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*.**

$$S = Silmälulujen\ summa$$

Tulos 1/Tulos 2	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

1. $P(S\ parillinen) =$

$$P(S = 2 \cup S = 4 \cup 6 \cup S = 8 \cup S = 10 \cup S = 12) = \frac{18}{36}$$

2. $P(S\ vähintään\ 3) = P(S = 3 \cup S = 4 \cup S = 5 \cup S = 6 \cup S = 7 \cup S = 8 \cup S = 9 \cup S = 10 \cup S = 11 \cup S = 12) = \frac{35}{36}$
3. $P(S\ ei\ ole\ 6) = P(S = 2 \cup .. \cup S = 5 \cup S = 7 \cup .. \cup S = 12) = \frac{31}{36}$
4. $P(molemmat\ nopat\ 6) = P(Tulos\ 1 = 6 \cap Tulos\ 2 = 6) = \frac{1}{36}$

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?

1. $5 * 8 * 4 = 160$ eritavalla (tuloperiaate)

2. $\frac{10!}{(10-3)!} = 10 * 9 * 8 = 720$ eri tavalla (variaatiot)

3. $\binom{15}{6} = \frac{15!}{6!(15-6)!} = 5005$ eri tavalla (kombinaatiot)

4. $P(\text{"Kaikki uutta mallia"}) = \frac{\binom{20}{5}\binom{80}{0}}{\binom{100}{5}} = \frac{\frac{20!}{5!15!} * 1}{\frac{100!}{5!95!}} =$

$\frac{20*19*18*17*16}{100*99*98*97*96} \approx 0.0002$

TODENNÄKÖISYYSLASKENNAN LASKUSÄÄNNÖT: ESIMERKKEJÄ

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

1. $P(S\ parillinen) = P(S = 2 \cup S = 4 \cup S = 6 \cup S = 8 \cup S = 10 \cup S = 12) =$

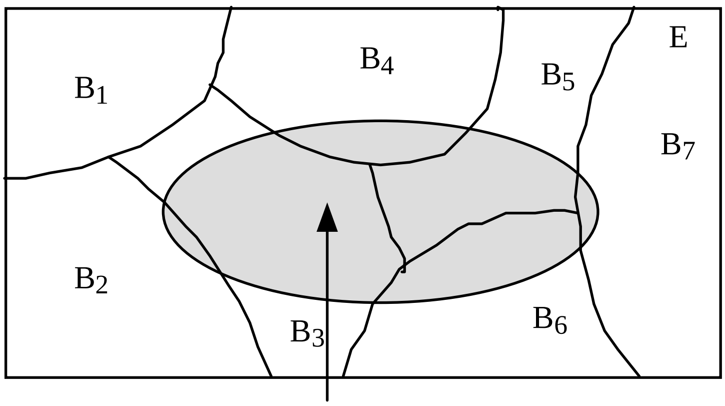
   $$\frac{1}{6} * \frac{1}{6} + 3 * \frac{1}{6} * \frac{1}{6} + 5 * \frac{1}{6} * \frac{1}{6} + 5 * \frac{1}{6} * \frac{1}{6} + 3 * \frac{1}{6} * \frac{1}{6} + \frac{1}{6} * \frac{1}{6} = \frac{18}{36}$$

2. $P(S\ vähintään\ 3) = 1 - P(S = 2) = 1 - \frac{1}{6} * \frac{1}{6} = \frac{35}{36}$
3. $P(S\ ei\ ole\ 6) = 1 - P(S = 6) = 1 - 5 * \frac{1}{6} * \frac{1}{6} = \frac{31}{36}$
4. $P(molemmat\ nopat\ 6) = P(Tulos\ 1 = 6 \cap Tulos\ 2 = 6) =$

$$\frac{1}{6} * \frac{1}{6} = \frac{1}{36}$$

KOKONAISTODENNÄKÖISYYDEN JA BAYESIN LAUSEET

Seuraavassa esiteltävät tulokset koskevat tilannetta, jossa perusjoukko E on jaettu toisensa poissulkeviin tapahtumiin B_i, $i = 1, \ldots, n$, Lisäksi oletetaan tunnetuiksi todennäköisyydet $P(B_i) > 0$. Olkoon myös A tarkasteltavan satunnaisilmiön tapahtuma (ks. seuraava kuvio).



Joukko A voidaan nyt kirjoittaa muodossa

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$$

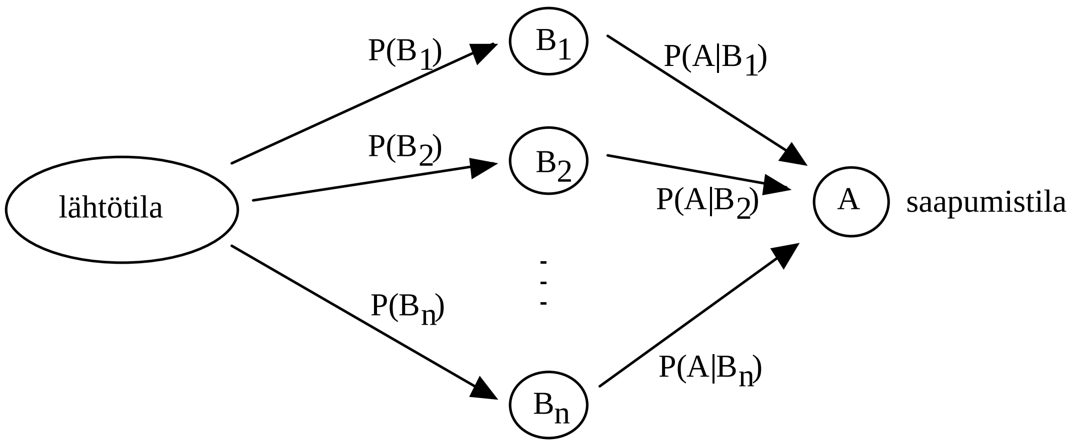
missä unionin joukot ovat toisensa poissulkevia kaikilla ij. Siis

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_n)$$

Yleisen kertolaskusäännön mukaan $P\left(A \cap B_j\right) = P\left(B_j\right) * P(A|B_j)$ kaikilla $i = 1, \ldots, n$, joten

$$P(A) = P(B_1) * P(A|B_1) + P(B_2) * P(A|B_2) + \cdots + P(B_n) * P(A|B_n)$$

Tätä lauseketta sanotaan *kokonaistodenäköisyyden* kaavaksi. Kaavaa voidaan tulkita mm. seuraavasti: “Tiloilla” B_i on tunnetut todennäköisyydet $P(B_i)$, ja tilaan A päästään vain jonkin tilan B_i kautta. Ehdollinen todennäköisyys ilmaisee ns. siirtymätodennäköisyyden tilasta B_i tilaan A. Kokonaistodennäköisyys $P(A)$ on siirtymätodennäköisyyksien painotettu keskiarvo painojen ollessa luvut $P(B_i)$ (joiden summa on 1). Seuraava kaavio havainnollistaa tätä tulkintaa.



Kokonaistodennäköisyyden kaavaa käytetään myös laskettaessa tapahtuman A todennäköisyys, kun tunnetaan ne eri reitit, joiden kautta tilaan A päädytään. Jos halutaan vastaus käänteiseen kysymykseen eli halutaan tietää, millä todennäköisyydellä tilaan A on tultu tietyn tilan B_i kautta, voidaan apuna käyttää Bayesin (“käänteistodennäköisyyden”) kaavaa

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i) * P(A|B_i)}{P(B_1) * P(A|B_1) + P(B_2) * P(A|B_2) + \cdots + P(B_n) * P(A|B_n)}.$$

Esimerkki.

Kuljetuksella on perille paikkaan A on kolme reittivaihtoehtoa (1, 2 ja 3). Kunkin reitin valintatodennäköisyydet ovat 0.2, 0.5 ja 0.3. Lisäksi tiedetään, että kuhunkin reittivaihtoehtoon liittyy todennäköisyys tapahtumaan ”olla ajoissa perillä”, jotka ovat 0.6, 0.5 ja 0.1. Määritä kokonaistodennäköisyys tapahtumalle A=”olla ajoissa perillä paikassa A”.

$$P(A) = 0.2 * 0.6 + 0.5 * 0.5 + 0.3 * 0.1 = 0.40$$

Jos tiedetään, että kuljetus saapui perille ajoissa, millä todennäköisyydellä näin tapahtui käyttäen reittiä 1?

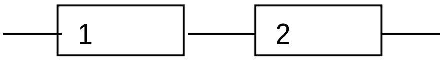
$$P(Reittiä\ 1\ perille|A) = \frac{0.2 * 0.6}{0.40} = 0.3$$

LUOTETTAVUUSANALYYSI

Systeemin luotettavuudella (R) tarkoitetaan sen toimimisen todennäköisyyttä. Systeemiin voidaan liittää tietokoneita, pumppuja, ihmisiä organisaatiossa jne. ja systeemissä voi kulkea tietoa, sähköä, vettä jne.

Esimerkki 1.

Kaksi tietokonetta on kytketty sarjaan (peräkkäin) seuraavasti:



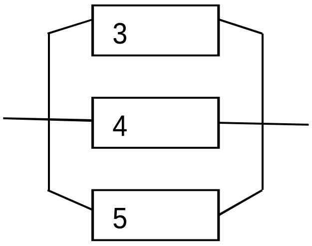
Jos molemmilla laitteilla on sama luotettavuus, on sarjasysteemin luotettavuus

$$R_1 = P(1 \cap 2) = p\,p = p^2$$

Jos molempien komponenttien luotettavuus on 0.99 on systeemin luotettavuus 0.9801.

Esimerkki 2.

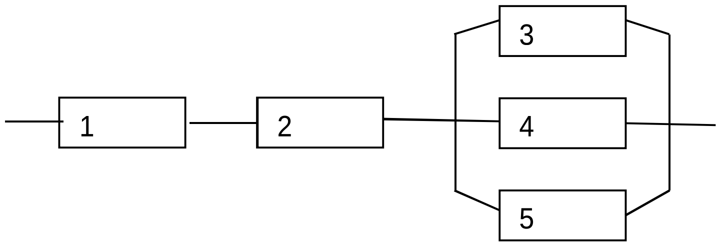
Kolme tietokonetta on kytketty rinnakkain seuraavasti:



$$R_2 = P(1 \cup 2 \cup 3) = 1 - P(1^* \cap 2^* \cap 3^*) = 1 - (1-p)^3$$

Jos komponenttien luotettavuus on 0.99 on systeemin luotettavuus 0,9999990.

Esimerkki 3.



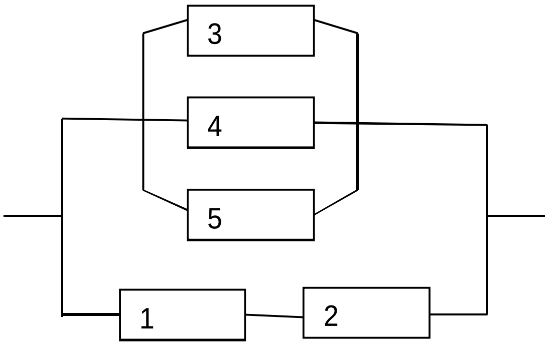
Kytketään edellisten esimerkkien systeemit sarjaan:

$$R = R_1 R_2 = p^2(1-(1-p)^3)$$

Jos komponenttien luotettavuus on 0.99 on systeemin luotettavuus 0,98009902.

Esimerkki 4.

