattaa χ^2 -jakaumaa vapausastein $(n - 1)$:

$$\chi^2 \sim \chi^2(n - 1)$$

- Todistus: ks. lukua **Otos ja otosjakaumat**.

Testi varianssille

Testi

- Testisuureen

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

normaaliarvo = $(n - 1)$, koska *nollahypoteesin*

$H_0 : \sigma^2 = \sigma_0^2$ pätiessä $E(s^2) = \sigma_0^2$, jolloin

$$E(\chi^2) = n - 1$$

- Siten sekä *pienet* että *suuret* testisuureen χ^2 arvot sen normaaliarvoon $(n - 1)$ nähden viittaavat siihen, että *nollahypoteesi ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin *p*-arvo on *kyllin pieni*.

Testin hylkäysalueen valinta 1/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_1 : \sigma^2 > \sigma_0^2$$

niin *kriittinen raja* χ_α^2 saadaan ehdosta

$$\Pr(\chi^2 \geq \chi_\alpha^2) = \alpha$$

jossa

$$\chi^2 \sim \chi^2(n - 1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(\chi_\alpha^2, +\infty)$$

Testin hylkäysalueen valinta 2/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_1 : \sigma^2 < \sigma_0^2$$

niin *kriittinen raja* $\chi_{1-\alpha}^2$ saadaan ehdosta

$$\Pr(\chi^2 \leq \chi_{1-\alpha}^2) = \alpha$$

jossa

$$\chi^2 \sim \chi^2(n - 1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(0, \chi_{1-\alpha}^2)$$

Testin hylkäysalueen valinta 3/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_1 : \sigma^2 \neq \sigma_0^2$$

niin *kriittiset rajat* $\chi_{1-\alpha/2}^2$ ja $\chi_{\alpha/2}^2$ saadaan ehdoista

$$\Pr(\chi^2 \leq \chi_{1-\alpha/2}^2) = \alpha/2$$

$$\Pr(\chi^2 \geq \chi_{\alpha/2}^2) = \alpha/2$$

jossa

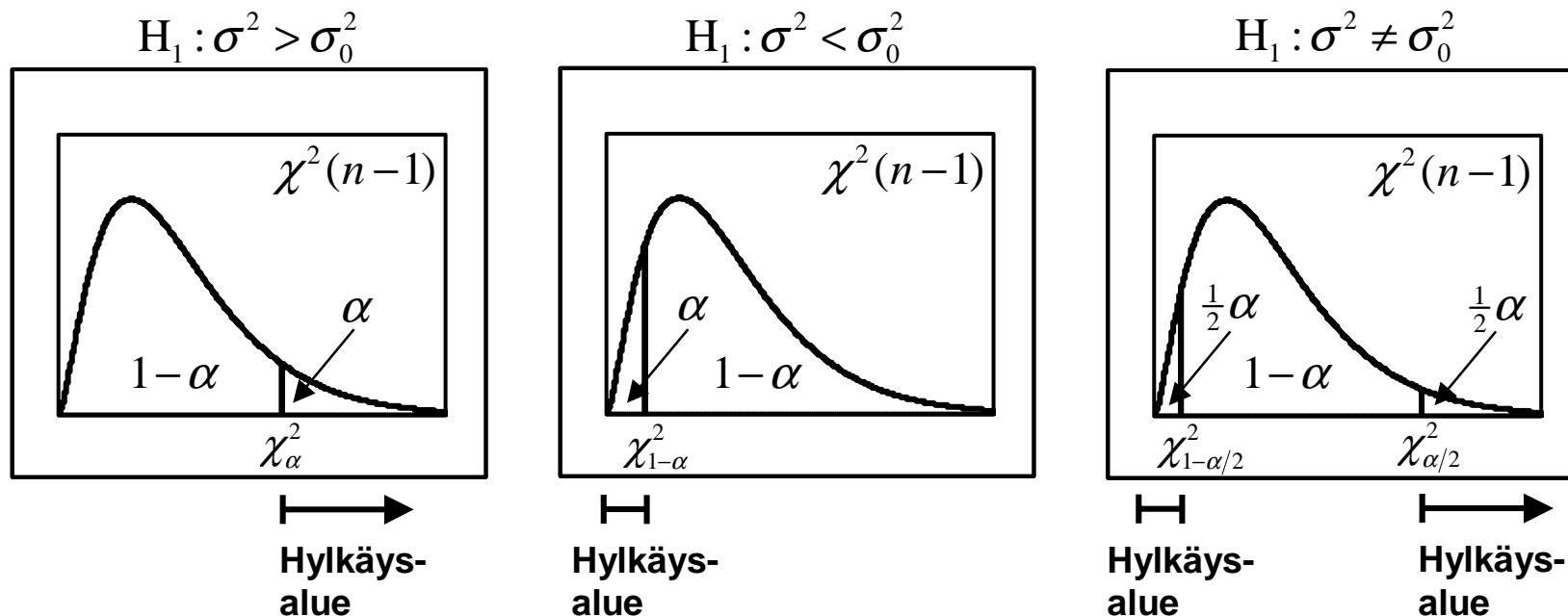
$$\chi^2 \sim \chi^2(n - 1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(0, \chi_{1-\alpha/2}^2) \cup (\chi_{\alpha/2}^2, +\infty)$$

