

PSYKOLOGIA  
KOGNITIOTIEDE

V A L I N T A K O E 2 0 1 2

TILASTOMATEMATIIKAN  
LISÄMATERIAALI



TAMPEREEN  
YLIOPISTO



HELSINGIN YLIOPISTO



Turun yliopisto  
University of Turku

PSYKOLOGIAN JA KOGNITIOTIETEEN VALINTAKOE 2012

# **Tilastomatematiikan lisämateriaali**

---

© Copyright

Helsingin yliopisto, käyttäytymistieteiden laitos,  
Tampereen yliopisto, yhteiskunta- ja kulttuuritieteiden yksikkö ja  
Turun yliopisto, käyttäytymistieteiden ja filosofian laitos.  
Materiaalin luvaton kopiointi kielletty.

# Aluksi

Tässä lisämateriaalissa on pyritty kuvaamaan muutamia täydennyksiä valintakoekirjaan, mutta keskeisemmän osan saavat psykologian valintakoetehtävät vuosien takaa. Vanhat koetehtävät on pyritty valitsemaan sopivan kattavasti niin, että niiden ratkaisemisesta voisi olla aidosti hyötyä valintakokeeseen valmistautumisessa. Tehtäviin esitetään myös ratkaisumallit, jotka eivät kaikissa tapauksissa ole ainoat oikeat, sillä kuten useinkin tilastomatematiikan tehtävissä, oikean ratkaisun voi esittää monella tavalla.

Esitystavassa on pyritty löytämään sopiva tasapaino tiiviin ja tyhjentävän ilmaisun välillä. Tästä johtuen lisämateriaalista saanee parhaimman hyödyn kun sitä lukee yhdessä valintakoekirjan, Holopainen & Pulkkinen Tilastolliset menetelmät, kanssa.

Koska materiaalia on otettu vanhojen valintakokeiden originaaliversioista, ei taitto – ikävä kyllä – ole aina paras mahdollinen. Ulkoasun puutteista huolimatta asiat toivottavasti selviävät lukijalle.

Tässä esitettyjen vanhojen koetehtävien lisäksi kannattaa opiskella Helsingin yliopiston käyttäytymistieteellisen tiedekunnan opiskelijavalintojen verkkosivulta löytyvät muutaman viime vuoden kokeet. Koska valintakoekirja oli noissa kokeissa eri kuin tänä vuonna, niin aivan kaikkien tilastomatematiikan tehtävien laskeminen noista kokeista sellaisenaan ei ole tarpeen.

Valintakoekirjan ja tämän lisämateriaalin mahdollisia virheitä tullaan korjaamaan Helsingin yliopiston käyttäytymistieteellisen tiedekunnan opiskelijavalintojen verkkosivulla kevään 2012 aikana. Verkkosivu on sama mistä tämä pdf-tiedosto on ladattavissa – opiskelijavalinnan etusivun ajankohtaiset-osio. Korjausten osalta verkkosivua päivitetään keskiviikkoina 15.2.2012, 14.3.2012 ja 11.4.2012 kello 9.00.

Tätä lisämateriaalia saavat käyttää ainoastaan yksityishenkilöt Helsingin, Tampereen ja Turun yliopistojen psykologian yhteisvalinnan ja Helsingin yliopiston kognitiotieteen valinnan valintakoetta 2012 varten. Kaikki muu käyttö on kielletty.

# Valintakoe kirjasta

*Tässä osassa on täydennyksiä ja huomioita valintakoe kirjasta  
Holopainen, M. & Pulkkinen, P. Tilastolliset menetelmät.*

Valintakoe kirjassa moni kohta käsitellään siten, että tulokset saadaan tietokoneohjelman avulla. Tässä lisämateriaalissa esitellään muutaman tärkeimmän tällaisen kohdan laskeminen 'käsini'. Lisäksi muutamia asioihin esitetään tarkennuksia ja huomioita.

Muutamia huomioita ovat sellaisista asioista, että tämä lisämateriaali ja valintakoe kirja ovat ristiriidassa. Näissä kohtaa tulee luottaa lisämateriaaliin. Valintakoe kessa on siis syytä muistaa, että jos tämän lisämateriaalin ja valintakoe kirjan välillä on jostakin asiasta ristiriita, niin lisämateriaalin mukainen vastaus on oikea.

Valintakoe kirjassa ei esitetä vastauksia kaikkiin tehtäviin. Yleensä vastaukset puuttuvat, jos vastauksen esittäminen olisi vienyt melko paljon tilaa. Esimerkiksi luvusta 3 ei vastauksia ole lainkaan. Tässäkään ei näitä vastauksia tilan rajallisuuden vuoksi esitetä, mutta on silti syytä tutkia nämäkin tehtävät valintakoe kirjasta tarkasti. Lukua 3 on hyvä kuvata esimerkkinä. Luvussa käsitellään muun muassa graafista esittämistä ja vaikka luvussa on kuvien tekemiseen Excel-ohjeita, on kuvien käsini piirtäminen tarpeen harjoitella. Käsini piirtämisen harjoittelu ei kuitenkaan ole kovin vaikeaa, minkä vuoksi sitä tässä lisämateriaalissa käsitellään vain vähän.