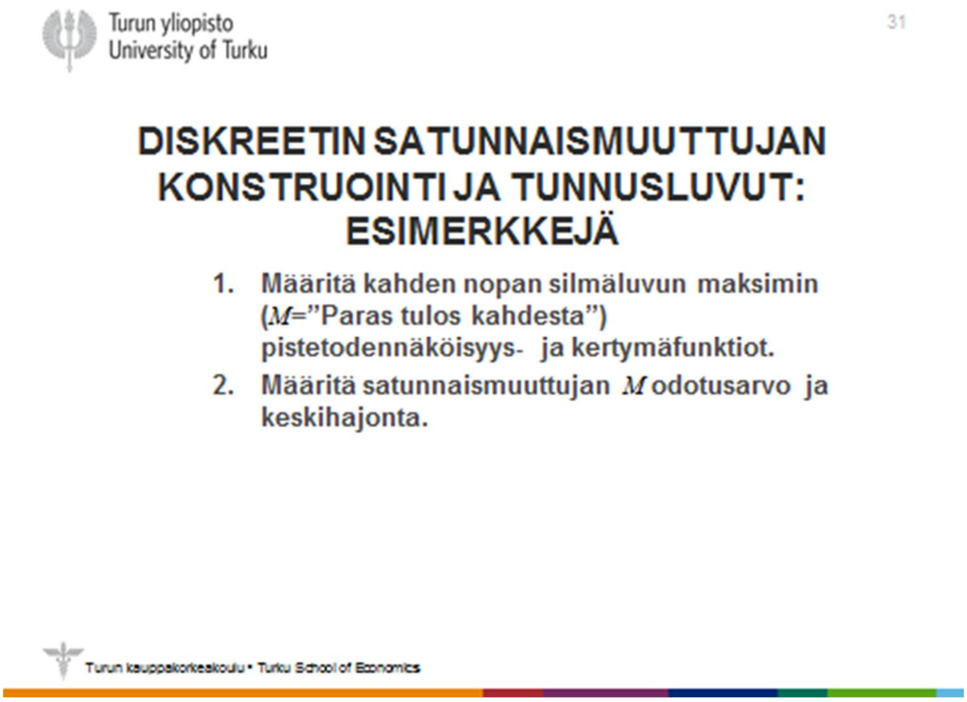
Vaihdetaan osasysteemien kytkentä rinnakkain:



$$R = 1 - R_1^* R_2^* = 1-((1-p^2)(1-(1-(1-p)^3)))$$

Jos komponenttien luotettavuus on 0.99 on systeemin luotettavuus 0,99999998.

Rinnakkaissysteemi on aina luotettavampi kuin sarjasysteemi.



Turun yliopisto
University of Turku

31

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.

Turun kauppakorkeakoulu • Turku School of Economics

Maksimin (kahdesta nopasta) erilaisten alkeistapahtumien taulukko:

Tulos 1 /Tulos 2	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Pistetodennäköisyysfunktio:

M="Paras tulos kahdesta"

$$p(m) = \begin{cases} \frac{1}{36}, kun\ m = 1 \\ \frac{3}{36}, kun\ m = 2 \\ \frac{5}{36}, kun\ m = 3 \\ \frac{7}{36}, kun\ m = 4 \\ \frac{9}{36}, kun\ m = 5 \\ \frac{11}{36}, kun\ m = 6 \\ 0, muulloin \end{cases}$$

Pistetodennäköisyysfunktion arvoille $p(m)$ pätee: $p(m) = P(M = m)$.

Kertymäfunktio:

$$F(m) = \begin{cases} 0, m < 1 \\ \frac{1}{36}, 1 \leq m < 2 \\ \frac{4}{36}, 2 \leq m < 3 \\ \frac{9}{36}, 3 \leq m < 4 \\ \frac{16}{36}, 4 \leq m < 5 \\ \frac{25}{36}, 5 \leq m < 6 \\ 1, m \geq 6 \end{cases}$$

Kertymäfunktion arvoille $F(m)$ pätee: $F(m) = P(M \leq m)$.

Odotusarvo ja keskihajonta:

$$EM = \frac{1}{36} * 1 + \frac{3}{36} * 2 + \frac{5}{36} * 3 + \frac{7}{36} * 4 + \frac{9}{36} * 5 + \frac{11}{36} * 6 \approx 4{,}47$$

Odotettavissa oleva keskimääräinen paras tulos kahdesta nopasta on 4,47.

$$DM = \sqrt{\frac{1}{36} * (1 - 4{,}472)^2 + \frac{3}{36} * (2 - 4{,}472)^2 + \frac{5}{36} * (3 - 4{,}472)^2 + \frac{7}{36} * (4 - 4{,}472)^2 + \frac{9}{36} * (5 - 4{,}472)^2 + \frac{11}{36} * (6 - 4{,}472)^2} \approx 1{,}97$$

Odotettavissa oleva poikkeama odotetusta on 1,97.

Esimerkki: Heitetään rahaa kaksi kertaa ja olkoon *X* = klaavojen lukumäärä. Alkeistapahtumat ja niihin liittyvät satunnaismuuttujan *X* arvot ovat *X*(*kr*,*kr*) = 0, *X*(*kr*,*kl*) = 1, *X*(*kl*,*kr*) = 1, *X*(*kl*,*kl*) = 2.

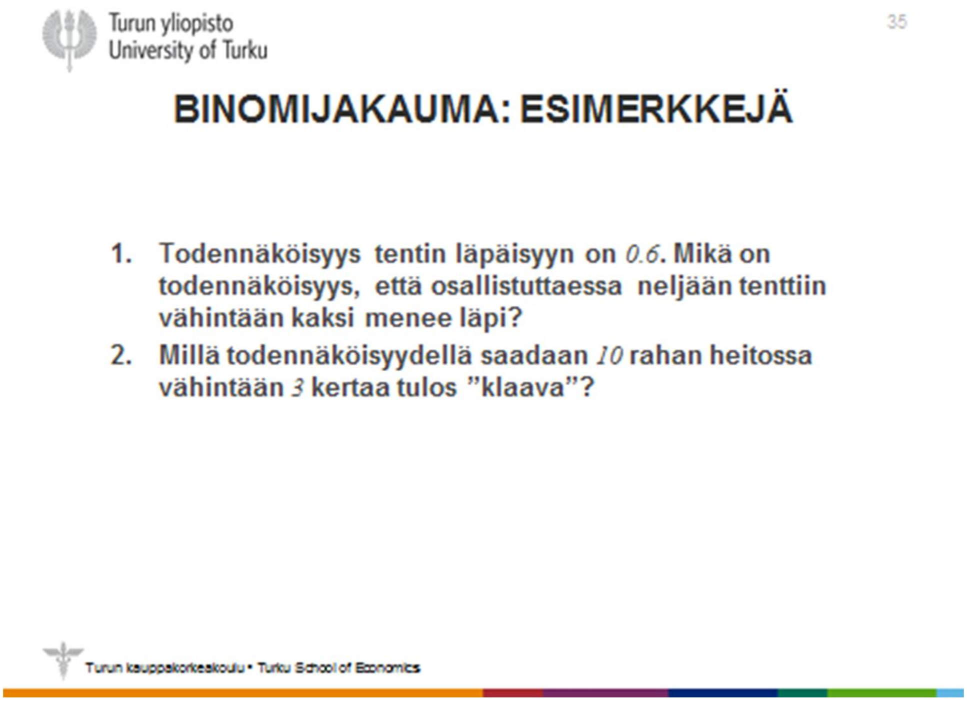
Vastaavat todennäköisyydet ovat siis
P(X = 0) = 1/4
P(X = 1) = 1/2
P(X = 2) = 1/4

josta pistetodennäköisyysfunktio (graafinen esitys pylväs- tai janakuviona):

$$P(x) = \begin{cases} 0.25, \; kun\, x = 0 \\ 0.5, \; kun\, x = 1 \\ 0.25, \; kun\, x = 2 \\ 0, \; muulloin \end{cases}$$

ja kertymäfunktio:

$$F(x) = \begin{cases} 0, \; x < 0 \\ 0.25, \; 0 \le x < 1 \\ 0.75, \; 1 \le x < 2 \\ 1, \; x \ge 2 \end{cases}.$$



X = läpäistyjen tenttien lukumäärä, neljä tenttiä.

$$X \sim Bin(4; 0{,}6)$$

$$P(X = k) = \binom{4}{k} * 0{,}6^k * (1 - 0{,}6)^{4-k}$$

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$$

$$= \binom{4}{2} * 0{,}6^2 * (1 - 0{,}6)^{4-2} + \binom{4}{3} * 0{,}6^3 * (1 - 0{,}6)^{4-3} + \binom{4}{4} * 0{,}6^4 * (1 - 0{,}6)^{4-4}$$

$$= 0{,}3456 + 0{,}3456 + 0{,}1296 \approx 0{,}82$$

X = klaavojen lukumäärä 10 heitossa.

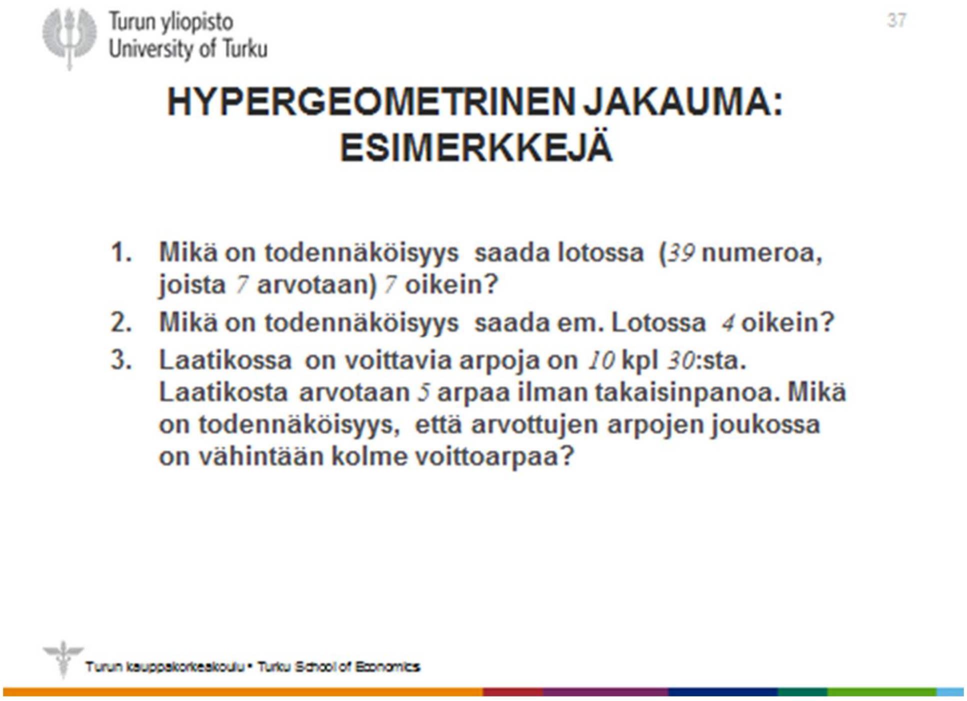
$$X \sim Bin(10; 0{,}5)$$

$$P(X = k) = \binom{10}{k} * 0{,}5^k * (1 - 0{,}5)^{4-k}$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - (P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 1 - \left(\binom{10}{0} * 0{,}5^0 * (1 - 0{,}5)^{10-0} + \binom{10}{1} * 0{,}5^1 * (1 - 0{,}5)^{10-1} + \binom{10}{2} * 0{,}5^2 * (1 - 0{,}5)^{10-2} \right)$$

$$= 1 - (0{,}000977 + 0{,}009766 + 0{,}043945) \approx 0{,}95$$



X=oikeiden lottonumeroiden lukumäärä 7:stä.

$$X \sim Hyperg(39,7,7)$$

$$P(X=7)=\frac{\binom{7}{7}\binom{32}{0}}{\binom{39}{7}} \approx 0{,}000000065 \quad (1/15380937)$$

$$P(X=4)=\frac{\binom{7}{4}\binom{32}{3}}{\binom{39}{7}} \approx 0{,}0113$$

X=voittoarpojen lukumäärä 5:stä.

$$X \sim Hyperg(30,10,5)$$

$$P(X \geq 3)=P(X=3)+P(X=4)+P(X=5)$$

$$=\frac{\binom{10}{3}\binom{20}{2}}{\binom{30}{5}}+\frac{\binom{10}{4}\binom{20}{1}}{\binom{30}{5}}+\frac{\binom{10}{5}\binom{20}{0}}{\binom{30}{5}}$$

$$=0{,}160+0{,}0295+0{,}00177 \approx 0{,}191$$

Binomijakauma

Binomijakaumaa voidaan käyttää sellaisten satunnaisilmiöiden kohdalla, joissa ilmiö toistuu tai toistetaan *n* kertaa ja kustakin toistosta havaitaan, esiintyykö tapahtuma *A* vai ei. Lisäksi oletetaan, että tapahtuman *A* todennäköisyys on sama jokaisessa toistossa ja kysytään, mikä on todennäköisyys sille, että *A* esiintyy tasan *k* kertaa kun ilmiö toistuu *n* kertaa.

Esimerkki: Heitetään arpakuutiota 3 kertaa (*n*=3). Millä todennäköisyydellä saadaan tasan kaksi kertaa kuutonen?

Määritellään satunnaismuuttuja X = kuutosten määrä kolmessa heitossa. Edelleen

$$P(X = 2) = \binom{3}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^1 \approx 0.069$$

Hypergeometrinen jakauma

Binomijakauman tapauksessa oli oleellista, että tapahtuman *A* todennäköisyys pysyi samana jokaisella toistokerralla. Binomijakauman kuvaama tilanne voi esiintyä otannan yhteydessä, jos poimittu yksikkö palautetaan havaintojen teon jälkeen takaisin perusjoukkoon. Jos sama otanta suoritetaankin siten, että poimitut yksiköt jätetään palauttamatta, käytetään ilmiön kuvaamiseen hypergeometrista jakaumaa.

