Turun yliopisto
University of Turku

TKMY3 TILASTOLLISEN PÄÄTTELYN PERUSTEET

Myös Porin PTKMY3 2015

KTT, VTM Janne Engblom

2015

KURSSIN SUORITTAMINEN

- ~~Riittävä osallistuminen harjoituksiin (25 kpl harjoitusmerkintöjä, joita saa 1.harjoituksissa osallistumisesta 2 kpl ja muissa harjoituksissa osallistumisesta ja kotitehtävien teosta) antaa tenttioikeuden.~~

- ~~Harjoitustunnille osallistuminen sisältyy harjoitusmerkintään.~~

- Tarkempi kurssi-info ja kurssimateriaali löytyy Moodlesta.

KURSSIN RAKENNE

1. ~~Johdanto ja todennäköisyyslaskennan perusteet (vkot 1-2)~~
2. ~~Satunnaismuuttujat ja niiden jakaumia (vko 2-3)~~
3. ~~Estimointi (vko 3)~~
4. ~~Tilastollisia testejä (vkot 4-7).~~

JOHDANTO: MITÄ ON TILASTOTIEDE?

- **Tilastotieteeseen kuuluvat reaalimaailman ilmiöitä koskevan, havaintoihin perustuvan numeerisen tiedon kerääminen ja tietojen hankinnan suunnittelu.**
- **Tilastotiedettä ovat myös näiden tietojen käsittely, analysointi, esittäminen ja tulkinta. Tilastotieteen keskeisin tavoite kuitenkin on sellaisten käsitteiden ja menetelmien kehittäminen, joita voidaan käyttää edellä mainituissa tehtävissä.**
- **Tilastotiede on luonteeltaan menetelmätiede, mikä tarkoittaa, että se tuottaa välineistöä soveltavien tieteiden tarpeisiin eli mm. ekonometrian, logistiikan, laadunhallinnan, rahoituksen ja markkinatutkimuksen sovelluksia.**

MITÄ ON TILASTOLLINEN PÄÄTTELY?

- **Kuvaileva tilastotiede tarkoittaa tilastollisten aineistojen yleispiirteiden kuvaamista ilman pyrkimystä yleistysten tekemiseen.**
- **Toinen tilastotieteen osa edustaa ns. tilastollista päättelyä, jossa pienempien aineistojen, otosten, avulla tehdään yleistyksiä suurempiin joukkoihin, perusjoukkoihin.**

TODENNÄKÖISYYSLASKENNAN PERUSTEET

- **Todennäköisyyslaskennan katsotaan saaneen alkunsa 1600-luvulla uhkapeleihin liittyvien ongelmien tutkimisesta.**
- **Todennäköisyydellä käsitteenä on erilaisia tulkintoja. Myös itse todennäköisyyslaskentaan on tapauksesta riippuen mahdollista ottaa erilaisia lähestymistapoja (ns. klassinen tai aksiomaattinen lähestymistapa). Ensin mainitussa todennäköisyyksiä määritellään lukumäärien avulla. Jälkimmäisessä todennäköisyyksiä määritellään muiden todennäköisyyksien avulla käyttäen laskentasääntöjä.**

TODENNÄKÖISYYSLASKENNAN PERUSTEITA

- Todennäköisyyslaskenta tarkastelee *satunnaisilmiöitä*.
- Tarkasteltavan satunnaisilmiön mahdollisia tuloksia sanotaan *alkeistapahtumiksi*.
- Kaikki alkeistapahtumat muodostavat perusjoukon.
- *Tapahtumiksi* sanotaan perusjoukon osajoukkoja.
- Tapahtumaan kuuluvia alkeistapahtumia sanotaan tapahtumalle *suotuisiksi alkeistapahtumiksi*.
- Koska tapahtumat esitetään joukkoina, eri tapahtumien yhdistelmiä voidaan kuvata joukko-opin operaatioiden avulla.

JOUKKO-OPIN MERKINTÖJÄ JA LYHENTEITÄ

E	perusjoukko, varma tapahtuma
$e_1, e_2, \ldots$	alkeistapahtumat
$A, B, C, \ldots$	tapahtumat eli alkeistapahtumien joukot
$\varnothing$	Mahdoton tapahtuma
$A \cup B$	A tai B (tai molemmat) tapahtuu (=yhdiste)
$A \cap B$	A ja B tapahtuu (=leikkaus)
A^*	A ei tapahdu (=komplementti)

Tapahtumia voidaan kuvata graafisesti ns. venn-diagrammin avulla.

TODENNÄKÖISYYDEN TULKINNOISTA

- **Frekvenssitulkinta: ”Se mitä tavallisesti tapahtuu”**

$$P(A) = \lim_{n\to\infty} P_n(A) = \lim_{n\to\infty} \frac{f_n}{n}$$

- **Subjektiivinen tulkinta: ”Uskomuksen aste”.**

TODENNÄKÖISYYSLASKENTA

- Todennäköisyyttä koskevan matemaattisen teorian ensi vaihe oli *klassiseksi* kutsuttu lähestymistapa.
- Siinä jokaisen alkeistapahtuman todennäköisyys oletetaan yhtä suureksi eli todennäköisyysavaruus on symmetrinen.
- Edellisestä johtuen tapahtumien todennäköisyydet voidaan laskea suotuisten ja kaikkien alkeistapahtumien lukumäärien avulla käyttäen apuna kombinatoriikkaa.
- Nykyinen matemaattinen todennäköisyysteoria perustuu Kolmogorovin 1930-luvulla esittämiin *aksioomiin*. Siinä ei oleteta todennäköisyysavaruuden symmetriaa.

KLASSINEN LÄHESTYMISTAPA TODENNÄKÖISYYSLASKENTAAN

- **Edellytys: Kaikki alkeistapahtumat yhtä todennäköisiä eli symmetrisiä.**
- **Tällöin**

$$P(A) = \frac{k}{n}$$

jossa *k* on suotuisten ja n kaikkien alkeistapahtumien lukumäärä.

- **Apuna lukumäärien laskennassa käytetään kombinatoriikan sovelluksia.**

KLASSINEN LÄHESTYMISTAPA TODENNÄKÖISYYSLASKENTAAN: ESIMERKKEJÄ

Heitetään kahta noppaa. Määritä seuraavien tapahtumien todennäköisyydet:

1. Silmälukujen summa on parillinen.
2. Silmäluku summa on vähintään 3.
3. Silmälukujen summa ei ole 6.
4. Molemmat silmäluvut ovat 6.

KOMBINATORIIKKAA

- **Alkeistapahtumat eivät koostu välttämättä yksittäisistä luvuista tms. vaan joukoista.**
- **Permutaatioiden eli erilaisten jonojen lukumäärä:** $n! = n \cdot (n-1) \cdot (n-2) \cdots 1$
- **Em. yleisemmin: Ns. tuloperiaate**
- **Kombinaatioiden eli k:n suuruisten järjestämättömien osajoukkojen lukumäärä:**
  $$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$
- **Variaatioiden eli k:n suuruisten järjestettyjen osajoukkojen lukumäärä:**
  $$\frac{n!}{(n-k)!}$$

KOMBINATORIIKKAA: ESIMERKKEJÄ

1. **Kuinka monella eri tavalla voidaan ravintolassa valita kolmen ruokalajin yhdistelmä, kun alku-pää- ja jälkiruokavaihtoehtoja on 5, 8 ja 4 kpl?**
2. **Kuinka monella eri tavalla voidaan valita laulukilpailujen voittaja, 2.sijalle sekä 3. sijalle tulleet, kun kilpailussa on 10 osallistujaa?**
3. **Kuinka monella eri tavalla voidaan 15 ihmisestä poimia 6 pelaajan lentopallojoukkue?**
4. **Yrityksen varastossa on 100 tuotetta, joista 20 on uutta mallia. Valitaan varastosta satunnaisesti 5 tuotetta. Millä todennäköisyydellä kaikki ovat uutta mallia?**

TODENNÄKÖISYYSLASKENNAN LASKUSÄÄNNÖT ELI AKSIOMAT

- **Eivät edellytä todennäköisyydeltään symmetrisiä alkeistapahtumia. Tällöin laskenta voidaan laskenta suorittaa seuraavien laskusääntöjen avulla.**

Laskusäännöt eli aksioomat edellyttävät seuraavaa

1. $0 \leq P(A) \leq 1$ kaikille tapahtumille
2. $P(E) = 1$
3. $P(A \cup B) = P(A) + P(B)$
toisensa poissulkeville tapahtumille
(Eli joille $P(A \cap B) = 0$).