Lisämateriaalissa käytetään kahta tapaa summamerkin indeksissä. Ne kuvataan alla ja tarkoittavat siis täysin samaa. Joskus – kun indeksointi on selvä ja yksikäsitteinen – indeksointi on jätetty kokonaan merkitsemättä.

$$\sum_{i=1}^n x_i = \sum_{i=1}^n x_i$$

ja jos indeksointi on selvä, niin yksinkertaisesti  $\sum x_i$  tai jopa  $\sum x$  on käytössä.

Valintakoe kirjan kaikista luvuista ei ole tässä lisämateriaalissa huomioita tai täydennyksiä ja yksittäisistä valintakoe kirjan painovirheistä tullaan ilmoittamaan aikaisemmin mainitulla tavalla – siis verkkosivulla kolme kertaa kevään 2012 aikana. Täydennyksiä ja huomioita esitetään lukuihin 1, 3, 4, 6 ja 8.

Valintakoe kirjassa kuvatut aineistot ASIAKAS ja KYSELY löytyvät Helsingin yliopiston käyttäytymistieteellisen tiedekunnan opiskelijavalintojen verkkosivulta tekstitiedostoina 15.2.2012 kello 9.00 mennessä.

## 1 – Tiede, tutkimus ja tilastollinen ajatteluprosessi

Osassa 1.2 ”Mittaaminen, validiteetti ja reliabiliteetti” sanotaan (s. 15, viimeinen kpl): ”**Välimatka-asteikolla** (interval scale) luokat ovat samanleveyisiä ja välimatka luokasta toiseen on yhtä suuri asteikon joka kohdassa”. Tässä puhuminen luokasta saattaa mutkistaa välimatka-asteikon ymmärtämistä. Tulee muistaa, että yksi tilastoyksikkö voi kuulua mitattavan ominaisuuden suhteen vain yhteen luokkaan, mutta ominaisuuden kaikkiin luokkiin ei tarvitse kuulua tilastoyksiköitä. Ajatellaan tilastoyksiköiksi vuoden 2012 jokaisen kuukauden 15. päivä ja mitattavaksi ominaisuudeksi lämpötila celsiusasteina kello 12. On aivan hyvin mahdollista, että toukokuussa ja elokuussa lämpötila olisi +20 astetta, mutta yhdessäkään kuukaudessa ei lämpötila olisi -20 astetta.

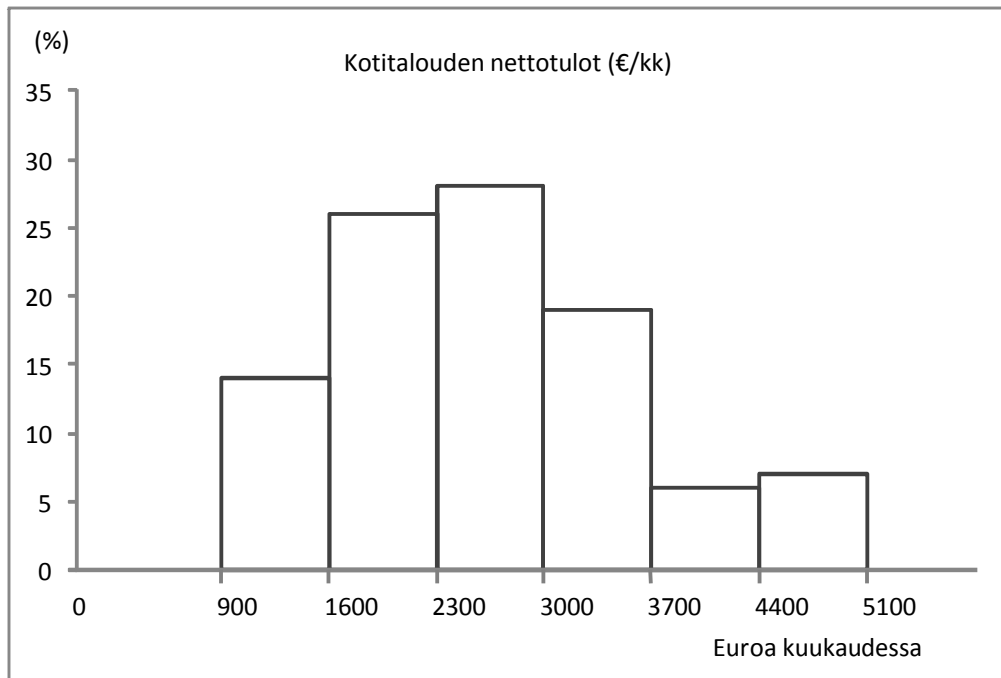
Sillä, että välimatka luokasta toiseen on yhtä suuri, tarkoitetaan että kun muuttujaa mitataan välimatka-asteikolla, niin havaintojen väliset etäisyydet voidaan laskea. Esimerkiksi jos havainto A saa tietyllä muuttujalla arvon 4, havainto B arvon 5.5, havainto C arvon 8 ja havainto D arvon 9.5, niin havaintojen A ja B välinen etäisyys on yhtä suuri kuin havaintojen C ja D välinen etäisyys (1.5 kummassakin tapauksessa), mutta havaintojen B ja C välinen etäisyys on suurempi (eli 2.5). Sen sijaan ei voida sanoa, että havainto C olisi muuttujalla kaksi kertaa niin suuri kuin havainto A, jos kyseessä on välimatka-asteikollinen muuttuja.