Esimerkki: *(Helenius s.237*) Olkoon 15 tuotteen joukossa 10 virheetöntä ja 5 virheellistä tuotetta. Valitaan tästä joukosta satunnaisesti 3 tuotetta

i) palauttamalla poimittu tuote takaisin ennen seuraavan valintaa
ii) palauttamatta tuotetta

ja määritetään todennäköisyys, että valittujen tuotteiden joukossa on korkeintaan yksi virheellinen. Määritellään satunnaismuuttuja X = virheellisten lukumäärä kolmen tuotteen joukossa.

Tapauksessa i) X on binomijakautunut, parametreina $n = 3$ ja $p = \frac{5}{15} = \frac{1}{3}$. Siis

P(korkeintaan yksi virheellinen) = $P(X \leq 1) = \binom{3}{0}\left(\frac{1}{3}\right)^0 \cdot \left(\frac{2}{3}\right)^3 + \binom{3}{1}\left(\frac{1}{3}\right)^1 \cdot \left(\frac{2}{3}\right)^2 \approx 0.741.$

Tapauksessa ii) X noudattaa hypergeometrista jakaumaa. $P(X=1)$ on

$$\mathrm{P(X=1)} = \frac{\binom{5}{1}\binom{10}{2}}{\binom{15}{3}} = \frac{225}{455}.$$

Todennäköisyys $P(X=0)$ on vastaavasti

$$\mathrm{P(X=0)} = \frac{\binom{5}{0}\binom{10}{3}}{\binom{15}{3}} = \frac{120}{455},$$

joten

$$P(X \leq 1) = P(X=0) + P(X=1) = \frac{120+225}{455} \approx 0.758.$$

Jatkuvat todennäköisyysjakaumat: Normaalijakauma

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

Jossa $\mu = EX$ ja $\sigma^2 = D^2X$.

Jakaumataulukossa 1 on jakauman $Z \sim N(0; 1)$ kertymäfunktion arvoja. Tätä jakaumaa kutsutaan standardoiduksi normaalijakaumaksi ja sitä päästään käyttämään tekemällä muunnos

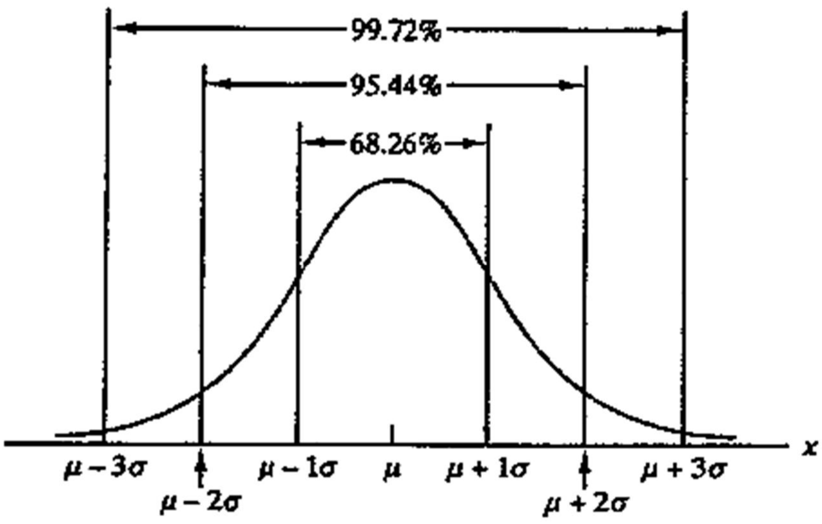
$$Z = \frac{X - \mu}{\sigma}$$

Johtuen normaalijakauman ominaisuuksista

$$P(X \leq a) = P\left(Z \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

Ns. standardoidun normaalijakauman kertymäfunktion tunnuksena käytetään kirjaimen F sijasta kreikkalaista kirjainta Φ.

(Kuvan lähde: Anderson, D.R., Sweeney D.J. ja Williams T.A.: Statistics for Business and Economics)



Esimerkkejä.

$$X \sim N(5{,}0; 0{,}5^2)$$

$$P(X \leq 6{,}2) = P(Z \leq 2{,}40) = \Phi(2{,}40) = 0{,}9918$$

jossa

$$Z = \frac{6{,}2 - 5{,}0}{0{,}5} = 2{,}40$$

$$P(X > 5{,}5) = 1 - P(X \le 5{,}5) = 1 - P(Z \le 1{,}00) = 1 - \Phi(1{,}00) = 1 - 0{,}8413 = 0{,}1587$$

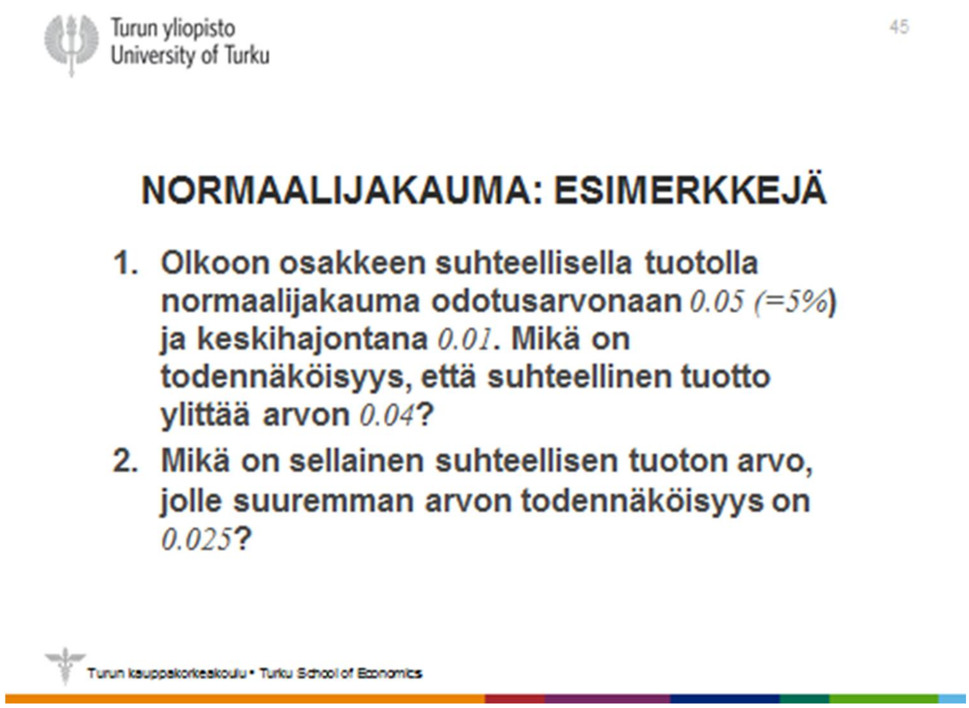
jossa

$$Z = \frac{5{,}5 - 5{,}0}{0{,}5} = 1{,}00$$

$$P(X \le 4{,}2) = P(Z \le -1{,}60) = \Phi(-1{,}60) = 1 - \Phi(1{,}60) = 1 - 0{,}9452 = 0{,}0548$$

jossa

$$Z = \frac{4{,}2 - 5{,}0}{0{,}5} = -1{,}60$$



X = Osakkeen suhteellinen tuotto

1.

$$X \sim N(0{,}05; 0{,}01^2)$$

$$P(X > 0{,}04) = 1 - P(X \leq 0{,}04) = 1 - \Phi\left(\frac{0{,}04 - 0{,}05}{0{,}01}\right)$$

$$= 1 - \Phi(-1{,}00) = 1 - (1 - \Phi(1{,}00)) = \Phi(1{,}00) = 0{,}8413$$

V: Todennäköisyys, että osakkeen suhteellinen tuotto ylittää arvon 0,04 (4%) on 0,8413.

2.

Määritetään a jolle pätee $P(X > a) = 0{,}025$

eli $P(X \leq a) = 0{,}975$

eli $\Phi(z_a) = 0{,}975$

$z_{0{,}025} = 1{,}96$ (eli taulukosta 1: $\Phi(1{,}96) = 0{,}9750$)

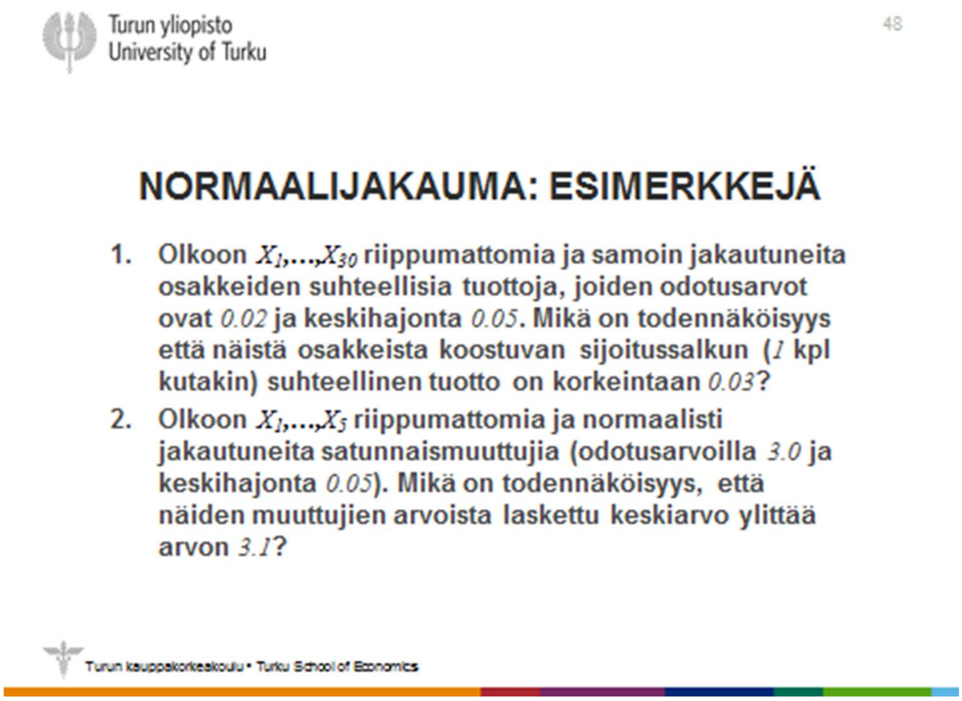
$$\frac{a - 0{,}05}{0.01} = 1{,}96$$

$$a \approx 0{,}0696$$

eli

$$P(X > 0{,}0696) = 0{,}025$$

V: Todennäköisyys, että sijoituksen tuotto ylittää arvon 0,0696 (6,96%) on 0,025.



1. Koska salkussa jokaista sijoitusta 1 kpl, on salkun suhteellinen tuotto sama kuin tuottojen keskiarvo, koska jokaisen paino on 1/30.

$$EX_i = 0{,}02,\; i = 1,\ldots,30$$

$$DX_i = 0{,}05,\; i = 1,\ldots,30$$

joten (n = 30)

$$\bar{X} \sim N\left(0{,}02; \frac{0{,}05^2}{30}\right)$$

$$P(\bar{X} < 0{,}03) = \Phi\left(\frac{0{,}03 - 0{,}02}{0{,}05/\sqrt{30}}\right) = \Phi(1{,}10) = 0{,}8643$$

2.

$$X_i \sim N(3{,}0; 0{,}05^2), i = 1, \dots, 5$$

joten

$$\bar{X} \sim N(3{,}0; \frac{0{,}05^2}{5})$$

$$P(\bar{X} > 3{,}1) = 1 - P(\bar{X} < 3{,}1) = 1 - \Phi\left(\frac{3{,}1 - 3{,}0}{0{,}05/\sqrt{5}}\right) = 1 - \Phi(4{,}47) < 0{,}001$$

Summan todennäköisyysjakauma

Jos

$$X_1 \sim N(EX1; DX1^2)$$

ja

$$X_2 \sim N(EX2; DX2^2)$$

on

$$X_1 + X_2 \sim N(EX1 + EX2; DX1^2 + DX2^2)$$

jos X_1ja X_2keskenään riippumattomia. Ominaisuuden voi yleistää usealle satunnaismuuttujalle.