TODENNÄKÖISYYSLASKENNAN LASKUSÄÄNNÖT

Aksiomista johdettuja laskusääntöjä:

- **Mahdottoman tapahtuman todennäköisyys**
- **Komplementtitapahtuman todennäköisyys**
- **Toisensa poissulkevien tapahtumien yhteenlaskusääntö**
- **Yleinen yhteenlaskusääntö**
- **Kertolaskusäännöt.**

MAHDOTTOMAN TAPAHTUMAN TODENNÄKÖISYYS

$$P(\emptyset) = 0$$

KOMPLEMENTTITAPAHTUMAN TODENNÄKÖISYYS

A***="**A **ei tapahdu"**

$$P(A^*) = 1 - P(A)$$

TOISENSA POISSULKEVIEN TAPAHTUMIEN YHTEENLASKUSÄÄNTÖ

$A \cup B =$ **”A tai B” tapahtuu**

$$P(A \cup B) = P(A) + P(B)$$

Yleisemmin: ”Yksi tapahtumista”

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

YLEINEN YHTEENLASKUSÄÄNTÖ

Tapahtumat eivät toisensa poissulkevia:

”A tai B tai molemmat tapahtuu”

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

KERTOLASKUSAANTÖ

- $A \cap B =$ "*A ja B*" tapahtuu. Jos A ja B ovat toisistaan riippumattomia:

$$P(A \cap B) = P(A) \cdot P(B)$$

- Yleisemmin (useampi tapahtuma):

$$P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

- Edelleen $P(A \cap B) = P(A) \cdot P(B \mid A)$

  , jos A ja B eivät ole toisistaan riippumattomia. Jälkimmäinen merkintä viittaa ns. ehdolliseen todennäköisyyteen, jossa ilmaistaan B:n todennäköisyys A:n tapahtuessa eli "ehdolla A".

TODENNÄKÖISYYSLASKENNAN LASKUSÄÄNNÖT: ESIMERKKEJÄ

1. **Määritä klassisen todennäköisyyslaskennan esimerkkien (*2* noppaa) tulokset käyttäen em. laskusääntöjä.**

TOISTOKOKEET

- Toistokoe tarkoittaa saman satunnaiskokeen toistamista tai satunnaisilmiön havainnoimista samoissa olosuhteissa useita kertoja. Tarkasteltavat tapahtumat ovat tällöin yleensä eri satunnaisilmiöihin liittyviä.
- Em. tilanteissa voidaan tarkastella esim. jonkin tapahtuman esiintymisen lukumääriin liittyviä todennäköisyyksiä.
- Toistokokeisiin liittyivät myös ne klassisen todennäköisyyslaskennan sovellukset, joissa alkeistapahtumat olivat joukkoja.
- Olennaista on pohtia, ovatko toistot toisistaan riippumattomia vai riippuvia.

SATUNNAISMUUTTUJA

- **Satunnaismuuttujaa rakennettaessa alkeistapahtumiin liitetään lukuarvot ja sen jälkeen määritetään satunnaismuuttujan arvoihin liittyviä todennäköisyyksiä.**
- **Samaan satunnaisilmiöön/ilmiöihin voidaan liittää useita erilaisia satunnaismuuttujia.**

ESIMERKKEJÄ SATUNNAISMUUTTUJISTA

Satunnaisilmiö	Satunnaismuuttuja	Satunnaismuuttujan mahdolliset arvot
Tarkastetaan 100:n tuotteen erä	Viallisten tuotteiden lukumäärä	0,1,…,100
Palvelutapahtumat pankissa	Palveluaika	[0, a] a= aukioloaika
Yrityksen arvon vaihtelu pörssissä	Osakekurssi	[0, ∞]
Tutkitaan tavara-talon asiakkaita	Asiakkaiden lukumäärä	0,1,2,3,…

SATUNNAISMUUTTUJAN TODENNÄKÖISYYSJAKAUMAN ESITYSTAPOJA

- Seuraavassa käsitellään ns. diskreettejä satunnaismuuttujia (myöhemmin käsitellään myös ns. jatkuvia satunnaismuuttujia).
- Diskreetin todennäköisyysjakauman funktiomuotoisia esitystapoja: Pistetodennäköisyys- ja kertymäfunktiot.
- Em. funktioita voidaan kuvata myös graafisesti.
- Kuvailevia tunnuslukuja: Odotusarvo ja varianssi/keskihajonta.

DISKREETIN SATUNNAISMUUTTUJAN PISTETODENNÄKÖISYYSFUNKTIO

$$P(x) = \begin{cases} P(X = x_1), kun\ x = x_1 \\ .. \\ P(X = x_K), kun\ x = x_K \\ \\ \qquad 0, muulloin \end{cases}$$

, jossa jono $x_1, x_2, \ldots, x_K$ viittaa muuttujan mahdollisiin arvoihin. Funktio kertoo todennäköisyyden, että satunnaismuuttujan X arvo on x. Käytännössä rivejä funktioon tule $K+1$ kappaletta, jos satunnaismuuttujalla on K erilaista arvoa.

Pistetodennäköisyysfunktion arvojen summa laskettuna kaikkien mahdollisten X:n arvojen yli on 1.

DISKREETIN SATUNNAISMUUTTUJAN KERTYMÄFUNKTIO

Kertoo todennäköisyyden, että satunnaismuuttujan X arvo on korkeintaan x:

$$F(x) = P(X \leq x)$$

Kertymäfunktiolla on seuraavat ominaisuudet:

1. $P(X \leq a) = F(a)$
2. $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$
3. $P(a < X \leq b) = F(b) - F(a)$.

Kertymäfunktion arvot määritellään kaikille reaaliluvuille.

TODENNÄKÖISYYSJAKAUMAN GRAAFINEN ESITTÄMINEN

- **Todennäköisyysjakaumia voidaan havainnollistaa periaatteessa samoin kuin empiirisiä jakaumia.**
- **Esim. jana- tai pylväsdiagrammit havainnollistavat pistetodennäköisyysfunktiota.**
- **Summakäyrällä voidaan havainnollistaa kertymäfunktiota.**

DISKREETIN SATUNNAISMUUTTUJAN TUNNUSLUKUJA

Muuttujan X odotusarvo EX sekä varianssi / keskihajonta D^2X/DX (kaikkien mahdollisten arvojen yli),

$$EX = \sum P(x) \cdot x$$

$$D^2X = \sum P(x) \cdot (x - EX)^2$$

$$DX = \sqrt{D^2X}$$

DISKREETIN SATUNNAISMUUTTUJAN KONSTRUOINTI JA TUNNUSLUVUT: ESIMERKKEJÄ

1. **Määritä kahden nopan silmäluvun maksimin (*M*=”Paras tulos kahdesta”) pistetodennäköisyys- ja kertymäfunktiot.**
2. **Määritä satunnaismuuttujan *M* odotusarvo ja keskihajonta.**

SATUNNAISMUUTTUJIEN MUUNNOKSISTA

- **Jos X on satunnaismuuttuja ja g on jokin funktio reaalilukujen joukolta reaalilukujen joukolle, on**

  $Y = g(X)$ myös satunnaismuuttuja. Merkintä tarkoittaa, että Y saa arvon $y = g(x)$ silloin kun X saa arvon x.
- **Eräs muunnoksista on lineaarinen muunnos**

  $Y = a + bX$, missä a ja b ovat reaalilukuvakioita. Muunnoksien tunnuslukujen laskemisessa voidaan käyttää apuna mm. seuraavia yleisiä laskusääntöjä:

  1) $E(a) = a$ ***ja*** $D^2(a) = 0$

  2) $EY = E(a + bX) = a + bEX$ **ja**

  $D^2Y = D^2(a + bX) = b^2D^2X$

  3) $E(X+Y) = EX + EY$ **ja kun X ja Y ovat riippumattomia** $D^2(X+Y) = D^2X + D^2Y$

JOTAKIN DISKREETTEJÄ TODENNÄKÖISYYSJAKAUMIA

- **Binomijakauma: Satunnaismuuttuja, jossa määritetään jonkin tapahtuman esiintymisten lukumäärä n toistossa. Tapahtuman todennäköisyys yksittäisessä toistossa pysyy samana.**
- **Hypergeometrinen jakauma: Ks. em. paitsi tapahtuman todennäköisyys yksittäisessä toistossa ei pysy samana.**
- **Poisson-jakauma: Harvinaiset tapahtumat.**
- **Näitä jakaumia määrittävät tarkemmin ns. parametrit. Näitä ovat esim. binomijakaumassa n ja p.**

DISKREETIT SATUNNAISMUUTTUJAT: BINOMIJAKAUMA

- **Pistetodennäköisyysfunktio:**

$$P(X=k)=\binom{n}{k}\cdot p^{k}\cdot(1-p)^{n-k}, k=0,\ldots,n$$

- **Tällöin sanotaan, että muuttujan X jakauma on binomijakauma parametreina n ja p ja merkitään**

$X \sim Bin(n,p)$

- **Binomijakautuneen satunnaismuuttujan kertymäfunktio on**

$$F(k)=P(X\le k)=\sum_{i=0}^{k}\binom{n}{i}\cdot p^{i}\cdot(1-p)^{n-i}$$

- **Sen odotusarvo on $EX = n \cdot p$ ja varianssi**

$D^2X = n\cdot p\cdot(1-p)$

BINOMIJAKAUMA: ESIMERKKEJÄ

1. **Todennäköisyys tentin läpäisyyn on 0.6. Mikä on todennäköisyys, että osallistuttaessa neljään tenttiin vähintään kaksi menee läpi?**
2. **Millä todennäköisyydellä saadaan 10 rahan heitossa vähintään 3 kertaa tulos ”klaava”?**