Psykologiassa – ja käyttäytymistieteissä yleensä – järjestys- tai välimatka-asteikolliseen mittaamiseen liittyy tärkeitä ja olennaisia kysymyksiä, joita ei tässä kuitenkaan ole mahdollista eikä tarpeellistakaan tarkemmin käsitellä. Riittää toistaiseksi todeta, että kun jonkin psykologisen ominaisuuden (esim. älykkyyden) mittausta tehdään hyvin (esim. luotettava ja tarkka älykkyystesti), niin mittauksen taso on vähintään välimatka-asteikollinen.

Tehtävässä 1.3 (s. 27 ja vastaukset s. 344) tuli ratkaista, millä mitta-asteikolla muuttujia on mitattu. Vastauksissa älykkyydosamäärän mittaamisen tasoksi on mainittu suhdeasteikko. Älykkyydosamäärä on kuitenkin välimatka-asteikollinen muuttuja. Tehtävän vastauksessa kerrotaan kouluarvosanan (1-5) mittaamisen tasoksi välimatka-asteikko. Kouluarvosanan kohdalla välimatka-asteikollisuus on mahdollista, mutta tämä riippuu siitä millaisten mittaamistoimenpiteiden perusteella kouluarvosana määräytyy. Ajatellaan, että kouluarvosana kuvastaa jonkin asian (vaikkapa tilastomatematiikan) osaamista ja ilmaistaan kokonaisluvulla, joka on 1, 2, 3, 4 tai 5 (voidaan ajatella yhtä hyvin myös arvosanoja 4, 5, 6, 7, 8, 9 ja 10 – itse asiassa arvosanan ei tarvitse olla kokonaisluku, mutta ainakin koulutodistuksissa näin yleensä on). Välimatka-asteikollisuus edellyttää, että osaamisessa tapahtuu samansuuruinen muutos arvosanan noustessa yhden numeron. Useinkaan kouluarvosanoihin tapahtuva mittaaminen ei tapahdu näin, vaan kouluarvosanan mittaamisen taso on järjestysasteikko.

### 3 – Tilastojen tutkiminen

Osassa 3.2 'Tilastojen graafinen esittäminen' käydään läpi havainnollistavien kuvien teon periaatteita. Kuvien piirtämisen käytännön ohjeita käydään läpi tietokoneohjelman toiminnoin. Suurimmaksi osaksi kuvien piirtäminen käsin onnistuu kirjassa kerrottujen periaatteiden pohjalta, mutta histogrammin laatimista on tarpeen käsitellä jonkin verran. Näin erityisesti siksi että kirjassa sivun 60 esimerkin histogrammissa on paitsi ilmeinen virhe pylväiden korkeuksissa myös epätyydyttävästi laadittu x-akseli ja hieman erikoisesti valittu luokkaväli. Luokkaväliksi on valittu 700 € kirjan ss. 49–51 ohjeiden ja esimerkkien pohjalta. X-akselilla luvut on merkitty pylväiden keskikohtille ja vaikka taulukossa luokittelu alkaa 900:sta ja suurin frekvenssi (28) on luokassa 2300–3000, niin kuviossa luokittelu alkaa ilmeisesti 700:sta, jolloin suurin frekvenssi (29) on luokassa 1400–2100. Alla on ensin sivun 60 taulukon mukainen kuva ja seuraavalla sivulla oikea taulukko sivulla 60 olevaan histogrammiin sillä poikkeuksella, että viimeistä luokkaa ei ole piirretty aivan oikean levyisenä ja korkuisena - aineistossa on yksi havainto, jonka tulot (5100 €) ylittävät kirjan kuvaan merkityn ylärajan 4900. Jos histogrammissa käytetään epätasavälistä luokittelua, niin pylväiden pinta-alojen tulee olla verrannollisia luokkien prosentiosuuksiin. Näin on myös tasavälisessä luokittelussa, mutta koska tällöin pylväiden kannat ovat saman levyisiä, niin pylväiden korkeudet kuvaavat luokkien kokoa oikein. Epätasavälisesti luokitellun muuttujan histogrammin piirtämistä ei käsitellä tässä.