Esimerkki. Olkoon kaksi riippumatonta satunnaismuuttujaa

$$X_1 \sim N(100; 5^2)$$

ja

$$X_2 \sim N(50; 3^2)$$

Mikä on todennäköisyys, että satunnaismuuttujien summa ylittää arvon 137,4?

$$X_1 + X_2 \sim N(150; 34)$$

$$P(X_1 + X_2 > 137{,}4) = 1 - P(X_1 + X_2 \leq 137{,}4) = 1 - \Phi\left(\frac{137{,}4 - 150}{\sqrt{34}}\right)$$
$$= 1 - \Phi(-2{,}16) = 1 - \big(1 - \Phi(2{,}16)\big) = \Phi(2{,}16) = 0{,}9846$$

Binomijakauman approksimointi normaalijakauman avulla

Jos

$$X \sim Bin(n; p)$$

ja kun n on suuri, lähestyy X:n jakauma

$$X \to Y \sim N(n * p; n * p * (1 - p))$$

Tarkempi tulos approksimoinnille saadaan kun

$$P(a \leq X \leq b) \approx P(a - \tfrac{1}{2} \leq Y \leq b + \tfrac{1}{2})$$

Esimerkki. Heitetään rahaa 300 kertaa. mikä on todennäköisyys, että klaavojen lukumäärä heitoissa on vähintään 131 ja korkeintaan 155?

$$X \sim Bin(300; 0{,}5)$$

eli likimain

$$Y \sim N(150; 75)$$

jossa

$$EY = n * p = 300 * 0{,}5 = 150$$

$$D^2Y = n * p * (1 - P) = 300 * 0{,}5 * 0{,}5 = 75$$

$$\begin{aligned} P(131 \leq X \leq 155) &\approx P(130{,}5 \leq Y \leq 155{,}5) \\ &= \Phi\left(\frac{155{,}5 - 150}{\sqrt{75}}\right) - \Phi\left(\frac{130{,}5 - 150}{\sqrt{75}}\right) = \Phi(0{,}64) - \Phi(-2{,}25) \\ &= \Phi(0{,}64) - \big(1 - \Phi(2{,}25)\big) = 0{,}7389 - (1 - 0{,}9878)) = 0{,}7267 \end{aligned}$$

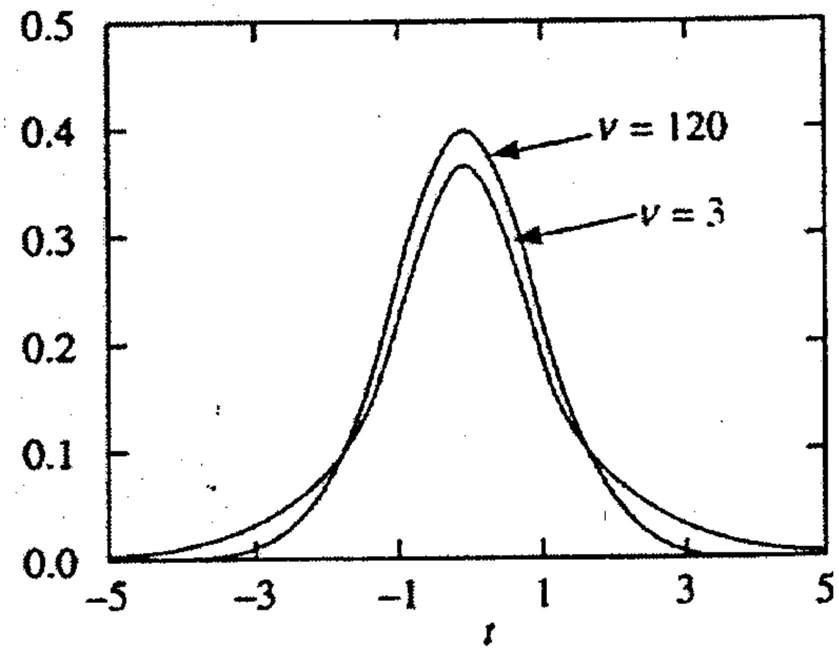
Tarkka arvo binomijakaumaa käyttäen olisi 0,7252, mitta siinä pitäisi laskea kaikki pistetodennäköisyydet yhteen käyttäen X:n arvoja 131-155.

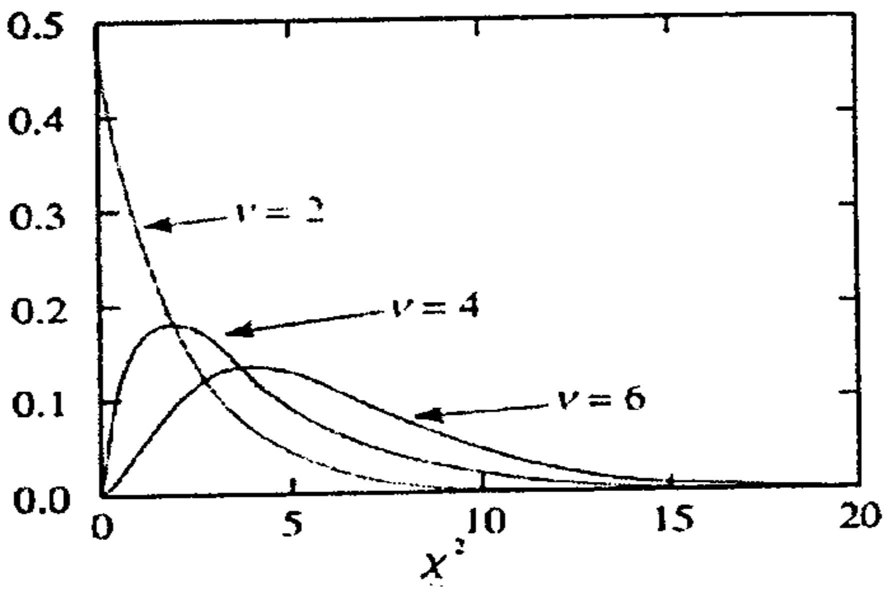
Jos satunnaismuuttujan arvovälillä olisi vain yksi raja, saadaan likimääräinen arvo

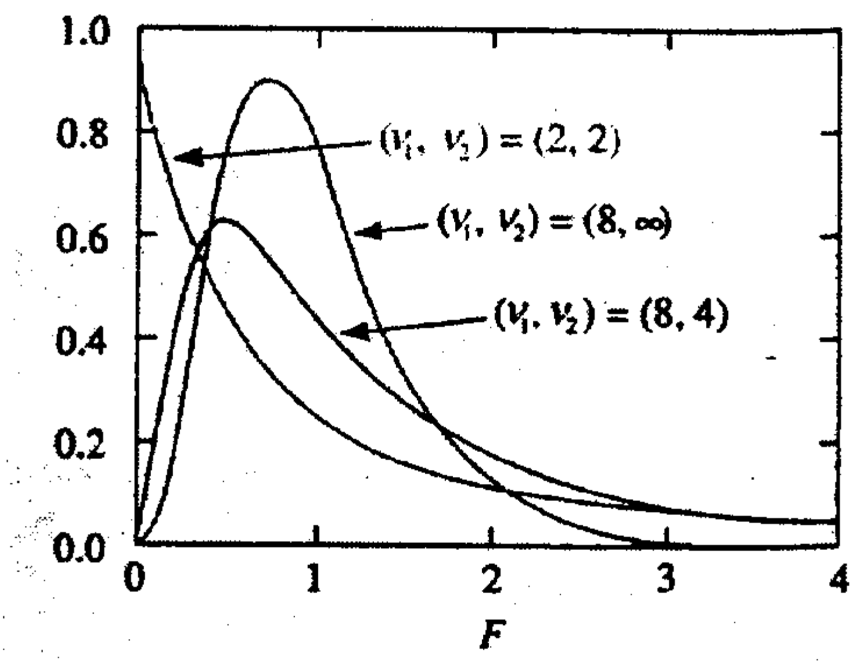
$$P(X \leq a) \approx P\left(Y \leq a + \frac{1}{2}\right)$$

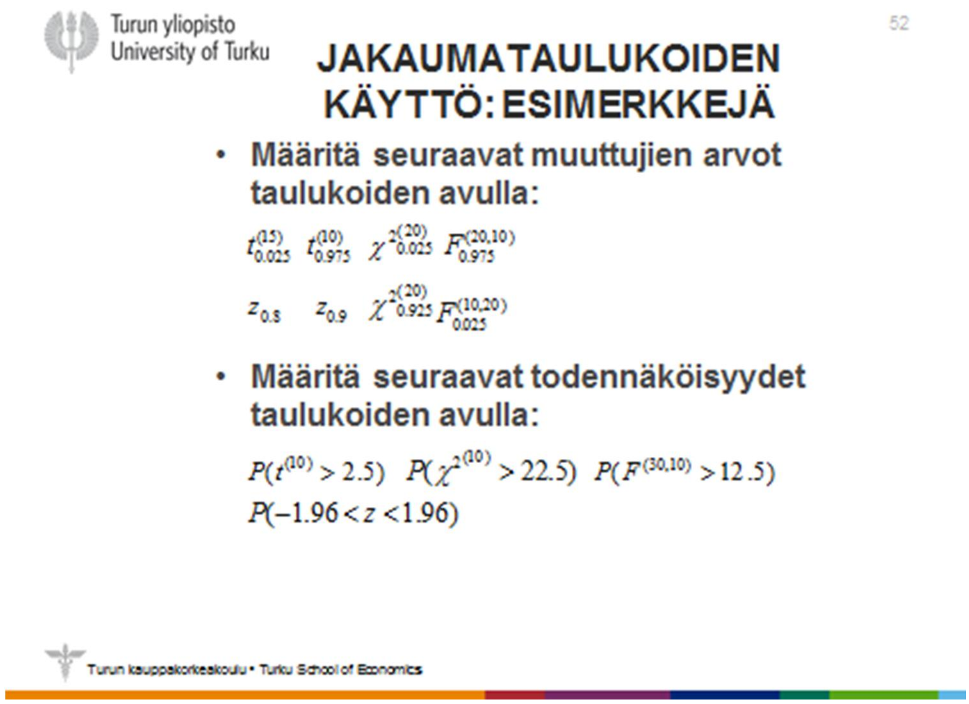
Jatkuvat todennäköisyysjakaumat: Muita jakaumia

(Kuvien lähde: Neter, J. , Wasserman W. ja Whitmore G.: Applied Statistics)









Ylemmän esimerkin kertoimien alaindekseissä oleva todennäköisyys tarkoittaa suuremman arvon todennäköisyyttä eli esim.

$P(Z > z_{0,5}) = 0{,}5$

Alemmissa esimerkeissä satunnaismuuttujien niminä on käytetty niiden jakauman kuvausta.

Kertoimien ominaisuuksia:

$$z_p = -z_{1-p}$$

$$t_p^{(v)} = -t_{1-p}^{(v)}$$

$$F_p^{(v_1,v_2)} = \frac{1}{F_{1-p}^{(v_2,v_1)}}$$

$t_{0,025}^{(15)}$=2,131

$0{,}01 < P\left(t^{(10)} > 2{,}5\right) < 0{,}025$

$t_{0,975}^{(10)} = -t_{0,025}^{(10)} = -2{,}228$

$0{,}01 < P\left(\chi^{2^{(10)}} > 22{,}5\right) < 0{,}025$

${\chi^2}_{0,025}^{(20)} = 34{,}17$

$P\left(F^{(30,10)} > 12{,}5\right) < 0{,}025$

$F_{0,975}^{(20,10)} = \frac{1}{F_{0,025}^{(10,20)}} = \frac{1}{2{,}77}$

$$P(-1{,}96 < z < 1{,}96)) = \Phi(1{,}96) - \Phi(-1{,}96) = \Phi(1{,}96) - \left(1 - \Phi(1{,}96)\right)$$
$$= 0{,}9750 - (1 - 0{,}9750) = 0{,}95$$

$z_{0,5} = 0$

$z_{0,9} = -z_{0,1} = -1{,}28$

$10{,}85 < {\chi^2}_{0,925}^{(20)} < 28{,}41$

$F_{0,025}^{(10,20)} = 2{,}77$

Lisäesimerkkejä:

$t_{0,000001}^{(15)} > 4{,}073$

$P\left(t^{(10)} > 6{,}2\right) < 0{,}0005$

$P\left(t^{(10)} > 1{,}2\right) > 0{,}1$

$2{,}021 < t_{0,025}^{(35)} < 2{,}042$

Tilastollinen päättely, väliestimointi

Populaatiokeskiarvon luottamusvälin kaavat:

Pieni otos (n<30), normaalijakaumaoletus tarvitaan

$$V_{100(1-\alpha)} = [\bar{x} - t_{\alpha/2}^{(n-1)} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}^{(n-1)} \frac{s}{\sqrt{n}}]$$

Suuri otos

$$V_{100(1-\alpha)} = [\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}]$$

Esimerkkejä:

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.