DISKREETIT SATUNNAISMUUTTUJAT: HYPERGEOMETRINEN JAKAUMA

- **Pistetodennäköisyysfunktio:**

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, k = 0, \ldots, n$$

- **Tällöin sanotaan, että muuttujan X jakauma on hypergeometrinen jakauma parametreina N, K ja n ja merkitään $X \sim Hyperg(N,K,n)$.**
- ***X*:n odotusarvo on tällöin $EX = \frac{n \cdot K}{N}$ ja varianssi**

$$D^2X = \frac{n \cdot K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}$$

HYPERGEOMETRINEN JAKAUMA: ESIMERKKEJÄ

1. **Mikä on todennäköisyys saada lotossa (39 numeroa, joista 7 arvotaan) 7 oikein?**
2. **Mikä on todennäköisyys saada em. Lotossa 4 oikein?**
3. **Laatikossa on voittavia arpoja on 10 kpl 30:sta. Laatikosta arvotaan 5 arpaa ilman takaisinpanoa. Mikä on todennäköisyys, että arvottujen arpojen joukossa on vähintään kolme voittoarpaa?**

JATKUVAT SATUNNAISMUUTTUJAT

- **Jatkuva satunnaismuuttuja voi saada joltain tietyltä reaalilukuväliltä periaatteessa minkä hyvänsä arvon.**
- **Jatkuvan satunnaismuuttujan todennäköisyysjakauman antaa jatkuva tai ainakin paloittain jatkuva funktio, jota sanotaan *tiheysfunktioksi (f)*, jolle pätee mm.**

$$P(a \le X \le b) = \int_a^b f(x)dx$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- **Jatkuvan satunnaismuuttujan kertymäfunktio F määritellään samoin kuin diskreetin muuttujan tapauksessa, ts.**

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx$$

JATKUVAT SATUNNAISMUUTTUJAT

- **Koska X:n yksittäisen arvon todennäköisyys on nolla, on myös esim.**

  $F(b) - F(a) = P(a < X \leq b) = P(a \leq X < b) =$

  $P(a < X < b) = P(a \leq X \leq b).$

- **Jatkuvan satunnaismuuttujan X odotusarvo ja varianssi ovat**

$$EX = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

$$D^2X = \int_{-\infty}^{\infty} (x - EX)^2 \cdot f(x)dx$$

JATKUVAT SATUNNAISMUUTTUJAT: NORMAALIJAKAUMA

- **Normaalijakaumaa kutsutaan toisen kehittäjänsä mukaan myös Gaussin jakaumaksi.**
- **Normaalijakauma soveltuu monien empiiristen ilmiöiden kuvaamiseen ja on jakaumista ehkä käytetyin.**
- **Normaalijakaumaa voidaan lisäksi käyttää diskreettien muuttujien jakaumien approksimointiin.**
- **Käytetään tilastollisessa päättelyssä mm. keskiarvoon liittyen.**

JATKUVAT SATUNNAISMUUTTUJAT: NORMAALIJAKAUMA

- **Olkoon X satunnaismuuttuja, joka voi saada kaikki reaalilukuarvot. Sanotaan, että X on normaalisti jakautunut parametrein μ ja σ^2 , merkitään $X \sim N(\mu, \sigma^2)$, jos sen tiheysfunktio on muotoa**

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- **Normaalisti jakautuneen satunnaismuuttujan odotusarvo ja varianssi ovat $EX = \mu$ *ja* $D^2X = \sigma^2$. Nämä määräävät jakauman sijainnin ja muodon.**

JATKUVAT SATUNNAISMUUTTUJAT: NORMAALIJAKAUMA

- **Normaalijakauman tiheysfunktion kuvaajan huippu sijaitsee keskiarvon μ kohdalla. Hajonta σ määrää, kuinka leveälle satunnaismuuttujan arvot ovat hajonneet. Mitä suurempi σ, sitä litteämpi ja leveämpi on tiheysfunktion kuvaajan muoto.**
- **Todennäköisyyden keskittyminen tietyn hajonnanmitan etäisyydelle keskiarvon ympärille on vakio kaikille normaalijakaumille parametrien arvosta riippumatta. Eli kun**

  $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ **ja** $X_2 \sim N(\mu_2, \sigma_2^2)$ **niin esim.**

  $P(X_1 \leq a) = P(X_2 \leq b)$**, jos** $\dfrac{a-\mu_1}{\sigma_1} = \dfrac{b-\mu_2}{\sigma_2}$

JATKUVAT SATUNNAISMUUTTUJAT: NORMAALIJAKAUMA

- **Normaalijakaumaan liittyvät todennäköisyydet lasketaan, kuten muillakin jatkuvilla jakaumilla, kertymäfunktion avulla.**
- **Normaalijakauman kertymäfunktion lauseketta ei kuitenkaan voida muodostaa, koska tiheysfunktion integraalia ei pystytä määrittämään muutoin kuin numeerisesti. Kertymäfunktion arvoja on taulukoitu yleensä jakaumalle $N(0,1)$ eli *standardoidulle* normaalijakaumalle.**
- **Jos $X \sim N(\mu, \sigma^2)$, on standardoitu muuttuja**

  $$Z = \frac{X - \mu}{\sigma}$$

  normaalisti jakautunut parametrein 0 ja 1, eli $Z \sim N(0,1)$, Tämän jakauman arvot ovat "hajonnanmittoja". Kuten aikaisemminkin esim.

  $P(X \leq a) = P(Z \leq z)$, **jossa** $z = \dfrac{a - \mu}{\sigma}$

JATKUVAT SATUNNAISMUUTTUJAT: NORMAALIJAKAUMA

- **Standardoidun normaalijakauman kertymäfunktiosta käytetään yleensä lyhenteen F sijasta tunnusta Φ.**
- **Johtuen siitä, että jakauma on symmetrinen pätee kertymäfunktiolle**

  $\Phi(-z)=1-\Phi(z)$
- **Jakauman arvoja ja niihin liittyviä todennäköisyyksiä on taulukoitu.**
- **Aikaisemman ominaisuuden vuoksi taulukoista ei löydy negatiivisia muuttujan arvoja vastaavia kertymäfunktion arvoja.**

NORMAALIJAKAUMA: ESIMERKKEJÄ

1. **Olkoon osakkeen suhteellisella tuotolla normaalijakauma odotusarvonaan *0.05 (=5%*) ja keskihajontana *0.01*. Mikä on todennäköisyys, että suhteellinen tuotto ylittää arvon *0.04*?**
2. **Mikä on sellainen suhteellisen tuoton arvo, jolle suuremman arvon todennäköisyys on *0.025*?**

JATKUVAT SATUNNAISMUUTTUJAT: NORMAALIJAKAUMA: KESKEINEN RAJA-ARVOLAUSE

- Olkoon $X_1, X_2, \ldots, X_n$ joukko samoin jakautuneita, riippumattomia (***i.i.d***) satunnaismuuttujia, joiden yhteinen odotusarvo μ ja yhteinen varianssi σ^2 ovat äärellisiä.
- Tällöin niiden keskiarvon todennäköisyysjakauma lähestyy normaalia, kun n kasvaa rajatta, eli $\bar{X} \to N(\mu, \sigma^2 / n)$.
- Tällöin niiden summan todennäköisyysjakauma lähestyy normaalia, kun n kasvaa rajatta, eli $S \to N(\mathrm{n}\mu, \mathrm{n}\sigma^2)$
- Edellisissä keskiarvon ja summan odotusarvot ja varianssit on määritelty niiden yleisten ominaisuuksien avulla.

JATKUVAT SATUNNAISMUUTTUJAT: NORMAALIJAKAUMA: YHTEENLASKUOMINAISUUS

- **Olkoon X_1, X_2, . . . , X_n joukko normaalisti jakautuneita, riippumattomia satunnaismuuttujia, joiden yhteinen odotusarvo μ ja yhteinen varianssi σ^2 ovat äärellisiä.**
- **Tällöin niiden keskiarvon todennäköisyysjakauma on normaali, eli** $\bar{X} \sim N(\mu, \sigma^2 / n)$
- **Tällöin niiden summan todennäköisyysjakauma on normaali, kun n kasvaa, eli** $S \sim N(\mathrm{n}\mu, \mathrm{n}\sigma^2)$
- **Edellisissä keskiarvon ja summan odotusarvot ja varianssit on määritelty niiden yleisten ominaisuuksien avulla.**

NORMAALIJAKAUMA: ESIMERKKEJÄ

1. **Olkoon $X_1,\ldots,X_{30}$ riippumattomia ja samoin jakautuneita osakkeiden suhteellisia tuottoja, joiden odotusarvot ovat 0.02 ja keskihajonta 0.05. Mikä on todennäköisyys että näistä osakkeista koostuvan sijoitussalkun (1 kpl kutakin) suhteellinen tuotto on korkeintaan 0.03?**
2. **Olkoon $X_1,\ldots,X_5$ riippumattomia ja normaalisti jakautuneita satunnaismuuttujia (odotusarvoilla 3.0 ja keskihajonta 0.05). Mikä on todennäköisyys, että näiden muuttujien arvoista laskettu keskiarvo ylittää arvon 3.1?**