**Frequency Distribution - Quantitative**

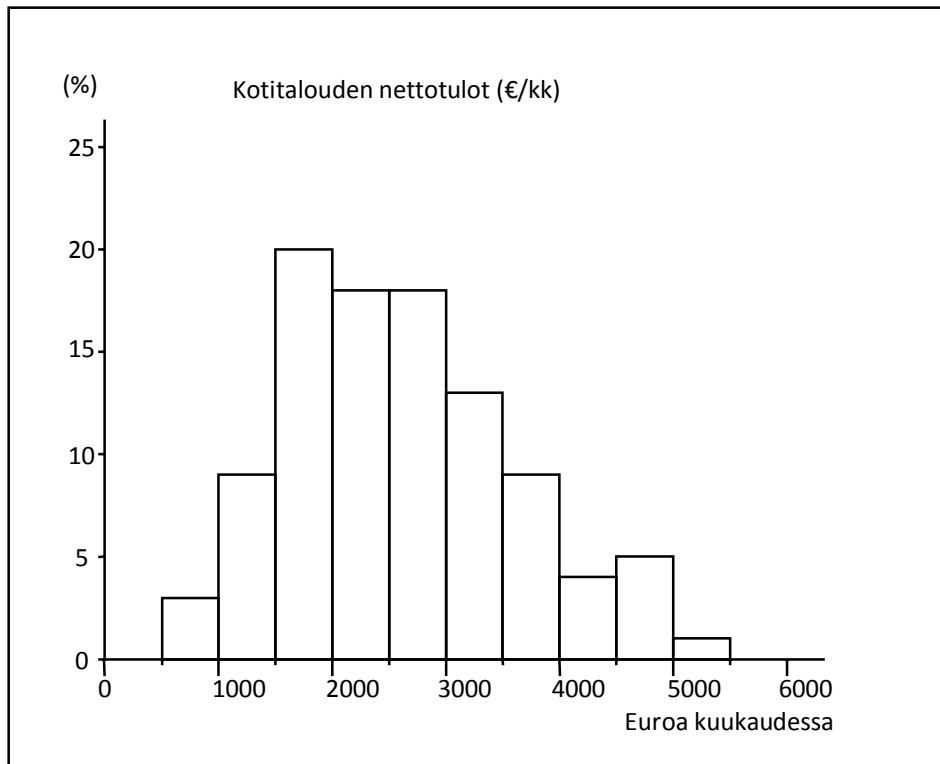
<i>tulot</i>					<i>cumulative</i>			
<i>lower</i>		<i>upper</i>	<i>midpoint</i>	<i>width</i>	<i>frequency</i>	<i>percent</i>	<i>frequency</i>	<i>percent</i>
700	<	1 400	1 050	700	14	14,0	14	14,0
1 400	<	2 100	1 750	700	29	29,0	43	43,0
2 100	<	2 800	2 450	700	24	24,0	67	67,0
2 800	<	3 500	3 150	700	21	21,0	88	88,0
3 500	<	4 200	3 850	700	6	6,0	94	94,0
4 200	≤	5 100	4 650	900	6	6,0	100	100,0
					100	100,0		
2 missing or invalid data values								

Tässä kyseisessä tapauksessa olisi kannattanut valita luokkaväliksi 500. Niin sanotut pyöreät luvut ovat monesti käytännöllisempiä kuvassa eikä luokkien lukumäärä tulojen kohdalla kuitenkaan kasva liian suureksi. X-akselien merkintöjen tulee olla oikeilla kohdilla, koska histogrammia käytetään vain kvantitatiivisille muuttujille. Kaikkia lukuja ei tarvitse välttämättä kirjoittaa, kunhan luokkarajat käyvät yksiselitteisesti ilmi. Alla on frekvenssitaulukko ja histogrammi, josta x-akselilta on jätetty ”viidet sadat” merkitsemättä, mutta asteikko käy silti selvästi ilmi, kun käytetään erimittaisia asteikkomerkkejä (pitkät viivat asteikon ulkopuolella tuhansille ja lyhyet viisille sadoille). Kun vertaat kuvaa kirjan sivun 60 kuvaan tai edellisellä sivulla esitettyyn, huomaa y-akselin muutokset.

**Frequency Distribution - Quantitative**

<i>tulot</i>					<i>cumulative</i>			
<i>lower</i>		<i>upper</i>	<i>midpoint</i>	<i>width</i>	<i>frequency</i>	<i>percent</i>	<i>frequency</i>	<i>percent</i>
500	<	1 000	750	500	3	3,0	3	3,0
1 000	<	1 500	1 250	500	9	9,0	12	12,0
1 500	<	2 000	1 750	500	20	20,0	32	32,0
2 000	<	2 500	2 250	500	18	18,0	50	50,0
2 500	<	3 000	2 750	500	18	18,0	68	68,0
3 000	<	3 500	3 250	500	13	13,0	81	81,0
3 500	<	4 000	3 750	500	9	9,0	90	90,0
4 000	<	4 500	4 250	500	4	4,0	94	94,0
4 500	<	4 200	4 750	500	5	5,0	99	99,0
5 000	≤	5 500	5 250	500	1	1,0	100	100,0
					100	100,0		
2 missing or invalid data values								

**VALINTAKOEKIRJASTA**



## 4 – Suoran jakauman tunnusluvut

Osassa 4.1 'Sijaintiluvut' esitetään, että jos järjestysasteikollisen muuttujan havaintoarvoja on parillinen määrä, niin mediaaniksi valitaan jompikumpi keskimmäisistä arvoista (tai jos nämä ovat hyvin kaukana toisistaan, niin ilmoitetaan joko molemmat tai jätetään mediaani kokonaan ilmoittamatta). **Jos havaintoja on parillinen määrä, niin myös järjestysasteikollisen muuttujan kohdalla ilmoitetaan mediaanina kahden keskimmäisen havainnon aritmeettinen keskiarvo.** Sivun 80 esimerkin 2b mediaani on siis  $(14+17) / 2 = 15.5$ . Mediaanin tulkinnassa tulee luonnollisesti muistaa muuttujan mittauksen taso.