Testin hylkäysalueen valinta 4/4

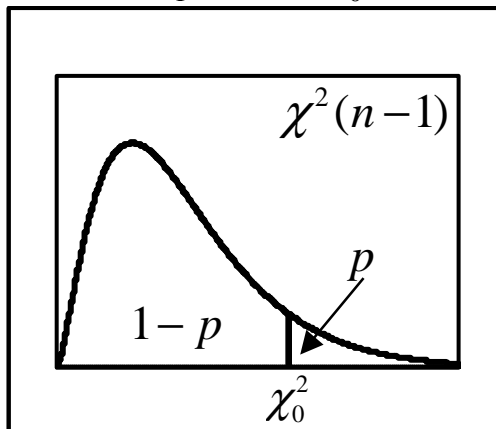
- Oletetaan, että testin *merkitsevyystasoksi* on valittu α .
- Testin **hylkäysalueen** määrittäminen voidaan havainnollistaa alla olevilla kuvioilla.



Testin p -arvo 1/2

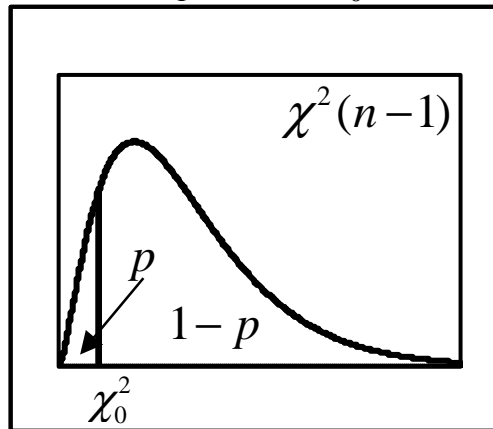
- Olkoon χ^2 -testisuureen havaittu arvo χ_0^2 .
- Jos vaihtoehtoinen hypoteesi on 1-suuntainen, testin p -arvon määrittäminen voidaan havainnollistaa alla olevilla kuvioilla.

$$H_1 : \sigma^2 > \sigma_0^2$$



Testin p -arvo = p

$$H_1 : \sigma^2 < \sigma_0^2$$



Testin p -arvo = p

Testin p -arvo 2/2

- Olkoon vaihtoehtoinen hypoteesi *2-suuntainen*:

$$H_1 : \sigma^2 \neq \sigma_0^2$$

- Tällöin testin **p -arvo** on

$$p = 2 \times \min \left\{ \Pr(\chi^2 \geq \chi_0^2), \Pr(\chi^2 \leq \chi_0^2) \right\}$$

jossa

$$\chi^2 \sim \chi^2(n-1)$$

Normaalisuusoletuksen merkitys

- Tässä esitetyn varianssitestin yleisessä hypoteesissa oletetaan, että havainnot ovat *normaalijakautuneita*.
- Testi *on herkkä poikkeamille normaalisuudesta ja testi ei toimi kovinkaan hyvin*, jos havaintojen jakauma *on vino* tai havaintojen joukossa *on poikkeavia havaintoja*.
- Tällöin suuretkaan havaintojen lukumäärät eivät yleensä paranna tilannetta.