JATKUVAT SATUNNAISMUUTTUJAT: *t*-JAKAUMA

- **Jos satunnaismuuttujalla X on t-jakauma, merkitään**

  $X \sim t^{(v)}$**,**

  jossa v on t-jakauman muotoa määrittävä ns. vapausasteluku. Jakauman tiheysfunktion muoto on sitä lähempänä standardoitua normaalijakaumaa, mitä suurempi v on.
- **t-jakaumaa käytetään vastaavasti kuin standardoitua normaalijakaumaa.**
- **Jakauman arvoja ja niihin liittyviä todennäköisyyksiä on taulukoitu.**

JATKUVAT SATUNNAISMUUTTUJAT: χ^2-JAKAUMA

- **Jos satunnaismuuttujalla X on χ^2-jakauma, merkitään**

  $X \sim \chi^2_{(v)}$ **,**

  jossa v on χ^2-jakauman muotoa määrittävä vapausasteluku, joka määrittää jakauman muotoa.
- **Jakauman arvoja ja niihin liittyviä todennäköisyyksiä on taulukoitu.**

JATKUVAT SATUNNAISMUUTTUJAT: F-JAKAUMA

- **Jos satunnaismuuttujalla X on F-jakauma, merkitään**
  $X \sim F^{(v_1, v_2)}$,
  jossa v_1 ja v_2 ovat F-jakauman muotoa määrittävät vapausasteluvut.
- **Käytetään tilastollisessa päättelyssä mm. keskihajontaan/varianssiin liittyen.**
- **F-jakauman ominaisuus: jos satunnaismuuttuja X on F-jakautunut vapausasteilla v_1 ja v_2, niin satunnaismuuttuja $Y = \frac{1}{X}$ on $\boldsymbol{F}$-jakautunut vapausasteilla v_2 ja v_1. Tästä seuraa, että** $F_p^{(v_1, v_2)} = \dfrac{1}{F_{1-p}^{(v_2, v_1)}}$.
- **Jakauman arvoja ja niihin liittyviä todennäköisyyksiä on taulukoitu.**

JAKAUMATAULUKOIDEN KÄYTTÖ: ESIMERKKEJÄ

- **Määritä seuraavat muuttujien arvot taulukoiden avulla:**

$$t_{0.025}^{(15)} \quad t_{0.975}^{(10)} \quad \chi^{2(20)}_{0.025} \quad F_{0.975}^{(20,10)}$$

$$z_{0.8} \quad z_{0.9} \quad \chi^{2(20)}_{0.925} \quad F_{0.025}^{(10,20)}$$

- **Määritä seuraavat todennäköisyydet taulukoiden avulla:**

$$P(t^{(10)} > 2.5) \quad P(\chi^{2(10)} > 22.5) \quad P(F^{(30,10)} > 12.5)$$

$$P(-1.96 < z < 1.96)$$

TILASTOLLINEN PÄÄTTELY

- **Monisteen alussa tutustuttiin tilastollisen tutkimuksen lähtökohtiin. Tutkimuksen kohteena on tilastoyksiköistä muodostuva perusjoukko eli populaatio ja tutkimuksessa tarkastellaan tilastoyksiköihin liittyviä ominaisuuksia, tilastollisia muuttujia. Jos tutkitaan kaikki populaatioon kuuluvat yksiköt, on kyseessä kokonaistutkimus.**
- **Tavallisempaa kuitenkin on otantatutkimuksen suorittaminen. Tällöin populaatiosta poimitaan jollakin otantamenetelmällä satunnaisotos, tutkitaan otokseen kuuluvat tilastoyksiköt ja yleistetään tutkimuksen tulokset koskemaan koko perusjoukkoa. Tämänkaltainen menettely edustaa ns. *induktiivista päättelyä*. Tilastollisia tunnuslukuja voidaan em. Syystä johtuen käsitellä satunnaismuuttujina.**

TILASTOLLINEN PÄÄTTELY

- Tilastollinen päättely voidaan jakaa kahteen toisiaan täydentävään lähestymistapaan. Nämä ovat (1) *estimointi* ja (2) *hypoteesien testaus* eli *tilastollinen merkitsevyystestaus*.
- Estimoinnissa on kyse perusjoukon tuntemattomien parametrien arvioimisesta eli estimoinnista otoksen antaman informaation perusteella. Estimointi voidaan jakaa *piste-estimointiin* ja *väliestimointiin*.

TILASTOLLINEN PÄÄTTELY: TUNNUSLUKUJEN MERKINTÖJÄ (ESIM.)

Tunnusluku	Populaatio	Otos
Keskiarvo Kahden keskiarvon ero	μ $\mu_1 - \mu_2$	$\bar{x}$ $\bar{x}_1 - \bar{x}_2$
Keskihajonta	σ	s
Suhteellinen frekvenssi Kahden suhteellisen frekvenssin ero	π $\pi_1 - \pi_2$	p $p_1 - p_2$
Korrelaatiokerroin	ρ	r

PISTE-ESTIMOINTI

- Estimoinnissa pyritään otoksen perusteella antamaan mahdollisimman hyvä arvio tutkimuksen kohteena olevalle populaation parametrille.
- Parametri on usein jonkin populaatiosta mitattavan ominaisuuden empiirisen jakauman tunnusluku, kuten keskiarvo, mutta se voi olla jokin muukin suure, esim. tietyn ominaisuuden omaavien yksiköiden suhteellinen osuus populaatiossa.
- Estimoinnin lähtökohtana on populaatiosta poimittu satunnaisotos.

PISTE-ESTIMOINTI

- **Arvio eli *estimaatti* populaation tunnusluvulle lasketaan mitatuista arvoista tapaukseen soveltuvan laskukaavan eli *estimaattorin* avulla.**
- **Estimaattorin arvo riippuu paitsi mitatuista havaintoarvoista, myös otoskoosta.**
- ***Otantavirheeksi* e sanotaan otoksesta lasketun arvion $\hat{T}$ ja populaation parametrin oikean arvon T erotusta, eli**
  $e = \hat{T} - T$.
- **Jos halutaan tietää, kuinka luotettavia estimaatteja tietyllä estimaattorilla saadaan, on tunnettava ko. estimaattorin *otantajakauma*. Otantajakaumaksi sanotaan yleisesti otossuureen todennäköisyysjakaumaa. Tässä käytetään hyväksi aikaisemmin opittuja todennäköisyyslaskennan tuloksia.**

PISTE-ESTIMOINTI: HYVÄN ESTIMAATTORIN OMINAISUUKSIA.

- Estimaattorin hyvyyttä voidaan tarkastella useista eri näkökulmista. Näitä ovat harhattomuus, varianssin pienuus ja tarkentuvuus.

VÄLIESTIMOINTI

- Yksittäisen estimaattiarvon sijasta määritetäänkin sellainen populaation tunnusluvun T arvoväli, joka sisältää oikean T:n arvon halutun suuruisella todennäköisyydellä

  $1-\alpha$.
- α:aa kutsutaan myös merkitsevyystasoksi ja se määrätään itse (yleensä 0.05).
- Toinen tulkinta: Väli ei sisällä T:tä tietyllä ”riskillä” α. Tulkinta on aina sama tunnusluvusta riippumatta.
- Tällaista väliä kutsutaan luottamusväliksi. Esim. Jos $\alpha=0.05$, kutsutaan väliä 95% luottamusväliksi.
- Luottamusvälit ovat otantajakaumista laskettuja välejä.

VÄLIESTIMOINTI (KESKIARVO)

- **Esimerkiksi jos** $X_i \sim N(\mu, \sigma^2)$ **eli** $\bar{X} \sim N(\mu, \sigma^2 / n)$ **sisältää väli**

  $$[\bar{x} - 1{,}96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96 \frac{\sigma}{\sqrt{n}}]$$

  perusjoukon keskiarvon μ **todennäköisyydellä** 0.95 **(Tai: Riski, että ei sisällä on** 0.05**).**
- **Väliä kutsutaan perusjoukon keskiarvon** μ 95% **luottamusväliksi. Yleisemmin**

  $$V_{100(1-\alpha)} = \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$
- **Usein perusjoukon keskihajontaa ei tunneta, vaan se joudutaan estimoimaan otoksesta (**s**). Tällöin keskihajonnan kertoimet määritellään käyttäen** t**-jakaumaa (**$t_{\alpha/2}^{(n-1)}$ **). Kyseinen tapa sisältää pienillä otoksilla perusjoukon normaalijakaumaoletuksen.**

LUOTTAMUSVÄLIN MÄÄRITTÄMINEN: ESIMERKKEJÄ (KESKIARVO)

1. Otoksessa, jonka koko oli *40* henkilöä, keski-ikä oli *32.3* vuotta ja keskihajonta *3.3* vuotta. Määritä 95% luottamusväli populaation keski-iälle ja tulkitse se.
2. Tuotteen paino tulisi olla *10.0* g. Otoksessa (*n=15*) keskipaino oli *10.2* g ja keskihajonta *0.2* g. Määritä keskipainolle *95%* luottamusväli ja tulkitse se. Painojen oletetaan olevan normaalisti jakautuneita.
3. Määritä edellisestä esimerkistä *99%* luottamusväli.