Vaikka aritmeettisen keskiarvon laskeminen aineistosta on helppoa, niin kenties siksi siitä ei yksinkertaista esimerkkiä kirjassa ole. Seuraavassa lasketaan sivun 80 esimerkin 2a aineiston keskiarvo varmuuden vuoksi. Nyt lukuja ei ajatella järjestyslukuiksi, vaan välimatka-asteikollisen muuttujan arvoiksi.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \sum_{i=1}^n x_i/n$$

$$\bar{x} = \frac{2 + 3 + 3 + 4 + 8 + 8 + 12 + 23 + 25}{9} = \frac{88}{9} = 9\frac{7}{9} \approx 9.78$$

Osassa 4.2 'Hajontaluvut' kerrotaan otoksesta laskettavan keskihajonnan laskemisesta. Keskihajonta kuvaa havaintoarvojen keskimääräistä poikkeamista keskiarvosta ja se on sopiva tunnusluku kertomaan arvojen vaihtelusta välimatka- ja suhdeasteikollisilla muuttujilla. Kirjassa esitetään otoskeskihajonnalle kaava, jossa keskihajonnan merkitys keskimääräisenä poikkeamana keskiarvosta käy ilmi selkeästi. Joskus keskihajonta on kuitenkin kätevämpi laskea siitä johdetulla alla esitetyllä kaavalla.

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

Tätä kaavaa voidaan soveltaa, myös kun lasketaan keskihajontaa luokitellusta aineistosta.

$$s = \sqrt{\frac{\sum_{i=1}^k f_i x_i^2 - \frac{(\sum_{i=1}^k f_i x_i)^2}{n}}{n-1}}$$

$k$  = luokkien lukumäärä,  $f_i$  = luokan  $i$  frekvenssi,  $x_i$  = luokan  $i$  luokkakeskus

Seuraavassa esitetään kummallakin tavalla laskien kirjan s. 91 oleva esimerkki 11 (jossa hajonnan laskemisessa on pieni virhe kirjassa). Käsien laskemisessa käytetään apuna taulukkoon laskettuja välivaiheita.

VALINTAKOEKIRJASTA

nettopaino	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$x_i^2$
983	-4.2	17.64	966289
976	-11.2	125.44	952576
990	2.8	7.84	980100
997	9.8	96.04	994009
983	-4.2	17.64	966289
969	-18.2	331.24	938961
986	-1.2	1.44	972196
995	7.8	60.84	990025
978	-9.2	84.64	956484
996	8.8	77.44	992016
998	10.8	116.64	996004
986	-1.2	1.44	972196
999	11.8	139.24	998001
988	0.8	0.64	976144
992	4.8	23.04	984064
979	-8.2	67.24	958441
$\Sigma$	15795	1168.44	15593795

$$\bar{x} = \frac{15795}{n} = 987.1875 \approx 987.2$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{1\,168.44}{15}} = \sqrt{77.896} \approx 8.83$$

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1}} = \sqrt{\frac{15\,593\,795 - \frac{15\,795^2}{16}}{15}}$$

$$= \sqrt{\frac{15\,593\,795 - 15\,592\,626.5625}{15}} = \sqrt{\frac{1\,168.4375}{15}} \approx 8.83$$

Pieni ero neliöjuurilausekkeen osoittajassa johtuu siitä, että ensin lasketulla tavalla keskiarvo oli pyöristetty yhteen desimaaliin.

Kun hajonta lasketaan perusjoukon arvoista, kaavoissa on jakajana  $(n-1)$  sijasta  $n$ .

## 6 – Todennäköisyysjakaumat

Osaan 6.2 'Jatkuvat todennäköisyysjakaumat' liittyen käydään tässä lyhyesti läpi normitetun normaalijakauman kertymäfunktion arvojen taulukon lukeminen. Kirjan taulukossa on  $z$  välillä  $-3.49$ – $3.49$ .  $Z$ -arvon kokonaisosa ja kymmenesosa luetaan sarakkeesta ja sadasosa riviltä  $z$ . Taulukosta sarakkeen ja rivin leikkauskohdasta löytyy kertymäfunktion arvo.