1. Tulkinta: Populaation keski-ikä on välillä 31,3-33,3 vuotta (5% virheen riski)

$$V_{100(1-\alpha)} = \left[\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}\right] = V_{95} = \left[32{,}3 - 1{,}96 \frac{3{,}3}{\sqrt{40}}; 32{,}3 + 1{,}96 \frac{3{,}3}{\sqrt{40}}\right]$$
$$= [31{,}28;\ 33{,}32]$$

2. Tulkinta: Tuotteen keskipaino on välillä 10,1-10,3 grammaa (5% virheen riski)

$$V_{100(1-\alpha)} = \left[\bar{x} - t_{\alpha/2}^{(n-1)} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}^{(n-1)} \frac{s}{\sqrt{n}}\right] = V_{95} = \left[10{,}2 - 2{,}145 \frac{0{,}2}{\sqrt{15}}; 10{,}2 + 2{,}145 \frac{0{,}2}{\sqrt{15}}\right]$$
$$= [10{,}09;\ 10{,}31]$$

3. Tulkinta: Tuotteen keskipaino on välillä 10,1-10,4 grammaa (1% virheen riski)

$$V_{100(1-\alpha)} = \left[\bar{x} - t_{\alpha/2}^{(n-1)} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}^{(n-1)} \frac{s}{\sqrt{n}}\right] = V_{99} = \left[10{,}2 - 2{,}977 \frac{0{,}2}{\sqrt{15}}; 10{,}2 + 2{,}977 \frac{0{,}2}{\sqrt{15}}\right]$$
$$= [10{,}05;\ 10{,}35]$$

SPSS-tuloste (Eri esimerkki):

			Statistic	Std. Error
Turnover	Mean		12.8354	1.42588
	95% Confidence Interval for Mean	Lower Bound	9.8701	
		Upper Bound	15.8007	

Populaatiokeskiarvon luottamusvälin ala- ja yläraja, 5% virheen riski.

Tulkinta: Keskimääräinen liikevaihto yritysten populaatiossa 9,9-15,8 M€ (5% virheen riski).

Tilastollinen päättely, merkitsevyystestaus (populaatiokeskiarvo)

Tarvittavat kaavat:

Pieni otos (n<30), normaalijakaumaoletus tarvitaan

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

Suuri otos

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

Havaitun merkitsevyystason laskeminen:

$$p = 2 \cdot P(t > |t_{hav}|)$$

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

Esimerkkejä:

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.

1. Otoskoko suuri→z-testi. H_0: μ=32; H_1:μ≠32.

$$z_{hav} = \frac{28,2-32}{1,2/\sqrt{32}} \approx -17,91$$

p-arvo (eli havaittu merkitsevyystaso):

$$p = 2\cdot(1-\Phi(17,91)) < 0,002 < 0,05 = \alpha$$

→H_0 hylätään, koska p<0,05. Tulkinta: Aineisto tuki olettamusta, jonka mukaan populaation keski-ikä ei ole 32 vuotta. 95% luottamusväli:

$$V_{95} = [28,2-1,96\frac{1,2}{\sqrt{32}};28,2+1,96\frac{1,2}{\sqrt{32}}] \approx [27,8;28,6]$$

Populaation keski-ikä on 27,8-28,6 vuotta (5% virheen riski).

2. Otoskoko pieni, normaalijakaumaoletus annettu→t-testi. H_0: μ=160; H_1:μ≠160.

$$t_{hav} = \frac{164,0-160}{6,0/\sqrt{16}} \approx 2,67$$

p-arvo (eli havaittu merkitsevyystaso):

$$0,01 < p = 2\cdot P(t > 2,67) < 0,02 < 0,05 = \alpha$$

→H_0 hylätään, koska p<0,05. Tulkinta: Aineisto tuki olettamusta, jonka mukaan tuotannon keskipaino ei ole 160g vuotta. 95% luottamusväli:

$$V_{95} = \left[164 - 2,145 * \frac{6,0}{\sqrt{16}}; 164 + 2,145 * \frac{6,0}{\sqrt{16}}\right] \approx [160,78; 167,22]$$

Tuotannon keskipaino 160,8-167,2 g (5% virheen riski).

SPSS-tuloste (Eri esimerkki):

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
etäisyys kotoa Turun keskustaan (km)	1623	8,0480	7,72471	,19174

One-Sample Test

	Test Value = 10					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
etäisyys kotoa Turun keskustaan (km)	-10,180	1622	,000	-1,95197	-2,3281	-1,5759

Testisuureen arvo

p-arvo

Tässä nollahypoteesi H_0: $\mu=10$ hylätään, koska p-arvo<0,05. Aineisto tuki väittämää, jonka mukaan alueen kotitalouksien keskimääräinen etäisyys Turun keskustaan ei ole 10 km.

Kahden riippumattoman otoksen keskiarvotesti

$$H_0: \mu_1 = \mu_2$$

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

Kolme ”versiota”:

(1) σ_1 ja σ_2 tunnettuja

<u>(2) σ_1 ja σ_2 tuntemattomia, mutta oletetaan $\sigma_1=\sigma_2$</u>

Pieni otos (n<30), normaalijakaumaoletus tarvitaan

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

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

Suuri otos

$$z_{hav} = \frac{\overline{X}_1 - \overline{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

Populaatiokeskiarvojen eron luottamusvälin kaavat:

Pieni otos (n<30), normaalijakaumaoletus tarvitaan

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

Suuri otos

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

<u>(3) σ_1 ja σ_2 tuntemattomia, mutta oletetaan $\sigma_1 \neq \sigma_2$</u>

Pieni otos (20<n<50), normaalijakaumaoletus tarvitaan

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

$$\frac{1}{v}=\frac{c^2}{n_1-1}+\frac{(1-c)^2}{n_2-1}$$

$$c=\frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Suuri otos

$$z_{hav}=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\sim \mathrm{N}(0,1)$$

Populaatiokeskiarvojen eron luottamusvälin kaavat:

Pieni otos (20<n<50), normaalijakaumaoletus tarvitaan

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

Suuri otos

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

Esimerkki.

Kaksi eri tuotantolinjaa, A ja B, tuottivat tietokoneen osia, joiden piti olla keskimäärin samanpainoisia (grammoissa). 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.

$$s=\sqrt{\frac{(35-1)*0{,}2^2+(34-1)*0{,}3^2}{35+34-2}}\approx 0{,}254$$

$$z_{hav}=\frac{12{,}5-12{,}1}{0{,}254*\sqrt{\frac{1}{35}+\frac{1}{34}}}\approx 6{,}53$$

Havaittu merkitsevyystaso:

$$p = 2 \cdot (1 - \Phi(6{,}53)) < 0{,}002 < 0{,}05 = \alpha$$

Tässä nollahypoteesi $H_0: \mu_1 = \mu_2$ hylätään, koska p<0,05. Aineisto tuki väittämää, jonka mukaan tuotantolinjoilta ei tule keskimäärin samanpainoisia tuotteita.

Populaatiokeskiarvojen eron 95% luottamusväli:

$$V_{100(1-\alpha)} = \left[12{,}5 - 12{,}1 - 1{,}96 * 0{,}254 * \sqrt{\frac{1}{35} + \frac{1}{34}}; 12{,}5 - 12{,}1 + 1{,}96 * 0{,}254 * \sqrt{\frac{1}{35} + \frac{1}{34}}\right] \approx [0{,}280; 0{,}520]$$

Tuotannon keskipainojen ero 0,28-0,52 grammaa (5% virheen riski).

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.

$$c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \approx 0{,}2$$

$$\frac{1}{v} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1} \approx 0.02 \rightarrow v = 50$$

$$t_{hav} = \frac{3{,}45 - 3{,}4}{\sqrt{\frac{0{,}1^2}{35} + \frac{0{,}2^2}{35}}} \approx 1{,}32$$

Havaittu merkitsevyystaso:

$$\alpha = 0{,}05 < 0{,}1 < p = 2 \cdot P(t > 1{,}32) < 0{,}2$$

Tässä nollahypoteesi $H_0: \mu_1 = \mu_2$ hyväksytään, koska p>0,05. Aineisto tuki väittämää, jonka mukaan miehet ja naiset asiakaskunnassa keskimäärin yhtä tyytyväisiä.

Populaatiokeskiarvojen eron 95% luottamusväli:

$$V_{100(1-\alpha)} = \left[3{,}45 - 3{,}4 - (\frac{2{,}021 + 2{,}000}{2})\sqrt{\frac{0{,}1^2}{35} + \frac{0{,}2^2}{35}}, 3{,}45 - 3{,}4 + (\frac{2{,}021 + 2{,}000}{2})\sqrt{\frac{0{,}1^2}{35} + \frac{0{,}2^2}{35}}\right]$$
$$\approx [-0{,}026; 0{,}126]$$

Miesten ja naisten asiakastyytyväisyyksien keskiarvojen ero välillä -0,03 +0,13 (5% virheen riski).

SPSS-tuloste (Eri esimerkki):

	Autojen lukumäärä taloudessa	N	Mean	Std. Deviation	Std. Error Mean
etäisyys kotoa Länsikeskukseen (km)	Ei	240	7,1321	4,93132	,31832
	Yksi tai enemmän	1140	10,1496	7,50259	,22221

Alla olevassa taulukossa on esitetty kahden riippumattoman otoksen keskiarvotestin versiot 2 ja 3.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
etäisyys kotoa Länsikeskukseen (km)	Equal variances assumed	56,303	,000	-5,965	1378	,000	-3,01752	,50591	-4,00996	-2,02509
	Equal variances not assumed			-7,773	503,594	,000	-3,01752	,38820	-3,78022	-2,25483

Testisuureen arvot (versiot 2 ja 3)

p-arvot (versiot 2 ja 3)

populaatiokeskiarvojen eron luottamusväli (versiot 2 ja 3)

Kumpi versio tulee valita, 2 vai 3?

Taulukon vasemmassa reunassa esitetään kahden populaatiovarianssin vertailun testin tulokset. Testissä

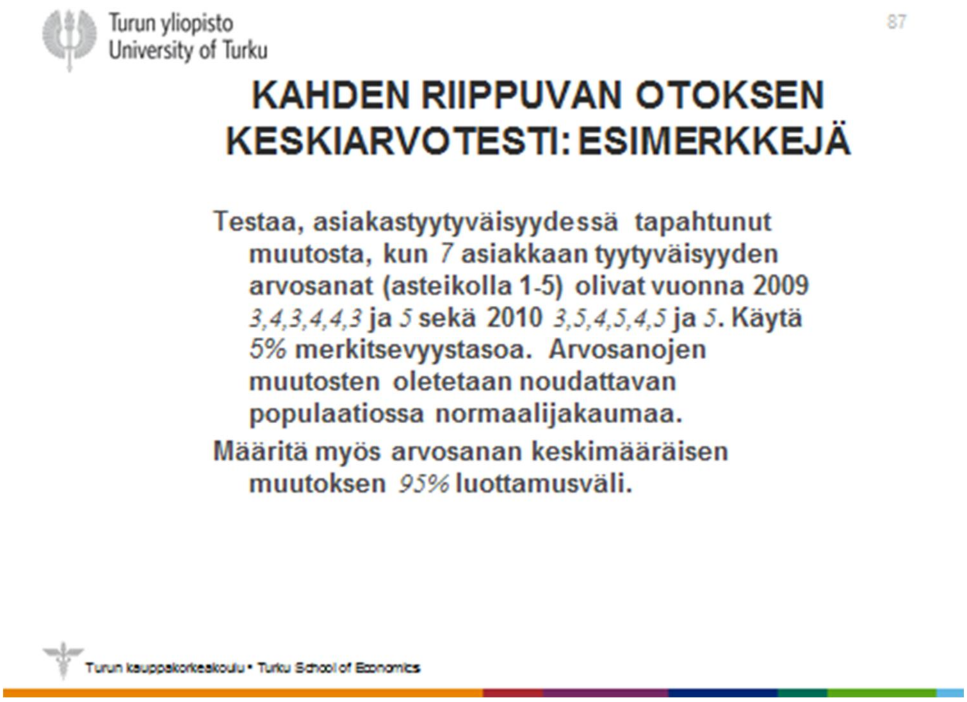
$$H_0: \sigma_1^2 = \sigma_1^2$$

Kyseisen testin p-arvo on <0,05, joten aineisto ei tue väitettä yhtä suurista hajonnoista. Näin ollen taulukosta valitaan keskiarvojen vertailuun alempi rivi (versio 3), vrt. ”Equal variances not assumed”. Tämän testin suorittaminen esitetään myöhemmin.