Testit suhdeasteikollisille muuttujille

Testit normaalijakauman parametreille

Yhden otoksen t -testi

Kahden otoksen t -testi A

Kahden otoksen t -testi B

t -testi parivertailuille

Testi varianssille

>> Varianssien vertailutesti

Varianssien vertailutesti

Avainsanat

Aritmeettinen keskiarvo

F-jakauma

Kahden otoksen testit

Nollahypoteesi

Normaalijakauma

Odotusarvo

Otos

Otosvarianssi

Parametri

Testisuure

Testisuureen jakauma

Testisuureen normaaliarvo

Varianssi

Vertailutesti

Yleinen hypoteesi

Testausasetelma 1/4

- Olkoon

$$X_{11}, X_{21}, \dots, X_{n_11}$$

yksinkertainen satunnaisotos perusjoukosta S_1 , joka noudattaa *normaalijakaumaa*

$$N(\mu_1, \sigma_1^2)$$

- Jakauma riippuu seuraavista parametreista:

$$\mu_1 = \text{jakauman } \textit{odotusarvo}$$

$$\sigma_1^2 = \text{jakauman } \textit{varianssi}$$

Testausasetelma 2/4

- Olkoon

$$X_{12}, X_{22}, \dots, X_{n_2 2}$$

yksinkertainen satunnaisotos perusjoukosta S_2 , joka noudattaa *normaalijakaumaa*

$$N(\mu_2, \sigma_2^2)$$

- Jakauma riippuu seuraavista parametreista:

$$\mu_2 = \text{jakauman } \textit{odotusarvo}$$

$$\sigma_2^2 = \text{jakauman } \textit{varianssi}$$

Testausasetelma 3/4

- Oletetaan lisäksi, että perusjoukosta S_1 poimittu otos

$$X_{11}, X_{21}, \dots, X_{n_11}$$

ja perusjoukosta S_2 poimittu otos

$$X_{12}, X_{22}, \dots, X_{n_22}$$

ovat toisistaan *riippumattomia*.

- Otosten riippumattomuus merkitsee sitä, että se mikä alkio poimitaan perusjoukosta S_1 *ei vaikuta* siihen mikä alkioista poimitaan perusjoukosta S_2 ja kääntäen.

Testausasetelma 4/4

- Asetetaan normaalijakaumien $N(\mu_1, \sigma_1^2)$ ja $N(\mu_2, \sigma_2^2)$ varianssiparametreille σ_1^2 ja σ_2^2 nollahypoteesi

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$$

- Testausongelma:

Ovatko havainnot *sopusoinnussa* nollahypoteesin H_0 kanssa?

- Ongelman ratkaisuna on **kahden riippumattoman otoksen F -testi variansseille.**

Yleinen hypoteesi

- *Yleinen hypoteesi* H :

(1) Havainnot $X_{i1} \sim N(\mu_1, \sigma_1^2)$, $i = 1, 2, \dots, n_1$

(2) Havainnot $X_{j2} \sim N(\mu_2, \sigma_2^2)$, $j = 1, 2, \dots, n_2$

(3) Havainnot X_{i1} ja X_{j2} ovat *riippumattomia* kaikille i ja j

- Huomautus:

Oletus (3) sisältää *kolme riippumattomuusoletusta*:

- Havainnot ovat riippumattomia otoksien 1 ja 2 *sisällä*.
- Havainnot ovat riippumattomia otoksien 1 ja 2 *välillä*.