Siis esimerkiksi jos tarvitsee selvittää  $P(Z < 1.75) = \Phi(1.75)$ , niin taulukon toiselta sivulta etsitään sarakkeesta  $z$  luku 1.7 ja riviltä  $z$  luku .05. Näiden leikkauskohdasta löytyy kertymäfunktion arvo 0.9599. Arvojen  $-\infty$  ja 1.75 väliin jää siis 95.99 % normitetun normaalijakauman tiheysfunktion kuvaajan ja  $x$ -akselin rajoittamasta alasta.

Kannattaa myös huomata, että koska normaalijakauma on symmetrinen, niin  $\Phi(-z) = 1 - \Phi(z)$

Siis esimerkiksi  $\Phi(-1.75) = 0.0401$  ja  $\Phi(1.75) = 0.9599 = 1 - 0.0401$ .

Koska näin on, niin jos normaalijakauman taulukko on annettu vain välille  $-3.49$ – $0$  tai välille  $0$ – $3.49$ , voi toisen välin arvot laskea.

## 8 – Tilastollinen riippuvuus

Osassa 8.2 'Korrelaatio' esitetään Pearsonin korrelaatiokerroin kahden muuttujan välisen yhteyden mittana. Kertoimen laskemiseen esitetään yksi kaava, mutta kertoimen voi laskea muillakin tavoin.

$$r = \frac{s_{xy}}{s_x \times s_y}$$

$s_{xy}$  on muuttujien x ja y välinen kovarianssikerroin,  $s_x$  on muuttujan x keskihajonta ja  $s_y$  on muuttujan y keskihajonta

Kovarianssikerroin on kahden muuttujan yhteisvaihtelun estimaatti, jota ei ole normitettu välille [-1, 1]. Kovarianssi lasketaan kaavalla

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Siis jokaisen havainnon kohdalla lasketaan havainnon x-muuttujan arvon ja x-muuttujan keskiarvon erotus sekä havainnon y-muuttujan arvon ja y-muuttujan keskiarvon erotus ja nämä arvot kerrotaan. Näin saadut tulot lasketaan yhteen otoksen kaikilta havainnoilta ja tämä summa jaetaan luvulla n-1.

Korrelaatiokertoimen voi laskea myös normittamalla kummankin muuttujan havaintoarvot ja käyttämällä kaavaa

$$r = \frac{\sum(z_x \times z_y)}{n-1}$$

$z_x$  on x-muuttujan normitettu arvo ( $z_x = \frac{x - \bar{x}}{s_x}$ ),  $z_y$  on y-muuttujan normitettu arvo ( $z_y = \frac{y - \bar{y}}{s_y}$ )

Joskus korrelaatiokertoimen voi laskea helpoiten 'käsien' seuraavien kaavojen avulla.

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y}$$

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{(n(\sum x_i^2) - (\sum x_i)^2) \cdot (n(\sum y_i^2) - (\sum y_i)^2)}}$$

Seuraavassa taulukossa on pieni kymmenen havainnon aineisto, jonka avulla eri kaavojen mukaisia laskuja on helppo harjoitella, sillä välivaiheissa saatavat luvut ovat 'pyöreitä'. Kummankin muuttujan keskiarvo on 4 ja hajonta 2. Kovarianssikertoimen arvo on 3 ja korrelaatiokertoimen arvo on 0.75.

nro	x	y
1	1	3
2	2	3
3	3	1
4	3	2
5	3	3

nro	x	y
6	4	4
7	5	5
8	5	7
9	7	5
10	7	7

## Vanhoja valintakoetehtäviä

*Tässä osassa on joitakin vanhoja valintakoetehtäviä. Ratkaisut niihin löytyvät lisämateriaalin viimeisestä osasta.*

### **Vuosi 1989, tehtävä 4**

Sankari oli päässyt tutkimusapulaiseksi erään lääkärin tutkimusprojektiin. Lääkäri oli kiinnostunut siitä, kuinka suuri vaikutus lämpötilan muutoksilla on päähineen käyttöön.

Sankarimme oli saanut seuraavat ohjeet tutkimuksen tekoon: Mene kaupungille johonkin baariin, istu siellä rauhallisessa nurkkapöydässä, josta on näköala vilkasliikenteiselle kadulle ja tutki päähineen käyttöä. Kirjaa seuraavat tiedot tutkimustilanteesta: Päivämäärä, ulkolämpötila, kadulla kulkevista jalankulkijoista kirjaa seuraavat tiedot: päähineen käyttö, sukupuoli, arvioitu ikä, kellonaika.

Sankari teki työtä käskettyä (pienen palkkion toivossa) ja seuraavalla sivulla on hänen tilastonsa. Rinnakkain esiintyvät kahtena eri päivänä tehdyt mittaukset. Ensimmäisen päivänä pakkasta ei ollut lainkaan. Toisella mittauskerralla lämpötila oli  $-20\text{ C}^\circ$ .

Aikaisemmissa tutkimuksissa oli todettu, että kylmemmällä säällä useammat käyttävät päähinettä, mikä kuulostaakin ihan luontevalta. Jos lasket päähineen käyttäjien lukumäärät  $0\text{ C}^\circ$  ja  $-20\text{ C}^\circ$  huomaat, että päähineen käyttäjiä on lähes yhtä paljon.