TILASTOLLINEN MERKITSEVYYSTESTAUS

- **Tilastollisessa merkitsevyystestauksessa, kuten estimoinnissakin, tehdään perusjoukkoa koskevia päätelmiä otoksen tuottaman havaintoaineiston pohjalta, mutta asetelma on erilainen kuin estimoinnissa.**
- **Lähtökohtana jokin perusjoukon ominaisuutta koskeva väittämä eli ns. nollahypoteesi, joka on jokin populaatiota koskeva väittämä tai ennakko-oletus.**
- **Testi sisältää päättelysäännön, jonka perusteella aineiston sisältämä informaatio on mahdollisesti sopusoinnussa nollahypoteesin kanssa.**

TILASTOLLINEN MERKITSEVYYSTESTAUS

- **Nolla- ja sille vastakkainen vastahypoteesi määräytyvät tutkimusongelman perusteella, mutta lisäksi on otettava huomioon, millaisia tilastollisia testejä on käytettävissä.**
- **Karkein jako testien välillä voidaan tehdä 1) *parametrisiin* ja 2) *ei-parametrisiin* eli *jakaumasta vapaisiin* testeihin.**
- **Parametrisia testejä voidaan tehdä vähintään välimatka-asteikon tasoisille muuttujille, kategorisille muuttujille taas soveltuvat ei-parametriset testit.**

TILASTOLLINEN MERKITSEVYYSTESTAUS: 1) TESTIN, HYPOTEESIEN JA MERKITSEVYYSTASON VALINTA

- Testin valinta riippuu käyttötarkoituksesta. Testauksen kohteena voivat olla mm. seuraavat tuntemattomat populaatiotunnusluvut (ks. aikaisempi taulukko s.55):
  - Keskiarvo μ, kahden keskiarvon ero
  - Suhteellisen osuus π, kahden suhteellisen osuuden ero
  - Korrelaatiokerroin ρ
- Em. Tunnusluvuille voidaan määrittää myös vastaavat luottamusvälit.

TILASTOLLINEN MERKITSEVYYSTESTAUS: 1) TESTIN, HYPOTEESIEN JA MERKITSEVYYSTASON VALINTA

- **Testissä muodostetaan nolla- ja sille vastakkainen vastahypoteesi. Molemmat ovat perusjoukkoa koskevia väittämiä eli ne lausutaan ed. sivun populaatiotunnuslukujen avulla.**
- **Testit ovat joko yksi- tai kaksisuuntaisia.**
- **Kaksisuuntainen testi populaation tunnusluvulle T:**

  $H_0: T = T_0$

  $H_1: T \neq T_0$

  Tällöin aineisto ei tue H_0 :ia, kun poikkeama otoksessa on sen mukaisesta arvosta riittävä jompaankumpaan suuntaan.

”Testaa, poikkeaako perusjoukon tunnusluku luvusta…”, tms.

TILASTOLLINEN MERKITSEVYYSTESTAUS: 1) TESTIN, HYPOTEESIEN JA MERKITSEVYYSTASON VALINTA

- **Merkitsevyystaso α valitaan yleensä joko 0.05 tai 0.01:ksi.**
- **Merkitsevyystason tulkinta: Suurin sallittu erehtymisen riski, kun H_0 hylätään eli todetaan, että aineisto ei sitä tue.**
- **Oikea testi valitaan hypoteesien pohjalta. Valittu testi voi olla joko z-, t-, χ^2- tai F-testi riippuen ns. testisuureen jakaumasta.**

TILASTOLLINEN MERKITSEVYYSTESTAUS: 2) TESTISUUREEN HAVAITUN ARVON SEKÄ HAVAITUN MERKITSEVYYSTASON LASKEMINEN

- Testisuureen havaittu arvo esittää otantavirheen muodossa, josta ns. havaittu merkitsevyystaso voidaan laskea. Tämä on usein standardoitu muoto. Esim. jos tarkasteltavana tunnuslukuna on populaation keskiarvo μ, toimii testisuureena standardoitu otoskeskiarvo.
- Testisuureen kaava ja todennäköisyysjakauma riippuvat valitusta testistä. Siitä käytetään jatkossa lyhennettä, josta on pääteltävissä myös sen todennäköisyysjakauma, esim. z_{hav}.
- Otantavirhe: Väitetyn populaatiotunnusluvun ja otostunnusluvun välinen ero.
- Havaittu merkitsevyystaso eli p-arvo: Lasketaan määrittämällä vähintään havaitun suuruisen otantavirheen todennäköisyys (lasketaan testisuureen arvon avulla).
- p-arvon tulkinta: Virheen riski, kun H_0 hylätään eli kun todetaan, että aineisto ei sitä tue.

TILASTOLLINEN MERKITSEVYYSTESTAUS: 3) HAVAITUN MERKITSEVYYSTASON JA MERKITSEVYYSTASON VERTAILU

- **Päätössääntö: Jos $p<\alpha$, aineiston ei katsota tukevan nollahypoteesia (tehtäessä näin erehtymisen riski on riittävän pieni).**

TILASTOLLINEN MERKITSEVYYSTESTAUS: 4) TULOSTEN TULKINTA JA KRITIIKKI.

- Tärkeää: tulosten esittäminen selväkielisesti testatun tunnusluvun avulla tulkittuna.
- Jos aineisto ei tue H_0:aa, vastaavan luottamusvälin laskenta.
- Poikkeaman merkittävyyden arviointi sekä otoskoon ja hajonnan arviointi.

KESKIARVOTESTEJÄ

- **Yhden otoksen keskiarvotesti: Yhden perusjoukon keskiarvon tarkastelu.**
- **Kahden riippumattoman otoksen keskiarvotesti: Kahden perusjoukon keskiarvon vertailu, perusjoukot toisensa poissulkevia.**
- **Kahden riippuvan otoksen keskiarvotesti: esim. sama perusjoukko kahtena eri ajanhetkenä.**

YHDEN OTOKSEN KESKIARVOTESTI

- **Kaksisuuntainen testi populaation keskiarvolle μ:**

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

YHDEN OTOKSEN KESKIARVOTESTI

- **Jos σ tunnettu, on testisuure**

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

(ks. ns. normaalijakauman yhteenlaskuominaisuus).

- **Jos σ tuntematon, on testisuure**

$$t_{hav} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \sim t^{(n-1)}$$

- **Kummassakin em. tapauksissa oletetaan havaintojen olevan peräisin normaalisti jakautuneesta perusjoukosta.**
- **Jos otoskoko on suuri ($n>30$), ei em. normaalijakaumaoletusta tarvita, ja testisuureen jakaumana voidaan käyttää $N(0,1)$ -jakaumaa myös jälkimmäisessä tapauksessa (ks. ns. keskeinen raja-arvolause).**

YHDEN OTOKSEN KESKIARVOTESTI

- **p-arvo 2-suuntaisessa testissä**

  $p = 2 \cdot P(Z > z_{hav})$, **jos** $z_{hav} > 0$

  $p = 2 \cdot P(Z < z_{hav})$, **jos** $z_{hav} < 0$

  eli

  $$p = 2 \cdot (1 - \Phi(|z_{hav}|))$$

- **Kuten aikaisemminkin on todettu, jos testisuureella on t-jakauma, lasketaan p-arvo käyttäen sitä, eli** $p = 2 \cdot P(t^{(n-1)} > |t_{hav}|)$**, jne. (vrt. myös μ:n luottamusvälin kaava).**

TILASTOLLINEN MERKITSEVYYSTESTAUS: ESIMERKKEJÄ (KESKIARVO)

1. **Voidaanko olettaa, että populaation keski-ikä on *32* vuotta, kun otoksessa (*n=32*) se oli *28.2* vuotta. Otoksen keskihajonta oli *1.2* vuotta. Käytä merkitsevyystasoa *0.05*.**
2. **Urheiluvälinetehdas ilmoittaa valmistamiensa 41-numeroisten kenkien painoksi *160* g. Asiaa tarkistettaessa saatiin *16* kengän otoksesta keskiarvoksi *164* g ja keskihajonnaksi *6* g. Testaa tehtaan ilmoittaman keskipainon paikkansapitävyyttä *5%* merkitsevyystasolla, kun oletetaan, että kenkien painon jakauma populaatiossa on normaali.**

MITEN JA MIKSI TARKASTELLA NORMAALIJAKAUMAOLETUSTA?

- Tarpeellista tehdä, jos otoskoko on pieni.
- Histogrammi otosjakaumasta.
- Normaalisuustesti.
- Jos ei perusteltu, testin tulokset eivät luotettavia.

YHTEENVETO TILASTOLLISEN TESTIN SUORITTAMISESTA

1a. Määritä nollahypoteesi ja tilanteeseen soveltuva vastahypoteesi.

1b. Valitse testissä käytettävä merkitsevyystaso α.

1c. Valitse käytettävä testi ja sen mukainen testisuure.