Keskiarvotestissä nollahypoteesi

$$H_0: \mu_1 = \mu_2$$

tässä hylätään koska p<0.05 eli keskimääräiset etäisyydet kahdessa autoilukategoriassa voidaan olettaa erilaisiksi. Keskiarvojen eron 95%-luottamusväli on [-3,78;-2,25]. Negatiiviset arvot tarkoittavat sitä, että talouksissa joissa on yksi auto tai enemmän on etäisyys Länsikeskukseen tuon verran enemmän (km) verrattuna autottomiin talouksiin.



$$H_0 : \mu_D = 0$$

$$H_0 : \mu_D \neq 0$$

2009	3	4	3	4	4	3	5
2010	3	5	4	5	4	5	5
D	0	-1	-1	-1	0	-2	0

$$\overline{D} = \frac{-5}{7} \approx -0{,}71$$

$$s_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}} \approx 0{,}76$$

$$t_{hav} = \frac{-0{,}71}{0{,}76/\sqrt{7}} \approx -2{,}50$$

$$p = 2 * P(t > -2{,}50) = 2 * P(t > 2{,}50)$$

, eli tällaisissa testeissä testisuureen etumerkillä ei ole väliä.

$$v = n - 1 = 7 - 1 = 6$$

$$0{,}02 < 2 * P(t > 2{,}50) < 0{,}05 = \alpha$$

eli H_0 hylätään. Aineisto tuki väittämää, jonka mukaan muutosta on tapahtunut asiakastyytyväisyydessä. Muutoksen 95% luottamusväli:

$$V_{95} = \left[-0{,}71 - 2{,}447 * \frac{0{,}76}{\sqrt{7}}; -0{,}71 + 2{,}447 * \frac{0{,}76}{\sqrt{7}}\right] \approx [-1{,}41; -0{,}007]$$

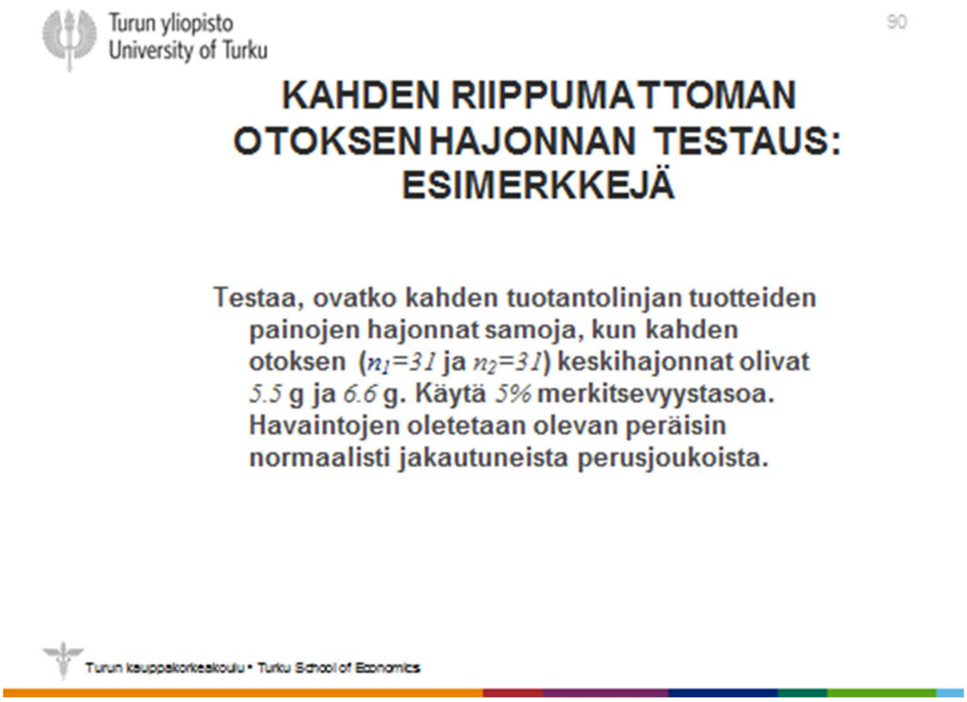
Asiakastyytyväisyyden keskimääräinen muutos asiakaskunnassa 0,007-1,41 pistettä parannusta (5% virheen riski).

Kahden riippuvan otoksen keskiarvotesti

SPSS-tuloste (eri esimerkki)

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	Weight at Start (kg) - Weight after 2 months (kg)	,98750	12,08293	3,82096	-7,65610	9,63111	,258	9	,802

Tässä muutos ei tilastollisesti merkitsevä (p=0,801>0,05).



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

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

Normaalijakaumaoletukset→F-testi

$$F_{hav} = \frac{s_1^2}{s_2^2} = \frac{5{,}5^2}{6{,}6^2} \approx 0{,}69$$

$$v_1 = n_1 - 1 = 31 - 1 = 30$$

$$v_2 = n_2 - 1 = 31 - 1 = 30$$

Määritetään p-arvo käyttäen ns. kriittisen arvon menetelmää eli määritellään ne kaksi testisuureen arvoa, jotka tuottaisivat p-arvon tasan 0,05:

$$F_{0{,}025}^{(30;30)} = 2{,}07$$

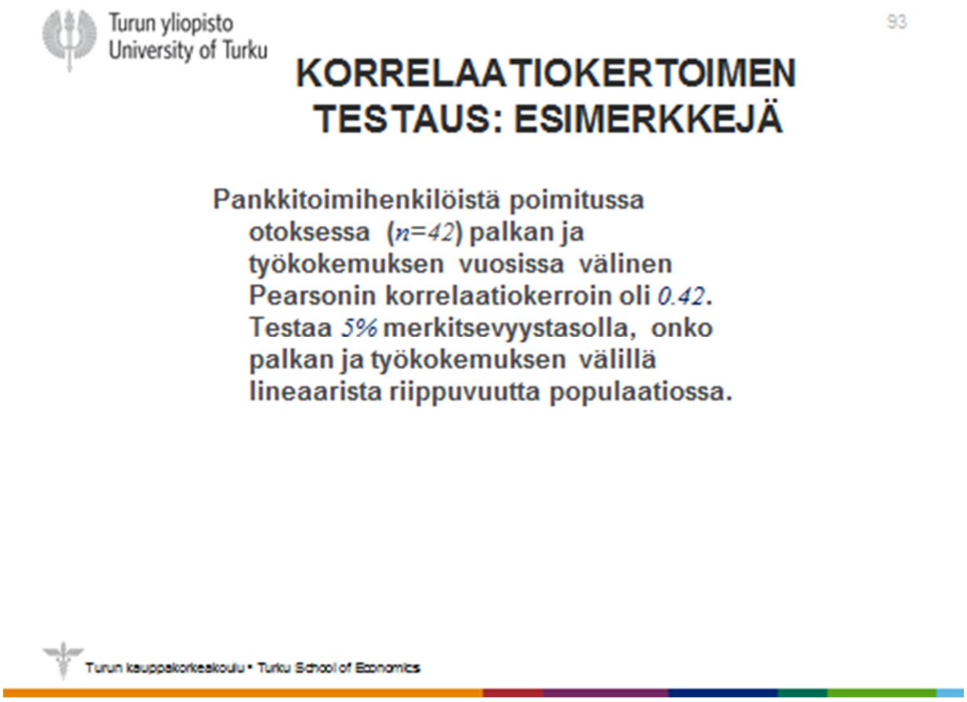
$$F_{0{,}975}^{(30;30)} = \frac{1}{F_{0{,}025}^{(30;30)}} = \frac{1}{2{,}07} \approx 0{,}48$$

Jos $F_{0{,}975}^{(30;30)} < F_{hav} < F_{0{,}025}^{(30;30)}$ on p-arvo automaattisesti suurempi kuin 0,05, muuten pienempi kuin 0,05. Tässä

$$0{,}48 < 0{,}69 < 2{,}07$$

joten p>0,05→H_0 hyväksytään. Aineisto tuki väittämää, jonka mukaan populaatiohajonnat (kahden tuotantolinjan tuotteiden painojen) ovat samoja (p>0,05).

SPSS-tuloste: Ks. aikaisempi kahden riippumattoman otoksen keskiarvotestin tuloste.



$$H_0: \rho = 0$$

$$H_0: \rho \neq 0$$

Normaalijakaumaoletus→t-testi. Testisuure

$$t_{hav} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}42 * \sqrt{42-2}}{\sqrt{1-0{,}42^2}} \approx 2{,}93$$

$$v = n - 2 = 42 - 2 = 40$$

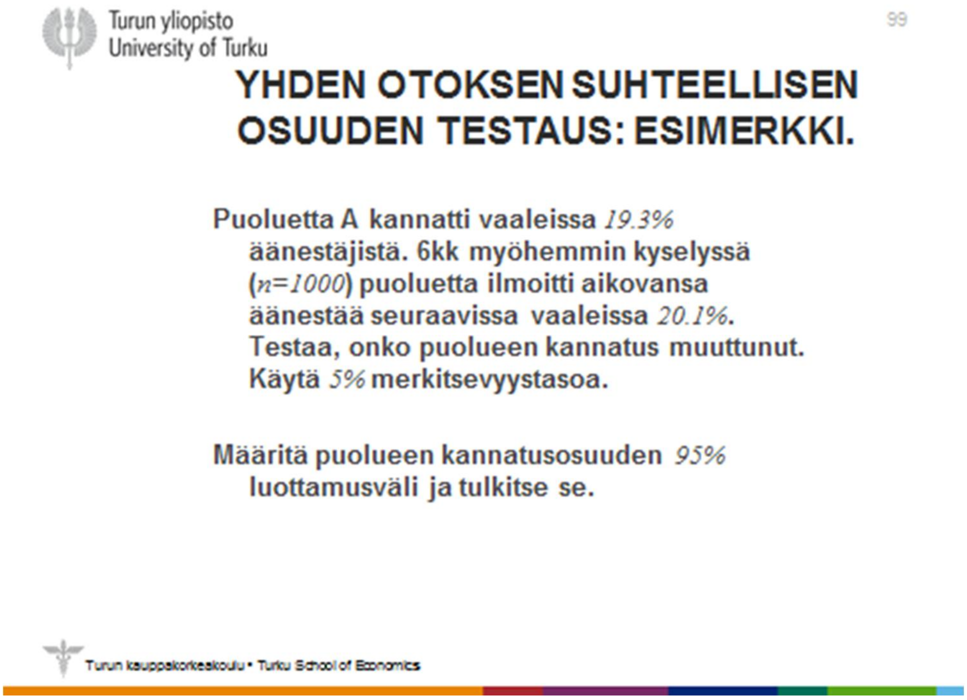
p-arvo

$$0{,}002 < 2 * P(t > 2{,}93) < 0{,}01 < 0{,}05 = \alpha$$

H_0 hylätään. Aineisto tuki väittämää, jonka mukaan pankkitoimihenkilöillä on lineaarista korrelaatiota palkan ja työkokemuksen vuosissa välillä (p<0,05). SPSS-tuloste (eri esimerkki).

		Age (Years)	Salary per Hour (£)
Age (Years)	Pearson Correlation	1	,397**
	Sig. (2-tailed)		,000
	N	231	231
Salary per Hour (£)	Pearson Correlation	,397**	1
	Sig. (2-tailed)	,000	
	N	231	231

**. Correlation is significant at the 0.01 level (2-tailed).



$$H_0\!:\pi = 0{,}193$$

$$H_0\!:\pi \neq 0{,}193$$

$$n\pi_0 > 5$$

$$n(1-\pi_0) > 5$$

toteutuu

$$1000 * 0{,}193 = 193 > 5$$

$$1000 * (1 - 0{,}193) = 807 > 5$$

Eli approksimaation tarkkuus riittävä ts. testi voidaan suorittaa z-testinä.

$$z_{hav} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}201 - 0{,}193}{\sqrt{\frac{0{,}193(1-0{,}193)}{1000}}} \approx 0{,}64$$

$$p = 2 * (1 - \Phi(z_{hav}) = 2 * (1 - 0{,}7389) \approx 0{,}52 > 0{,}05 = \alpha$$

H_0 hyväksytään. Aineisto tuki väittämää, jonka mukaan puolueen kannatusosuus on edelleen 19,3%. Luottamusväli kannatusosuudelle

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

$$V_{95} = \left[0{,}201 - 1{,}96 * \sqrt{\frac{0{,}201 * (1 - 0{,}201)}{1000}}; 0{,}201 + 1{,}96 * \sqrt{\frac{0{,}201 * (1 - 0{,}201)}{1000}} \right]$$
$$\approx [0{,}176; 0{,}225]$$

Puolueen kannatusosuus äänestäjäkunnassa 17,6-22,5% (5% virheen riski).