- a) Mistä johtuu, ettei Sankari saanut yllä mainitulla aineistolla aikaisempien tutkimustulosten mukaista tulosta? Perustele vastauksesi graafisin kuvin. Tilastollisia testejä tuloksen vahvistamiseen ei tämän tehtävän ratkaisuisissa edellytetä.
- b) Jos sinun pitäisi tutkia lämpötilan vaikutusta päähineen käyttöön, miten voisit ottaa huomioon kokeen järjestämisessä Sankarin saamia tuloksia?

Taulukossa esiintyvät muuttujat:

päähine: Päähine on päässä = 1, ei päähinettä päässä = 0

sukup: Sukupuoli, nainen = N, mies = M

ikä: Arvioitu ikä 10 vuoden tarkkuudella

kellonaika: Kellonaika, jolloin tutkittavan henkilön päähineen käyttö kirjattu

**VANHOJA VALINTAKOETEHTÄVIÄ**

Mittauskerta 1, 4.1.1989 Ulkolämpötila 0 C°					Mittauskerta 2, 6.1.1989 Ulkolämpötila -20 C°				
pää- hine	sukup	ikä	kellon- aika	ulko- lämpö- tila	pää- hine	sukup	ikä	kellon- aika	ulko- lämpö- tila
1	M	20	9:01	0	1	N	50	9:51	-20
0	M	40	9:01	0	1	M	20	9:51	-20
0	M	20	9:01	0	0	N	20	9:51	-20
0	N	20	9:01	0	1	M	20	9:51	-20
1	N	50	9:01	0	0	N	20	9:51	-20
1	N	50	9:01	0	1	N	20	9:51	-20
1	N	60	9:02	0	1	N	60	9:52	-20
0	M	60	9:02	0	0	M	20	9:52	-20
1	M	50	9:02	0	1	M	20	9:52	-20
0	M	50	9:02	0	0	M	20	9:52	-20
1	N	60	9:02	0	1	M	60	9:52	-20
0	M	20	9:03	0	1	N	20	9:52	-20
1	N	40	9:03	0	0	N	20	9:52	-20
0	N	20	9:03	0	0	N	20	9:53	-20
1	M	40	9:03	0	1	N	20	9:53	-20
1	M	40	9:03	0	0	M	20	9:53	-20
1	M	50	9:03	0	1	M	50	9:53	-20
0	M	20	9:04	0	0	M	20	9:54	-20
1	N	50	9:04	0	0	M	20	9:54	-20
0	N	20	9:04	0	0	N	20	9:54	-20
1	N	40	9:04	0	1	N	40	9:54	-20
0	N	50	9:04	0	0	N	20	9:54	-20
1	M	60	9:04	0	1	M	20	9:54	-20
1	N	60	9:04	0	0	M	20	9:54	-20
0	N	20	9:04	0	1	N	20	9:54	-20
1	M	50	9:04	0	1	N	50	9:55	-20

**Vuosi 1990, tehtävä 2**

Tutkija on kiinnostunut siitä, ehtiikö aamun lehti tulla kello viiteen mennessä, jolloin hän ehtisi lukea lehden. Tutkijalla on seuraavaa tietoa postinkannosta:

Postin kanto tapahtuu 12 päivän periodeissa. Yhden periodin aikana varsinainen postinkantaja tuo postin 10 kertaa ja sijainen kahtena satunnaisena päivänä. Kun varsinainen postinkantaja tuo postin, niin 9 kertana 10:stä postin saapuminen noudattaa normaalijakaumaa parametrein: odotusarvo kello 4 aamulla, hajonta 1 tunti.

Keskimäärin kerran yhden periodin aikana postinkantajamme unohtuu juhlimaan liian pitkään (silloin, kun hänellä on oma kantovuoro) ja tällöin postin saapuminen noudattaa normaalijakaumaa parametrein: odotusarvo kello 6 aamulla, hajonta 90 minuuttia.

## VANHOJA VALINTAKOETEHTÄVIÄ

Kun sijainen kantaa postia, postin saapuminen noudattaa tasaista jakaumaa kello puoli neljän ja kello 6 välillä (näiden aikojen ulkopuolella postin saapumisen todennäköisyys on 0).

**Millä todennäköisyydellä lehti saapuu kello viiteen mennessä jonakin satunnaisena aamuna, jos:**

- a) Varsinainen postinkantaja tuo lehden ja hän ei ole ollut juhlimassa.
- b) Varsinainen postinkantaja tuo lehden ja hän on ollut juhlimassa.
- c) Sijainen tuo postin.
- d) Millä todennäköisyydellä jonakin satunnaisena aamuna lehti on tipahtanut postiluukusta kello viiteen mennessä?

### **Vuosi 1990, tehtävä 3**

Tutkija on kiinnostunut polkupyöräkypärän käytöstä. 100 pyöräilijän otoksesta 20 käyttää kypärää.

Edellisenä vuonna on todettu, että kypärän käyttäjiä on 10 % polkupyöräilijöistä.