Nollahypoteesi ja vaihtoehtoiset hypoteesit

- *Nollahypoteesi* H_0 :

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$$

- *Vaihtoehtoinen hypoteesi* H_1 :

$$\left. \begin{array}{l} H_1 : \sigma_1^2 > \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{array} \right\} \text{1-suuntaiset vaihtoehtoiset hypoteesit}$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad \text{2-suuntainen vaihtoehtoinen hypoteesi}$$

Varianssien vertailutesti

Parametrien estimointi

- Olkoot

$$\bar{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ik}, \quad k = 1, 2$$

ja

$$s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)^2, \quad k = 1, 2$$

tavanomaiset *harhattomat estimaattorit* parametreille

$$E(X_{ik}) = \mu_k, \quad i = 1, 2, \dots, n_k, \quad k = 1, 2$$

ja

$$\text{Var}(X_{ik}) = \sigma_k^2, \quad i = 1, 2, \dots, n_k, \quad k = 1, 2$$

Testisuure ja sen jakauma

- Määritellään ***F*-testisuure**

$$F = \frac{s_1^2}{s_2^2}$$

- *Jos nollahypoteesi*

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$$

pätee, niin testisuure *F* noudattaa *Fisherin F-jakaumaa* vapausastein $(n_1 - 1)$ ja $(n_2 - 1)$:

$$F \sim F(n_1 - 1, n_2 - 1)$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 1/4

- Oletetaan, että testin yleinen hypoteesi H ja nollahypoteesi H_0 pätevät:

$$X_{11}, X_{21}, \dots, X_{n_1 1}, X_{12}, X_{22}, \dots, X_{n_2 2} \perp$$

$$X_{i1} \sim N(\mu_1, \sigma^2), i = 1, 2, \dots, n_1$$

$$X_{j2} \sim N(\mu_2, \sigma^2), j = 1, 2, \dots, n_2$$

- Tällöin (ks. lukua **Otos ja otosjakaumat**)

$$Y_1 = \sum_{i=1}^{n_1} \left(\frac{X_{i1} - \mu_1}{\sigma} \right)^2 \sim \chi^2(n_1)$$

$$Y_2 = \sum_{j=1}^{n_2} \left(\frac{X_{j2} - \mu_2}{\sigma} \right)^2 \sim \chi^2(n_2)$$

- Koska $Y_1 \perp Y_2$, niin

$$Y = \frac{Y_1/n_1}{Y_2/n_2} \sim F(n_1, n_2)$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 2/4

- Koska odotusarvot μ_1 ja μ_2 ovat *tuntemattomia*, satunnaismuuttujan

$$Y = \frac{Y_1 / n_1}{Y_2 / n_2}$$

lauseke on testisuurena *epäoperationaalinen*.

- Korvataan satunnaismuuttujien Y_1 ja Y_2 lausekkeissa odotusarvot μ_1 ja μ_2 vastaavilla *otossuureilla*

$$\bar{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ik}, \quad k = 1, 2$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 3/4

- Saamme satunnaismuuttujat (ks. lukua **Otos ja otosjakaumat**)

$$V_1 = \frac{(n_1 - 1)s_1^2}{\sigma^2} = \sum_{i=1}^{n_1} \left(\frac{X_{i1} - \bar{X}_1}{\sigma} \right)^2 \sim \chi^2(n_1 - 1)$$

$$V_2 = \frac{(n_2 - 1)s_2^2}{\sigma^2} = \sum_{j=1}^{n_2} \left(\frac{X_{j2} - \bar{X}_2}{\sigma} \right)^2 \sim \chi^2(n_2 - 1)$$

jossa

$$s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)^2, \quad k = 1, 2$$

- Lisäksi satunnaismuuttujat V_1 ja V_2 ovat *riippumattomia*:

$$V_1 \perp V_2$$

Testisuureen jakauma nollahypoteesin H_0 pätiessä: Perustelu 4/4

- Määritellään F -testisuure

$$F = \frac{V_1 / (n_1 - 1)}{V_2 / (n_2 - 1)} = \frac{s_1^2}{s_2^2}$$

- Jos nollahypoteesi H_0 pätee, testisuure F noudattaa Fisherin F -jakaumaa vapausastein $(n_1 - 1)$ ja $(n_2 - 1)$:

$$F \sim F(n_1 - 1, n_2 - 1)$$

Varianssien vertailutesti

Testi

- Testisuureen

$$F = \frac{s_1^2}{s_2^2}$$

normaaliarvo ≈ 1 , koska *nollahypoteesin* $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ *pätiessä* (ja jos n_2 on kyllin suuri)

$$E(F) = \frac{n_2 - 1}{n_2 - 3} \approx 1$$

- Siten sekä *pienet* että *suuret* testisuureen F arvot sen normaaliarvoon ≈ 1 nähden viittaavat siihen, että *nollahypoteesi ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.