(Ei löydy sellaisenaan SPSSstä).

Turun yliopisto
University of Turku

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.

$$H_0: \pi_1 = \pi_2$$

$$H_0: \pi_1 \neq \pi_2$$

Approksimoinnin hyvyyden kriteerit toteutuvat, joten testi voidaan tehdä z-testinä:

$$n_1 p_1 = 32 > 5$$
$$n_1(1 - p_{1)} = 68 > 5$$
$$n_2 p_2 = 39 > 5$$
$$n_2(1 - p_2) = 61 > 5$$

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

$$p = \frac{100 * 0{,}32 + 100 * 0{,}39}{100 + 100} = 0{,}355$$

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

$$z_{hav} = \frac{0{,}32 - 0{,}39}{\sqrt{0{,}355 * (1 - 0{,}355) * \left(\frac{1}{100} + \frac{1}{100}\right)}} \approx -1{,}03$$

$$p = 2 * (1 - 0{,}8485) = 0{,}303 > 0{,}05 = \alpha$$

H_0 hyväksytään. Aineisto tuki väittämää, jonka mukaan tuotemerkin tunnistavien osuuksissa ei ole ikäryhmissä eroa ts. ikäryhmä ei vaikuta tuotemerkin tunnistamiseen. Tuotemerkin tunnistavien osuuksien eron 95% luottamusväli

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

$$V_{95} = \left[0{,}32 - 0{,}39 \pm 1{,}96 * \sqrt{\frac{0{,}32 * (1 - 0{,}32)}{100} + \frac{0{,}39 * (1 - 0{,}39)}{100}} \right] \approx [-0{,}203; 0{,}062]$$

Tuotemerkin tunnistavien osuuksien ero ikäryhmien välillä on välillä -20,3% (vanhemmassa ikäryhmässä osuus suurempi) +6,2% (nuoremmassa ikäryhmässä osuus suurempi) (5% virheen riski).

(Ei löydy sellaisenaan SPSSstä).

Turun yliopisto
University of Turku

105

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

Turun kauppakorkeakoulu • Turku School of Economics

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

	Vastaus 1	*Vastaus 2*
Vastaus 1	$f_{11}=a=30$	$f_{12}=b=15$
Vastaus 2	$f_{21}=c=9$	$f_{22}=d=51$

$$z_{hav} \frac{15-9/105}{\sqrt{\dfrac{(15+9)-(15-9)^2/105}{105*(105-1)}}} \approx 1{,}23$$

$$p = 2*(1-0{,}8907) \approx 0{,}22 > 0{,}05 = \alpha$$

H_0 hyväksytään. Aineisto tuki väittämää, jonka mukaan vastauksien osuudet eivät ole muuttuneet (p>0,05). Luottamusväli

$$V_{100(1-\alpha)} = \left[(b-c)/n \pm z_{\alpha/2}\sqrt{\frac{(b+c)-(b-c)^2/n}{n(n-1)}} \right]$$

$$V_{95} = \left[(15-9)/105 \pm 1{,}96 * \sqrt{\frac{(15+9)-(15-9)^2/105}{105*(105-1)}} \right] \approx [-0{,}034; 0{,}148]$$

Vastaus 2:n osuuden muutos perusjoukossa välillä 3,4%-yksikköä laskua ja 14,8% nousua (5% virheen riski).

(Ei löydy sellaisenaan SPSSstä).

Yhden otoksen likimääräinen suhteellisen osuuden testi SPSSllä (luentoesimerkki)

Tehdään yhden otoksen keskiarvotestinä

Vaatimukset: Muuttujan arvot koodattu 0 ja 1

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Aanestaa	1000	,20	,401	,013

One-Sample Test

	Test Value = 0.193					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Aanestaa	,631	999	,528	,008	-,02	,03

Taulukosta saa suhteellisen osuuden 95% luottamusvälin [0,193-0,02; 0,193+0,03]=[0,173; 0,203].

Yhden otoksen suhteellisen osuuden testi (z-testi): Testin p-arvo on 0,52 ja suhteellisen osuuden 95% luottamusväli [0,176; 0,225].

Toinen esimerkki

Frequencies

Statistics

Konkurssi

N	Valid	39
	Missing	0

Konkurssi

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ei	18	46,2	46,2	46,2
	Kyllä	21	53,8	53,8	100,0
	Total	39	100,0	100,0	

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Konkurssi	39	,54	,505	,081

One-Sample Test

	Test Value = 0.5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Konkurssi	,476	38	,637	,038	-,13	,20

Taulukosta saa suhteellisen osuuden 95% luottamusvälin [0,5-0,13; 0,5+0,20]=[0,37; 0,70].

Yhden otoksen suhteellisen osuuden testi (z-testi): Testin p-arvo on 0,64 ja suhteellisen osuuden 95% luottamusväli [0,38; 0,69].

Kahden riippumattoman otoksen likimääräinen suhteellisen osuuden testi SPSSllä (luentoesimerkki)

Tehdään kahden riippumattoman otoksen keskiarvotestinä

Vaatimukset: Muuttujan arvot koodattu 0 ja 1

T-Test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TUNNISTI	Equal variances assumed	4,109	,044	-1,032	198	,303	-,070	,068	-,204	,064
	Equal variances not assumed			-1,032	197,608	,303	-,070	,068	-,204	,064

Kahden riippumattoman otoksen suhteellisen osuuden testi (z-testi): Testin p-arvo on 0,303 ja suhteellisen osuuksien eron 95% luottamusväli [-0,203; 0,062].

Toinen esimerkki

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Konkurssi * Omavaraisuusaste >50%	39	100,0%	0	0,0%	39	100,0%

Konkurssi * Omavaraisuusaste >50% Crosstabulation

			Omavaraisuusaste >50%		Total
			Ei	Kyllä	
Konkurssi	Ei	Count	10	8	18
		% within Omavaraisuusaste >50%	62,5%	34,8%	46,2%
	Kyllä	Count	6	15	21
		% within Omavaraisuusaste >50%	37,5%	65,2%	53,8%
Total		Count	16	23	39
		% within Omavaraisuusaste >50%	100,0%	100,0%	100,0%

T-Test

Group Statistics

	Omavaraisuusaste >50%	N	Mean	Std. Deviation	Std. Error Mean
Konkurssi	Ei	16	,38	,500	,125
	Kyllä	23	,65	,487	,102

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Konkurssi	Equal variances assumed	,110	,742	-1,729	37	,092	-,277	,160	-,602	,048
	Equal variances not assumed			-1,721	31,867	,095	-,277	,161	-,605	,051

Kahden riippumattoman otoksen suhteellisen osuuden testi (z-testi): Testin p-arvo on 0,09 ja suhteellisen osuuksien eron 95% luottamusväli [-0,58; 0,03].

Kahden riippuvan otoksen likimääräinen suhteellisen osuuden testi (luentoesimerkki) SPSSllä

Tehdään kahden riippuvan otoksen keskiarvotestinä

Vaatimukset: Muuttujan arvot koodattu 0 ja 1

Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Vastaus1	,63	105	,486	,047
	Vastaus2	,57	105	,497	,049

Paired Samples Correlations

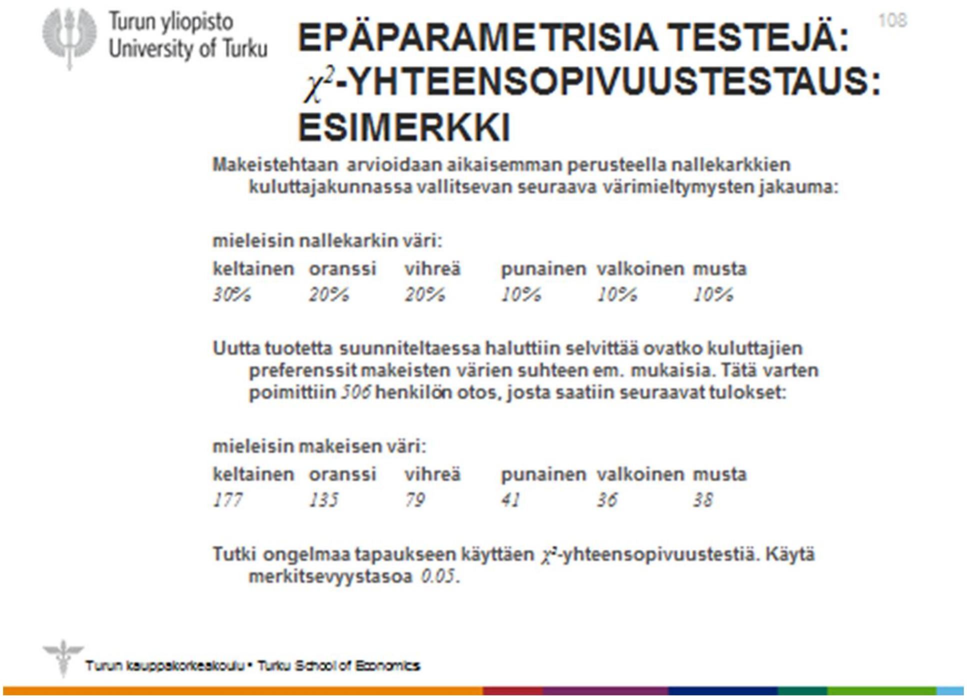
		N	Correlation	Sig.
Pair 1	Vastaus1 & Vastaus2	105	,529	,000

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Vastaus1 - Vastaus2	,057	,477	,047	-,035	,149	1,228	104	,222

Kahden riippuvan otoksen suhteellisen osuuden testi (z-testi): Testin p-arvo on 0,22 ja suhteellisen osuuden muutoksen 95% luottamusväli [-0,034; 0,148].

Kyseiset testit saa tehtyä SPSS:llä myös tarkkoina toista ”kiertotietä”, toinen tapa esitellään kurssilla myöhemmin.



Hypoteesit:

H_0: Havaintoaineisto on peräisin väitettyä suhteellista jakaumaa noudattavasta populaatiosta (ks. ensimmäinen taulukko).

H_1: Havaintoaineisto ei ole peräisin väitettyä suhteellista jakaumaa noudattavasta populaatiosta.

Testissä verrataan otoksesta mitattuja eli k-luokkaisen muuttujan havaittuja frekvenssejä *f* nollahypoteesissa oletetun suhteellisen jakauman mukaisiin eli odotettuihin frekvensseihin *e*. Tässä *k*=6. Odotetut frekvenssit *e*:

mieleisin makeisen väri

keltainen	oranssi	vihreä	punainen	valkoinen	musta
e_1	e_2	e_3	e_4	e_5	e_6
151.8	101.2	101.2	50.6	50.6	50.6

Eli nollahypoteesin mukainen suhteellinen jakauma otoksessa, jonka koko on 506 havaintoa. Tässä esim. 0.3·506 = 151.8.

Kriteerit approksimaation hyvyyden tarkasteluun: Jokaisen odotetun frekvenssin tulee olla >1 ja korkeintaan 20% odotetuista frekvensseistä saa olla <5. Ehdot toteutuvat tässä, koska jokainen odotettu frekvenssi e on suurempi kuin 5→ tehdään χ2-testi.

Testisuureen kaava ja jakauma:

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

Testisuureen arvo:

$$\chi^2_{hav} = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = \frac{(177-151.8)^2}{151.8} + \frac{(135-101.2)^2}{101.2} + \frac{(79-101.2)^2}{101.2} + \frac{(41-50.6)^2}{50.6} + \frac{(36-50.6)^2}{50.6}$$
$$+ \frac{(38-50.6)^2}{50.6} \approx 29.51$$

Testi suoritetaan aina yksisuuntaisena. Vapausasteet ja testin havaittu merkitsevyystaso (p-arvo):

$$v=k-1=6-1=5$$

$$p = P\left(\chi^2_{(5)} > 29{,}52\right) < 0{,}001$$

$$\rightarrow p<\alpha=0{,}05.$$

Tilastollinen päättely:

Johtuen edellisestä H_0 hylätään.

Tulosten tulkinta:

Aineisto tuki 5% merkitsevyystasolla olettamusta, jonka mukaan havaintoaineisto ei ole peräisin oletettua suhteellista jakaumaa noudattavasta populaatiosta (eli suhteellinen jakauma on jotain muuta kuin nollahypoteesissa väitettiin). Estimaatiksi sopii otoksesta laskettu prosenttijakauma:

mieleisin nallekarkin väri

keltainen	oranssi	vihreä	punainen	valkoinen	musta
35%	26.7%	15.6%	8.1%	7.1%	7.5%

SPSS-tuloste (toinen esimerkki)

NPar Tests

Chi-Square Test

Frequencies

Miten paljon seuraava tekijä vaikuttaa ostospaikan valintaan Turun alueella? Tässä testataan väitettä, jonka mukaan jokaisen mielipidekategorian osuus alueella on yhtä suuri.