2. Kerää tarvittaessa otoksesta mitattu data ja laske sen avulla testisuureen arvo (eli määritä testisuureen havaittu arvo) ja p-arvo.

3. Vertaa p-arvoa kohdassa 1b valittuun merkitsevyystasoon ja tämän perusteella toteat aineiston tukevan nollahypoteesia/et totea. Määritä tarvittaessa vastaava luottamusväli.

4. Tulkitse testauksen tulokset selväkielisesti. Tarkastele tulosten luotettavuutta.

KAHDEN RIIPPUMATTOMAN OTOKSEN KESKIARVOTESTI

- **Kaksisuuntainen testi kahden populaation keskiarvojen μ_1 ja μ_2 erolle:**

  H_0: $\mu_1 = \mu_2$
  H_1: $\mu_1 \neq \mu_2$

- **Testi voidaan suorittaa seuraavin vaihtoehtoisin oletuksin:**
  (1) σ_1 **ja** σ_2 **tunnettuja,**
  (2) σ_1 **ja** σ_2 **tuntemattomia, mutta oletetaan** $\sigma_1=\sigma_2$
  (3) σ_1 **ja** σ_2 **tuntemattomia, mutta oletetaan** $\sigma_1 \neq \sigma_2$

- **Otoksien oletetaan olevan peräisin normaalisti jakautuneista perusjoukoista. Suurissa otoksilla tätä havaintojen normaalisuusoletusta ei tarvita (riittävästä otoskoosta ks. seuraavat sivut).**

KAHDEN RIIPPUMATTOMAN OTOKSEN KESKIARVOTESTI: OLETUS: POPULAATIOHAJONNAT TUNTEMATTOMIA, MUTTA OLETETAAN YHTÄ SUURIKSI.

- **Testisuure perustuu otoskeskiarvojen erotukseen ja niiden otantajakaumaan:**

$$t_{hav} = \frac{\overline{X}_1 - \overline{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t^{(n_1+n_2-2)}$$

missä

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

- **Populaatiokeskiarvojen erotuksen** $\mu_1 - \mu_2$ $100(1-\alpha)\%$**:n luottamusväli on**

$$V_{100(1-\alpha)} = \left[\overline{X}_1 - \overline{X}_2 - t_{\alpha/2}^{(n_1+n_2-2)} s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \overline{X}_1 - \overline{X}_2 + t_{\alpha/2}^{(n_1+n_2-2)} s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

- **Mikä otoskoot ovat suuret,** $n_1 > 30$ **ja** $n_2 > 30$**, testi suoritetaan (ja luottamusväli lasketaan) käyttäen standardoitua normaalijakaumaa,** eikä **havaintojen normaalijakaumaoletusta tarvita.**

KAHDEN RIIPPUMATTOMAN OTOKSEN KESKIARVOTESTI: OLETUS: POPULAATIOHAJONNAT TUNTEMATTOMIA, MUTTA OLETETAAN YHTÄ SUURIKSI: ESIMERKKI.

Kaksi eri tuotantolinjaa, A ja B, tuottivat tietokoneen osia, joiden piti olla keskimäärin samanpainoisia. Tuotannosta poimittu satunnaisotos tuotti seuraavat tulokset:

	A	B
painon keskiarvo	*12.1*	*12.5*
painon keskihajonta	*0.2*	*0.3*
Otoskoko	*35*	*34*

Testaa, tuleeko tuotantolinjoilta keskimäärin samanpainoisia tuotteita. Käytä *5%* **merkitsevyystasoa. Perusjoukon hajonnat oletetaan yhtä suuriksi. Määritä myös keskiarvojen eron** *95%* **luottamusväli.**

KAHDEN RIIPPUMATTOMAN OTOKSEN KESKIARVOTESTI: OLETUS: POPULAATIOHAJONNAT TUNTEMATTOMIA, MUTTA OLETETAAN ERI SUURIKSI.

- **Testisuure perustuu otoskeskiarvojen erotukseen ja niiden otantajakaumaan. Jos otoskoot ovat kohtalaisen suuria,** $n_1>20$ **ja** $n_2>20$,

$$t_{hav=} \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t^{(v)}$$

missä

$$\frac{1}{v} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1} \text{ ja } \qquad c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

- **Populaatiokeskiarvojen erotuksen** $\mu_1 - \mu_2$ $100(1-\alpha)\%$:**n luottamusväli on**

$$V_{100(1-\alpha)} = \left[\bar{X}_1 - \bar{X}_2 - t_{\alpha/2}^{(v)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{X}_1 - \bar{X}_2 + t_{\alpha/2}^{(v)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

- **Mikä otoskoot ovat suuret,** $n_1>50$ **ja** $n_2>50$, **testi suoritetaan (ja luottamusväli lasketaan) käyttäen standardoitua normaalijakaumaa, eikä havaintojen normaalijakaumaoletusta tarvita.**

KAHDEN RIIPPUMATTOMAN OTOKSEN KESKIARVOTESTI: OLETUS: POPULAATIOHAJONNAT TUNTEMATTOMIA, MUTTA OLETETAAN YHTÄ SUURIKSI: ESIMERKKI.

Testaa, ovatko miehet ja naiset keskimäärin yhtä tyytyväisiä tuotteeseen X, kun *35* **miehen arvosanojen keskiarvo oli** *3.4* **ja** *35* **naisen** *3.45* **(keskihajonnat** *0.1* **ja** *0.2***). Hajonnat populaatiossa oletetaan eri suuriksi ja jakaumat normaaleiksi. Käytä** *5%* **merkitsevyystasoa.**

Määritä myös keskimääräisen arvosanan eron *95%* **luottamusväli.**

KAHDEN RIIPPUVAN OTOKSEN KESKIARVOTESTI

- Tarkastellaan koejärjestelyä, jossa halutaan verrata kahden eri menetelmän tai käsittelyn vaikutusta tilastoyksiköiden johonkin ominaisuuteen. Kokeen suorittamiseen tarvitaan kaksi tilastoyksikköjen ryhmää, joista toiseen sovelletaan käsittelyä A ja toiseen käsittelyä B.
- Käsittelyvaikutuksen selville saamiseksi on tärkeää eristää muiden samaan ominaisuuteen vaikuttavien muuttujien vaikutus kokeen aikana.
- Jos koeryhmät muodostetaan jakamalla käytettävissä olevien tilastoyksiköiden joukko kahteen ryhmään, voidaan tulosten analysointiin käyttää riippumattomien otosten keskiarvotestiä. Nyt ongelmaksi voi kuitenkin muodostua tilastoyksiköiden keskinäinen heterogeenisuus, joka luonnollisista syistä aiheuttaa eri ryhmistä tehtävien havaintojen jakaumiin vaihtelua ja haittaa täten käsittelyvaikutusten mittaamista.

KAHDEN RIIPPUVAN OTOKSEN KESKIARVOTESTI

- Edellisen kaltaisten häiriötekijöiden vaikutusta voidaan vähentää kahdella tavalla:

  1. Tilastoyksiköistä muodostetaan yhteen sovitettuja pareja siten, että kunkin parin yksiköt ovat keskenään mahdollisimman samankaltaisia niiden ominaisuuksien suhteen, joilla saattaisi olla vaikutusta tutkittavaan ominaisuuteen. Tämän jälkeen kunkin parin kohdalla toiseen sovelletaan käsittelyä A ja toiseen käsittelyä B.

  2. Samoille tilastoyksiköille suoritetaan kumpikin käsittelyistä A ja B (jokainen tilastoyksikkö on “itse itsensä pari”).

KAHDEN RIIPPUVAN OTOKSEN KESKIARVOTESTI

- **Käsittelyjen erojen selvittämiseksi mitataan tilastoyksiköistä käsittelyvaikutuksen indikaattorina olevan muuttujan arvo ja lasketaan kustakin parista arvojen parittaiset erotukset (D).**
- **Kaksisuuntainen testi populaation keskimääräiselle erotukselle μ_D:**

H_0: $\mu_D = 0$ **(käsittelyillä ei ole eroa)**

H_1: $\mu_D \neq 0$

KAHDEN RIIPPUVAN OTOKSEN KESKIARVOTESTI

- **Erotukset on poimittu normaalisti jakautuneesta perusjoukosta, jolloin testisuure:**

$$t_{hav} = \frac{\overline{D}}{\frac{s_D}{\sqrt{n}}} \sim t^{(n\text{-}1)}$$

missä

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \quad \textbf{ja} \quad s_D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2}$$

KAHDEN RIIPPUVAN OTOKSEN KESKIARVOTESTI

- **Keskimääräisen erotuksen** μ_D $100(1-\alpha)\%$**:n luottamusväli on**

$$V_{100(1-\alpha)} = \left[\overline{D} - t_{\alpha/2}^{(n-1)} \frac{s_D}{\sqrt{n}}, \overline{D} + t_{\alpha/2}^{(n-1)} \frac{s_D}{\sqrt{n}} \right]$$

- **Mikä otoskoko on suuri,** $n>30$**, testi suoritetaan (ja luottamusväli lasketaan) käyttäen standardoitua normaalijakaumaa, eikä havaintojen (erotusten) normaalijakaumaoletusta ei tarvita.**

KAHDEN RIIPPUVAN OTOKSEN KESKIARVOTESTI: ESIMERKKEJÄ

Testaa, asiakastyytyväisyydessä tapahtunut muutosta, kun *7* **asiakkaan tyytyväisyyden arvosanat (asteikolla 1-5) olivat vuonna 2009** *3,4,3,4,4,3* **ja** *5* **sekä 2010** *3,5,4,5,4,5* **ja** *5*. **Käytä** *5%* **merkitsevyystasoa. Arvosanojen muutosten oletetaan noudattavan populaatiossa normaalijakaumaa.**

Määritä myös arvosanan keskimääräisen muutoksen *95%* **luottamusväli.**

VARIANSSIN TESTAUS

- Yhden varianssin tarkastelu
- Kahden varianssin vertailu
- Kummassakin testissä oletetaan aina otoksen/otoksien olevan peräisin normaalisti jakautuneista perusjoukoista.
- Testit ovat analogisia keskihajonnan testien kanssa: Tietty varianssiin liittyvä nollahypoteesi vastaa aina tiettyä keskihajontaan liittyvää nollahypoteesia (koska keskihajonta on varianssin neliöjuuri).