Testin hylkäysalueen valinta 1/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_0 : \sigma_1^2 > \sigma_2^2$$

niin *kriittinen raja* F_α saadaan ehdosta

$$\Pr(F \geq F_\alpha) = \alpha$$

jossa

$$F \sim F(n_1 - 1, n_2 - 1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(F_\alpha, +\infty)$$

Testin hylkäysalueen valinta 2/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_0 : \sigma_1^2 < \sigma_2^2$$

niin *kriittinen raja* $F_{1-\alpha}$ saadaan ehdosta

$$\Pr(F \leq F_{1-\alpha}) = \alpha$$

jossa

$$F \sim F(n_1 - 1, n_2 - 1)$$

- Testin **hylkäysalue** on tällöin muotoa
 $(0, F_{1-\alpha})$

Testin hylkäysalueen valinta 3/4

- Valitaan testin **merkitsevyystasoksi** α .
- Jos *vaihtoehtoinen hypoteesi* on muotoa

$$H_0 : \sigma_1^2 \neq \sigma_2^2$$

niin *kriittiset rajat* $F_{1-\alpha/2}$ ja $F_{\alpha/2}$ saadaan ehdoista

$$\Pr(F \leq F_{1-\alpha/2}) = \alpha/2$$

$$\Pr(F \geq F_{\alpha/2}) = \alpha/2$$

jossa

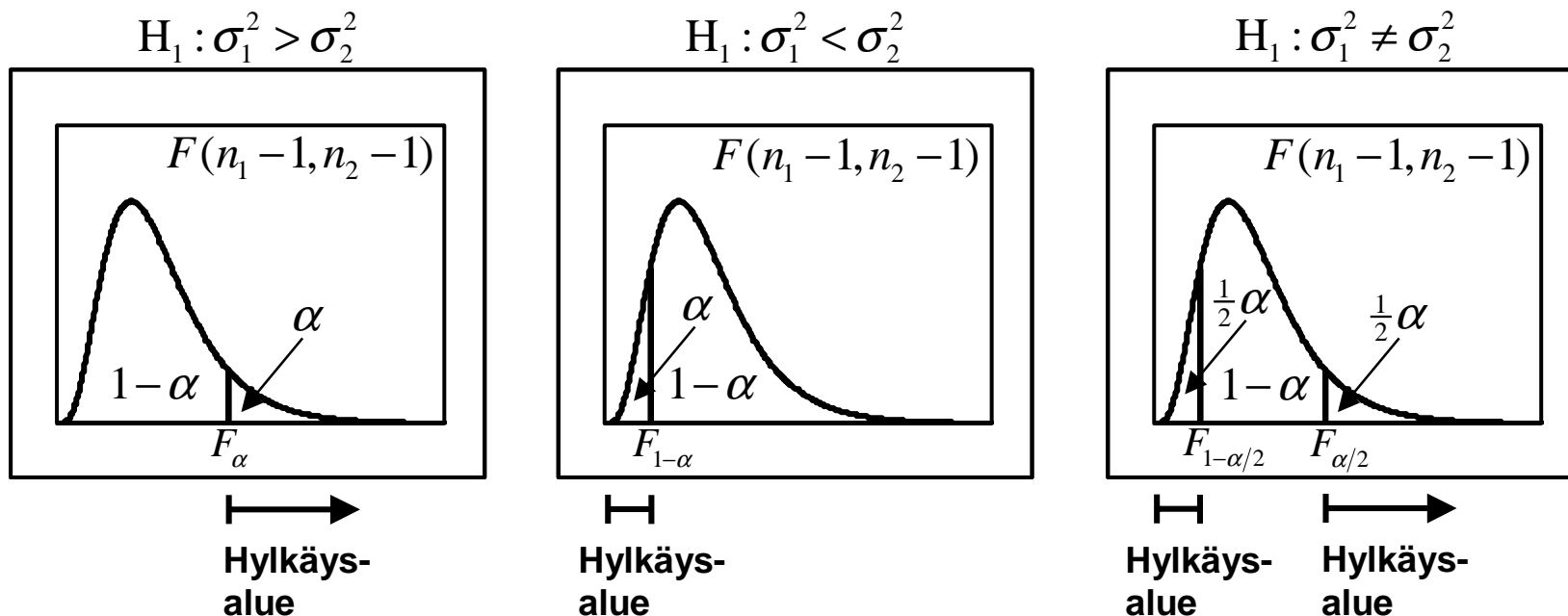
$$F \sim F(n_1 - 1, n_2 - 1)$$

- Testin **hylkäysalue** on tällöin muotoa

$$(0, F_{1-\alpha/2}) \cup (F_{\alpha/2}, +\infty)$$

Testin hylkäysalueen valinta 4/4

- Oletetaan, että testin *merkitsevyystasoksi* on valittu α .