Helppo liikkua/siirtyä liikkeestä toiseen

	Observed N	Expected N	Residual
erittäin vähän	71	390,2	-319,2
vähän	164	390,2	-226,2
jonkin verran	470	390,2	79,8
paljon	715	390,2	324,8
erittäin paljon	531	390,2	140,8
Total	1951		

Test Statistics

	Helppo liikkua/siirtyä liikkeestä toiseen
Chi-Square	729,736[a]
df	4
Asymp. Sig.	,000

a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 390,2.

Turun yliopisto
University of Turku

111

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.

Turun kauppakorkeakoulu • Turku School of Economics

Hypoteesit:

H_0: Muuttujat perusjoukossa riippumattomia

H_1: Muuttujat perusjoukossa riippuvia

Odotetut (riippumattomuusoletuksen mukaiset) frekvenssit e:

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

, jossa osoittajan kerrottavat ovat rivin *i* ja sarakkeen *j* frekvenssien summia (luettavissa edellisestä taulukosta). Rivejä on tässä 2 ja sarakkeita 3 kpl.

e_{ij}

	pesäpallo	koripallo	jalkapallo
naiset	18.80	17.72	21.48
miehet	16.28	15.28	18.52

Tässä esim. $18.80=\frac{58\cdot 35}{108}$

Approksimaation hyvyyden tarkastelu: Odotettuja frekvenssejä koskevat samat vaatimukset kuin χ2 -yhteensopivuustestissä. Tässä jokainen odotettu frekvenssi on suurempi kuin 5 eli kriteerit toteutuvat → tehdään χ2-testi.

Testisuureen kaava ja jakauma:

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

,jossa r ja s ovat muuttujien luokkien lukumääriä (2 ja 3).

Testisuureen arvo:

$$\chi^2_{hav}=\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}\approx 1.55$$

Yhteenlaskettavia on yhtä monta kuin taulukossa soluja eli kuin on muuttujien arvojen yhdistelmiä eli 6 kpl.

Testi suoritetaan aina yksisuuntaisena. Vapausasteet ja testin havaittu merkitsevyystaso (p-arvo):

$$v=(r-1)(s-1)=(2-1)(3-1)=2$$

$$p = P\left(\chi^2_{(2)} > 1{,}55\right)$$

$$0.1<p<0.95$$

$$\rightarrow p>\alpha.$$

Tilastollinen päättely:

Johtuen edellisestä H_0 hyväksytään.

Tulosten tulkinta:

Aineisto tuki 5% merkitsevyystasolla olettamusta, jonka mukaan sukupuoli ja mieliurheilulaji ovat perusjoukossa riippumattomia.

Huom! Jos nollahypoteesi olisi hylätty, olisi riippuvuutta tarkasteltu ristiintaulukoinnin rivi- tai sarakeprosenttien avulla:

	Pesäpallo	Jalkapallo	Koripallo	
Miehet	32 %	25 %	42 %	100%
Naiset	32 %	36 %	32 %	100%

Miesten suorituin mieliurheilulaji koripallo, naisilla jalkapallo jne.

SPSS-tuloste (toinen esimerkki).

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
talouden kuukausitulot 2011 * Hyvä asiakaspalvelu	1873	93,2%	137	6,8%	2010	100,0%

talouden kuukausitulot 2011 * Hyvä asiakaspalvelu Crosstabulation

			Hyvä asiakaspalvelu					
			erittäin vähän	vähän	jonkin verran	paljon	erittäin paljon	Total
talouden kuukausitulot 2011	alle 1000 €	Count	8	26	67	53	41	195
		% within talouden kuukausitulot 2011	4,1%	13,3%	34,4%	27,2%	21,0%	100,0%
	1000 - 1999 €	Count	13	28	96	144	81	362
		% within talouden kuukausitulot 2011	3,6%	7,7%	26,5%	39,8%	22,4%	100,0%
	2000 - 2999 €	Count	8	32	99	146	104	389
		% within talouden kuukausitulot 2011	2,1%	8,2%	25,4%	37,5%	26,7%	100,0%
	3000 - 3999 €	Count	9	19	69	95	83	275
		% within talouden kuukausitulot 2011	3,3%	6,9%	25,1%	34,5%	30,2%	100,0%
	4000 - 4999 €	Count	4	12	65	96	60	237
		% within talouden kuukausitulot 2011	1,7%	5,1%	27,4%	40,5%	25,3%	100,0%
	5000 - 5999 €	Count	3	12	44	78	32	169
		% within talouden kuukausitulot 2011	1,8%	7,1%	26,0%	46,2%	18,9%	100,0%
	6000 - 6999 €	Count	1	7	22	41	30	101
		% within talouden kuukausitulot 2011	1,0%	6,9%	21,8%	40,6%	29,7%	100,0%
	7000 - 7999 €	Count	1	4	17	21	15	58
		% within talouden kuukausitulot 2011	1,7%	6,9%	29,3%	36,2%	25,9%	100,0%
	8000 € tai enemmän	Count	2	5	24	30	26	87
		% within talouden kuukausitulot 2011	2,3%	5,7%	27,6%	34,5%	29,9%	100,0%
Total		Count	49	145	503	704	472	1873
		% within talouden kuukausitulot 2011	2,6%	7,7%	26,9%	37,6%	25,2%	100,0%

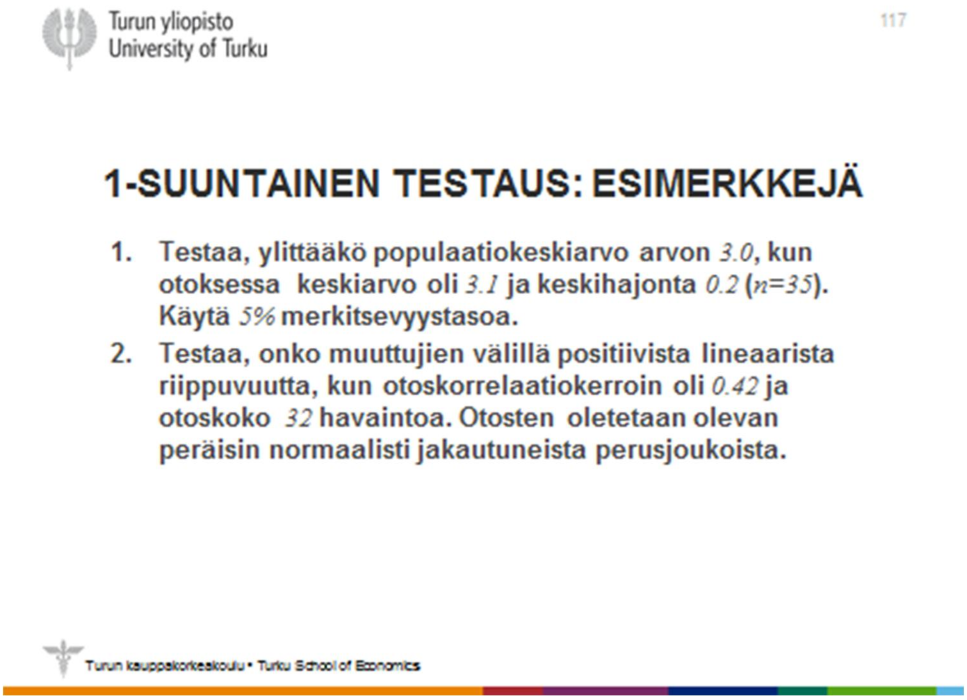
Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	44,229[a]	32	,074
Likelihood Ratio	43,559	32	,084
Linear-by-Linear Association	9,904	1	,002
N of Valid Cases	1873		

a. 5 cells (11,1%) have expected count less than 5. The minimum expected count is 1,52.

$$p = 0{,}074 > 0{,}05 = \alpha$$

Johtuen edellisestä H_0 hyväksytään (p>0,05). Aineisto tuki väittämää, jonka mukaan mielipide asiakaspalvelun tärkeydestä ostospaikan valinnassa ja tuloluokka eivät riipu toisistaan.



1.

$$H_0 \colon \mu \leq 3{,}0$$
$$H_0 \colon \mu > 3{,}0$$

$$z_{hav} = \frac{3{,}1 - 3{,}0}{0{,}2 \big/ \sqrt{35}} \approx 2{,}96$$

$$p = 1 - \Phi(2{,}96) \approx 1 - 0{,}9985 = 0{,}0015 < 0{,}05 = \alpha$$

H_0 hylätään. Aineisto tuki väittämää, jonka mukaan populaatiokeskiarvo ylittää arvon 3.0 (p<0,05).

2.

$$H_0 \colon \rho \leq 0$$

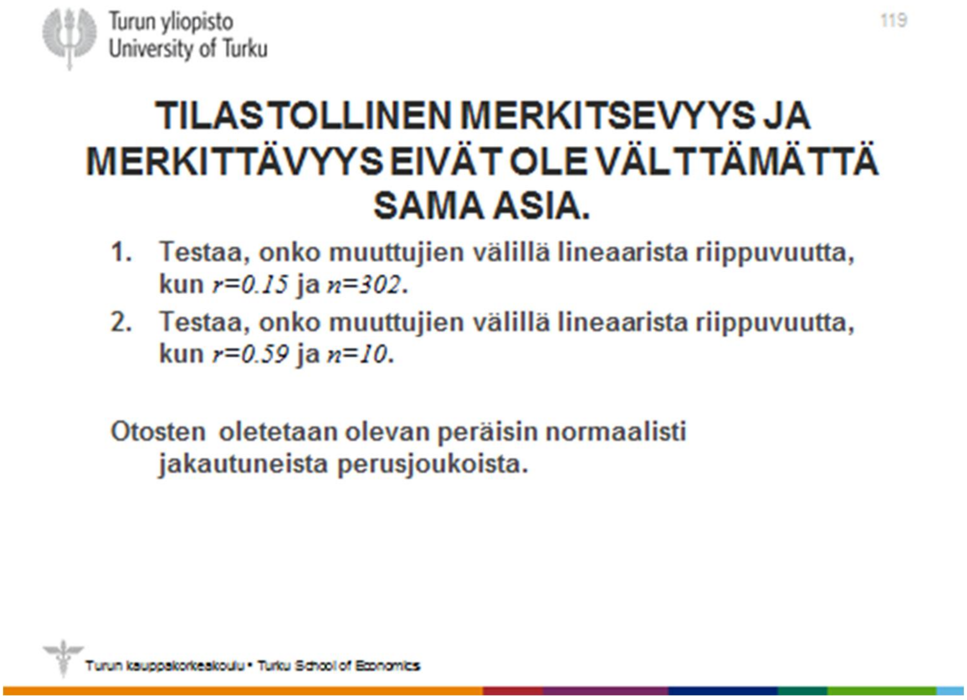
$$H_1 \colon \rho > 0$$

$$t_{hav} = \frac{0{,}42\sqrt{32-2}}{\sqrt{1-0{,}42^2}} \approx 2{,}53$$

$$v = n - 2 = 32 - 2 = 30$$

$$0{,}005 < p = P(t > 2{,}53) < 0{,}01 < 0{,}05 = \alpha$$

H_0 hylätään. Aineisto tuki väittämää, jonka mukaan muuttujien välillä on populaatiossa positiivista lineaarista riippuvuutta ($p<0{,}05$).



1. $t_{hav} = \frac{0{,}15 * \sqrt{302-2}}{\sqrt{1-0{,}15^2}} \approx 2{,}63$

$$v = n - 2 = 302 - 2 = 300$$

$$0{,}002 < p = 2 * P(t > 2{,}63) < 0{,}01 < 0{,}05 = \alpha$$

H_0 hylätään. Aineisto tuki väittämää, jonka mukaan muuttujien välillä on populaatiossa lineaarista riippuvuutta (p<0,05).

2. $t_{hav} = \frac{0{,}59 * \sqrt{10-2}}{\sqrt{1-0{,}59^2}} \approx 2{,}07$

$$v = n - 1 = 10 - 2 = 8$$

$$\alpha = 0{,}05 < p = 2 * P(t > 2{,}07) < 0{,}1$$

H_0 hylätään. Aineisto tuki väittämää, jonka mukaan muuttujien välillä ei ole populaatiossa lineaarista riippuvuutta (p>0,05).