KAHDEN RIIPPUMATTOMAN OTOKSEN HAJONNAN TESTAUS

- **Kaksisuuntainen testi kahden populaation varianssin σ_1^2 ja σ_2^2 suhteelle:**

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

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

- **Normaalijakaumaoletusten vallitessa on testisuure**

$$F_{hav} = \frac{s_1^2}{s_2^2} \sim F^{(n_1-1, n_2-1)}$$

KAHDEN RIIPPUMATTOMAN OTOKSEN HAJONNAN TESTAUS: ESIMERKKEJÄ

Testaa, ovatko kahden tuotantolinjan tuotteiden painojen hajonnat samoja, kun kahden otoksen ($n_1=31$ **ja** $n_2=31$**) keskihajonnat olivat** 5.5 **g ja** 6.6 **g. Käytä** 5% **merkitsevyystasoa. Havaintojen oletetaan olevan peräisin normaalisti jakautuneista perusjoukoista.**

KORRELAATIOKERTOIMEN TESTAUS

- **Kaksisuuntainen testi populaation korrelaatiokertoimelle ρ:**

  H_0: $\rho = 0$ **(ei lineaarista riippuvuutta)**

  H_1: $\rho \neq 0$ **(on lineaarista riippuvuutta)**

KORRELAATIOKERTOIMEN TESTAUS

- **Normaalijakaumaoletusten vallitessa on testisuure**

$$t_{hav} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t^{(n\text{-}2)}$$

- **(Tarkalleen ottaen normaalijakauma oletus kuuluu: Kahden muuttujan yhteisjakauma on kaksiulotteinen normaalijakauma)**

KORRELAATIOKERTOIMEN TESTAUS: ESIMERKKEJÄ

Pankkitoimihenkilöistä poimitussa otoksessa ($n=42$) palkan ja työkokemuksen vuosissa välinen Pearsonin korrelaatiokerroin oli 0.42. Testaa 5% merkitsevyystasolla, onko palkan ja työkokemuksen välillä lineaarista riippuvuutta populaatiossa.

EPÄPARAMETRISIA TESTEJÄ

- Epäparametrisiin testeihin ei liity normaalijakaumaoletuksia.
- Tarkasteltavat muuttujat ovat kategorisia.
- Tarkasteltavat testit liittyvät jonkin em. tyyppisen kategorisen ominaisuuden suhteelliseen osuuteen/osuuksiin perusjoukossa.
- Seuraavassa esitetyissä testeissä testisuureen tarkka jakauma on diskreetti.
- Johtuen diskreettien jakaumien luonteesta niitä approksimoidaan jatkuvilla jakaumilla.
- Approksimoinnin hyvyyttä tarkastellaan erilaisilla mm. otoskokoon liittyvillä kriteereillä.

SUHTEELLISEN OSUUDEN TESTAUS

- **Populaation suhteellisen osuuden π estimaattorina käytetään otoksen suhteellista osuutta p.**
- **π ja p ovat suhdelukuja ja ne saavat arvoja väliltä *[0,1]*.**
- **Populaation suhteellinen osuus voidaan tulkita myös ominaisuuden *A* esiintymistodennäköisyydeksi populaatiossa.**
- **Ominaisuuden *A* suhteellinen osuus π populaatiossa on ominaisuuden A omaavien yksiköiden lukumäärä K jaettuna populaation koolla N. Silloin p:n otantajakauma on binomijakauma, jossa**

  $Ep = \pi$ **ja** $D^2 p = \dfrac{\pi(1-\pi)}{n}$.

SUHTEELLISEN OSUUDEN TESTAUS

- **Binomijakaumaa voidaan approksimoida normaalijakaumalla, mikäli otoskoko on riittävän suuri.**
- **Riittävä approksimaation tarkkuus katsotaan yleensä saavutettavan, jos**

$n\pi \geq 5$ **ja** $n(1-\pi) \geq 5$ **. Tällöin on likimain**

$p \sim N(\pi, \frac{\pi(1-\pi)}{n})$**, eli**

$$z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1)$$ **.**

YHDEN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS

- **Kaksisuuntainen testi populaation suhteelliselle osuudelle π:**

$$H_0: \pi = \pi_0$$

$$H_1: \pi \neq \pi_0$$

YHDEN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS

- **Testisuure**

$$z_{hav} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1)$$

- **Huom! Approksimaatio on riittävän tarkka, kun** $n\pi_0 \geq 5$ **ja** $n(1-\pi_0) \geq 5$**.**
- **Luottamusväli π:lle**

$$V_{100(1-\alpha)} = \left[p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}},\, p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \right]$$

- **Approksimaation hyvyyttä tarkastellaan luottamusväliä laskettaessa p:n avulla.**

YHDEN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS: ESIMERKKI.

Puoluetta A kannatti vaaleissa *19.3%* **äänestäjistä. 6kk myöhemmin kyselyssä (**$n=1000$**) puoluetta ilmoitti aikovansa äänestää seuraavissa vaaleissa** *20.1%*. **Testaa, onko puolueen kannatus muuttunut. Käytä** *5%* **merkitsevyystasoa.**

Määritä puolueen kannatusosuuden *95%* **luottamusväli ja tulkitse se.**

KAHDEN RIIPPUMATTOMAN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS

- **Kaksisuuntainen testi populaation suhteellisen osuuden erolle π_1-π_2:**

H_0: $\pi_1 = \pi_2$

H_1: $\pi_1 \neq \pi_2$

KAHDEN RIIPPUMATTOMAN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS

- **Testisuure**

$$z_{hav} = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

jossa $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$.

- **Approksimaation hyvyyttä tarkastellaan tässä p_1:n ja p_2:n avulla.**
- **Luottamusväli suhteellisen osuuksien erolle**

$$V_{100(1-\alpha)} = \left[p_1 - p_2 \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right]$$

- **Approksimaation hyvyyttä tarkastellaan myös luottamusväliä laskettaessa p_1:n ja p_2:n avulla.**

KAHDEN RIIPPUMATTOMAN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS: ESIMERKKEJÄ

Erään tuotemerkin tunnistavien osuudet olivat kahdessa eri kyselyssä (ikäryhmät -18 v ja 19-vuotta, molemmat otokset *100* havaintoa) *32%* ja *39%*. Testaa *5%* merkitsevyystasolla, onko tunnistavien osuudessa eroa kahdessa eri ikä-ryhmässä.

Määritä *95%* luottamusväli tunnistavien osuuden erolle ja tulkitse se.

KAHDEN RIIPPUVAN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS

- **Esim. ajallista muutosta voidaan tarkastella samoin kuin keskiarvon tapauksessa:**

$$H_0\colon \pi_D = 0$$

$$H_1\colon \pi_D \neq 0$$

KAHDEN RIIPPUVAN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS

- **Ristiintaulukointi (sarakemuuttuja=1.kysely, rivimuuttuja=2.kysely)**

	Kyllä	*Ei*
Kyllä	$f_{11}=a$	$f_{12}=b$
Ei	$f_{21}=c$	$f_{22}=d$

- **Testisuure:**

$$z_{hav} = \frac{(b-c)/n}{\sqrt{\frac{(b+c)-(b-c)^2/n}{n(n-1)}}} \sim N(0,1)$$

- **Testi on luotettava suurilla otoksilla.**

KAHDEN RIIPPUVAN OTOKSEN SUHTEELLISEN OSUUDEN TESTAUS: ESIMERKKI

Tee kahden riippuvan otoksen suhteellisen osuuden testi, kun

$$a=30,\ b=15,\ c=9,\ d=51,\ n=105$$

EPÄPARAMETRISIA TESTEJÄ: χ^2-YHTEENSOPIVUUSTESTAUS

- **χ^2-yhteensopivuustestillä tutkitaan, noudattaako populaation frekvenssijakauma jotain tunnettua suhteellista jakaumaa.**
- **Yhden otoksen suhteellisen osuuden testin moniluokkainen yleistys.**
- **Hypoteesit:**

H_0: **Havaintoaineisto on peräisin oletettua suhteellista jakaumaa noudattavasta populaatiosta.**

H_1: **Havaintoaineisto ei ole peräisin oletettua suhteellista jakaumaa noudattavasta populaatiosta.**

EPÄPARAMETRISIA TESTEJÄ: χ^2-YHTEENSOPIVUUSTESTAUS

- Testissä verrataan otoksesta mitattuja eli k-luokkaisen muuttujan havaittuja frekvenssejä f nollahypoteesissa oletetun jakauman mukaisiin eli odotettuihin frekvensseihin e.
- Testisuure

$$\chi^2_{hav} = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \sim \chi^2_{(k-1)}$$

- Testi suoritetaan aina ns. yksisuuntaisena.
- Kriteerit approksimaation hyvyyden tarkasteluun: Jokaisen odotetun frekvenssin tulee olla >1 ja korkeintaan 20% odotetuista frekvensseistä saa olla <5.