- Testin **hylkäysalueen** määrittäminen voidaan havainnollistaa olevilla kuvioilla.

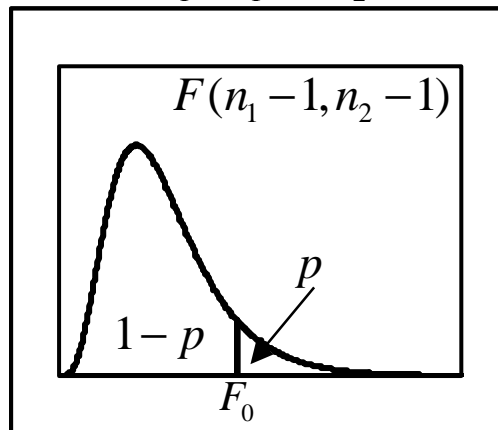


Varianssien vertailutesti

Testin p -arvo 1/2

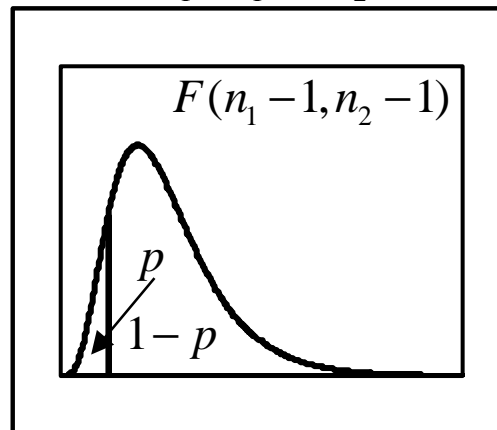
- Olkoon F -testisuureen havaittu arvo F_0 .
- Jos vaihtoehtoinen hypoteesi on 1-suuntainen, testin p -arvon määrittämistä voidaan havainnollistaa alla olevilla kuvioilla.

$$H_1 : \sigma_1^2 > \sigma_2^2$$



Testin p -arvo = p

$$H_1 : \sigma_1^2 < \sigma_2^2$$



Testin p -arvo = p

Varianssien vertailutesti

Testin p -arvo 2/2

- Olkoon vaihtoehtoinen hypoteesi *2-suuntainen*:

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

- Tällöin testin **p -arvo** on

$$p = 2 \times \min \{ \Pr(F \geq F_0), \Pr(F \leq F_0) \}$$

jossa

$$F \sim F(n_1 - 1, n_2 - 1)$$

Normaalisuusoletuksen merkitys

- Tässä esitetyn varianssien vertailutestin yleisessä hypoteesissa oletetaan, että havainnot ovat molemmissa otoksissa *normaalijakautuneita*.
- Testi *on herkkä poikkeamille normaalisuudesta ja testi ei toimi kovinkaan hyvin, jos havaintojen jakauma on vino tai havaintojen joukossa on poikkeavia havaintoja*.
- Tällöin suurelleen havaintojen lukumäärät eivät yleensä paranna tilannetta.