EPÄPARAMETRISIA TESTEJÄ: χ^2-YHTEENSOPIVUUSTESTAUS: ESIMERKKI

Makeistehtaan arvioidaan aikaisemman perusteella nallekarkkien kuluttajakunnassa vallitsevan seuraava värimieltymysten jakauma:

mieleisin nallekarkin väri:

keltainen	oranssi	vihreä	punainen	valkoinen	musta
30%	*20%*	*20%*	*10%*	*10%*	*10%*

Uutta tuotetta suunniteltaessa haluttiin selvittää ovatko kuluttajien preferenssit makeisten värien suhteen em. mukaisia. Tätä varten poimittiin *506* **henkilön otos, josta saatiin seuraavat tulokset:**

mieleisin makeisen väri:

keltainen	oranssi	vihreä	punainen	valkoinen	musta
177	*135*	*79*	*41*	*36*	*38*

Tutki ongelmaa tapaukseen käyttäen χ^2-yhteensopivuustestiä. Käytä merkitsevyystasoa *0.05***.**

EPÄPARAMETRISIA TESTEJÄ: χ^2-RIIPPUMATTOMUUSTESTAUS

- **χ^2-riippumattomuustestillä tutkitaan, ovatko muuttujan populaatiossa riippumattomia toisistaan.**
- **Kahden riippumattoman otoksen suhteellisen osuuden testin moniluokkainen yleistys.**
- **Hypoteesit:**

H_0: **Muuttujat perusjoukossa riippumattomia**

H_1: **Muuttujat perusjoukossa riippuvia**

EPÄPARAMETRISIA TESTEJÄ: χ^2-RIIPPUMATTOMUUSTESTAUS

- **Testisuure**

$$\chi^2_{hav} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} \sim \chi^2_{((r-1)(s-1))},$$

jossa r ja s ovat muuttujien luokkien lukumääriä, ja jossa

$$e_{ij} = \frac{f_{i.}f_{.j}}{n}$$

- **Approksimaation hyvyyden tarkastelu: Odotettuja frekvenssejä koskevat samat vaatimukset kuin χ^2-yhteensopivuustestissä.**
- **Testi suoritetaan aina yksisuuntaisena.**

EPÄPARAMETRISIA TESTEJÄ: χ^2-RIIPPUMATTOMUUSTESTAUS: ESIMERKKI

Aikakauslehti *Palloilija* teetti kyselyn lukijoidensa mieliurheilulajeista. Lukijoiden joukosta poimittiin satunnaisotos, josta saatiin seuraavat tulokset:

Mieliurheilulaji:

	pesäpallo	*koripallo*	*jalkapallo*	*yht.*
naiset	19	15	24	58
miehet	16	18	16	50
yht.	35	33	40	$108=n$

Testaa χ^2-riippumattomuustestillä 5% merkitsevyystasolla, onko lehden lukijakunnan keskuudessa riippuvuutta sukupuolen ja mieliurheilulajien välillä.

EPÄPARAMETRISIA TESTEJÄ: NORMAALIJAKAUMATESTIT

- **Shapiro-Wilkin testi: Sopii paremmin pienemmille otoksille.**
- **Kolmogorov-Smirnovin testi: Sopii paremmin suuremmille otoksille.**
- **Perustuvat väitetyn ja havaitun jakauman vertailuun.**
- **Hypoteesit:**

H_0: **Muuttuja noudattaa perusjoukossa normaalijakaumaa**

H_1: **Muuttuja ei noudata perusjoukossa normaalijakaumaa**

MUITA EPÄPARAMETRISIA TESTEJÄ

- Useimmille parametrisille testeille löytyy oma epäparametrinen vastineensa. Niitä voidaan käyttää, kun normaalijakaumaoletus ei ole perusteltu.
- Esimerkiksi kahden riippumattoman otoksen keskiarvotestiä vastaava epäparametrinen testi on nimeltään Mann-Whitneyn U-testi, joka perustuu järjestyslukuihin. Sen hypoteesit ovat

  H_0: Otokset ovat peräisin perusjoukoista, joiden jakauma on sama.

  H_1: Otokset ovat peräisin perusjoukoista, joiden jakauma ei ole sama.
- Em. testiä voidaan käyttää pienillä otoksilla, jotka eivät tue populaatioiden normaalijakaumaoletusta.

MUITA EPÄPARAMETRISIA TESTEJÄ

- **Yhden otoksen mediaanitesti (yhden otoksen keskiarvotestin vastine)**
- **Järjestyskorrelaatiokertoimen testi (lineaarisen korrelaatiokertoimen testin vastine).**

jne.

YKSISUUNTAINEN TESTAUS

- Yksisuuntainen testi tehdään, jos poikkeamasta nollahypoteesin mukaisesta arvosta ja erityisesti sen suunnasta on ennakkokäsitys/väittämä.
- Tällä on vaikutusta p-arvon laskemiseen ja tulosten tulkintaan.
- Edellisellä ei ole vaikusta testisuureen arvon ja jakauman määrittämiseen.

TILASTOLLINEN MERKITSEVYYSTESTAUS: TESTIN, HYPOTEESIEN JA MERKITSEVYYSTASON VALINTA

- **Yksisuuntainen testi:**

  (1) $H_0: T = T_0$ **tai** *(2)* $H_0: T = T_0$

  $H_1: T > T_0$ $H_1: T < T_0$

  Tällöin aineisto ei tue H_0 :ia, kun poikkeama otoksessa on sen mukaisesta arvosta riittävä tiettyyn suuntaan. Suoritettaessa 1-suuntainen testi, poikkeaman suunnasta on ennakko-oletus, mikä näkyy vastahypoteesissa.

”Testaa, ylittääkö/alittaako perusjoukon tunnusluku luvun…”, tms.

1-SUUNTAINEN TESTAUS: ESIMERKKEJÄ

1. **Testaa, ylittääkö populaatiokeskiarvo arvon** 3.0**, kun otoksessa keskiarvo oli** 3.1 **ja keskihajonta** 0.2 **(**$n=35$**). Käytä** 5% **merkitsevyystasoa.**
2. **Testaa, onko muuttujien välillä positiivista lineaarista riippuvuutta, kun otoskorrelaatiokerroin oli** 0.42 **ja otoskoko** 32 **havaintoa. Otosten oletetaan olevan peräisin normaalisti jakautuneista perusjoukoista.**

TILASTOLLINEN TESTAUS: LAADUNVALVONTAA

- Testisuureen havaittua arvoa laskettaessa kerätään käytännössä informaatiota nollahypoteesia vastaan. Em. arvoon vaikuttavat absoluuttisten erojen lisäksi otoskoko ja hajonta.
- Nämä tekijät vaikuttavat myös luottamusvälien pituuteen.
- Tarkastele aina analyysin jälkeen otoskoon ja hajonnan mahdollista merkitystä tulosten kannalta.
- Tilastollinen merkitsevyys ja merkittävyys eivät ole välttämättä sama asia.

TILASTOLLINEN MERKITSEVYYS JA MERKITTÄVYYS EIVÄT OLE VÄLTTÄMÄTTÄ SAMA ASIA.

1. **Testaa, onko muuttujien välillä lineaarista riippuvuutta, kun $r=0.15$ ja $n=302$.**
2. **Testaa, onko muuttujien välillä lineaarista riippuvuutta, kun $r=0.59$ ja $n=10$.**

Otosten oletetaan olevan peräisin normaalisti jakautuneista perusjoukoista.

TENTTIIN VALMISTAUTUMISESTA

- **Tutki etukäteen kaavakokoelman rakennetta.**
- **Opettele lukemaan kaavojen merkintöjä.**
- **Opettele käyttämään todennäköisyysjakaumataulukoita.**
- **Käy läpi luentoesimerkit ja kotitehtävät. Laske samoja tehtäviä hieman lähtöarvoja muuttaen.**

KIRJALLISUUTTA

Anderson, D.R., Sweeney D.J. , Williams T.A.: Statistics for Business and Economics. West Publishing Company.

Field, A. : Discovering Statistics Using IBM SPSS Statistics. Sage Publications.