**Voiko tutkija väittää, että kypärän käyttö on lisääntynyt 5 % riskitasolla.**

- a) Aseta  $H_0$  ja  $H_1$  hypoteesit.
- b) Ratkaise tehtävä.

**Miten tämän esimerkin tapauksessa määrittelet hypoteesin testauksessa:**

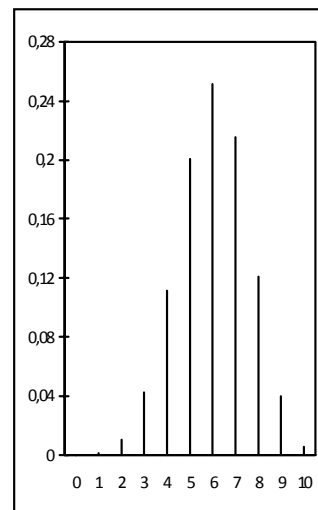
- c) I lajin virhe.
- d) II lajin virhe.

**Vuosi 1991, tehtävä 1**

Seuraavassa on piirretty kolme jakaumaa. Määrittele yksikäsitteisesti mistä todennäköisyys- tai tiheysfunktioista on kysymys. Piirrä pyydyt kertymäfunktio. Kaikkiin piirrettäviin kuvioihin on myös merkittävä asianmukaiset asteikot.

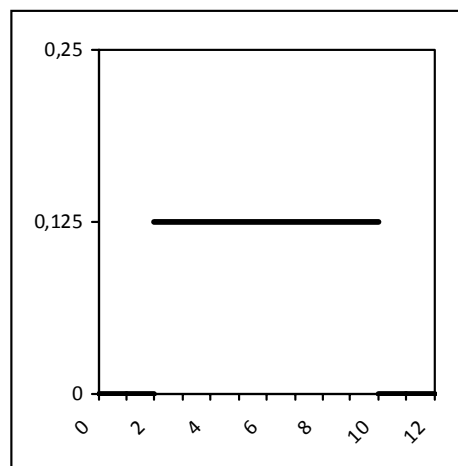
- a) Mikä on seuraava jakauma ja mitkä ovat jakauman tunnusluvut? Kuvioon on piirretty jakauman kaikkien mahdollisten arvojen todennäköisyydet.

pistemäärä	pistemäärän todennäköisyys
0	0.00010
1	0.00157
2	0.01061
3	0.04246
4	0.11147
5	0.20065
6	0.25082
7	0.21499
8	0.12093
9	0.04031
10	0.00604



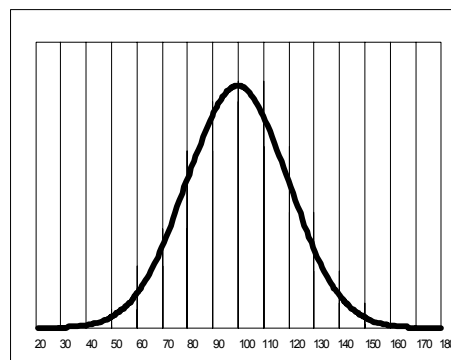
- b) Kuviossa on erään jakauman tiheysfunktio. Mikä todennäköisyysjakauma on kyseessä?

pistemäärä	osuus pinta-alasta	pistemäärä	osuus pinta-alasta
0-1	0	6-7	0.125
1-2	0	7-8	0.125
2-3	0.125	8-9	0.125
3-4	0.125	9-10	0.125
4-5	0.125	10-11	0
5-6	0.125	11-12	0



- c) Mikä jakauma kuvaan on piirretty? Mitkä ovat jakauman tunnusluvut? Tarkennukseksi ohessa pistemäärien väliin jäävän pinta-alan osuus jakauman kokonaispinta-alasta (kun kokonaispinta-alaa kuvataan luvulla 1).

pistemäärä	osuus pinta-alasta	pistemäärä	osuus pinta-alasta
10-20	0.00002	100-110	0.19146
20-30	0.00020	110-120	0.14988
30-40	0.00111	120-130	0.09184
40-50	0.00485	130-140	0.04405
50-60	0.01654	140-150	0.01654
60-70	0.04405	150-160	0.00485
70-80	0.09184	160-170	0.00111
80-90	0.14988	170-180	0.00020
90-100	0.19146	180-190	0.00002



## Vuosi 1991, tehtävä 2

Tutkija halusi tietää, kumpia pareja on Helsingin puhelinyhdistyksen (HPY) 1991 puhelinluettelossa (kotipuhelimet) enemmän: sellaisia pareja, joissa miehen nimi on ensimmäisenä (esim. Virtanen Ahti ja Liisa) vai pareja, joissa naisen nimi on ensimmäisenä (esim. Virtanen Aili ja Pentti). Mukaan otetaan myös sellaiset pariskunnat, joilla on eri sukunimi, mutta yhteinen puhelinnumero (esim. Virtanen Antero ja Turunen Leena). Sellaisista nimistä, joissa oli vain pelkästään miehen nimi (esim. Virtanen Aarre), pelkästään naisen nimi (esim. Virtanen Agnes), enemmän kuin kaksi nimeä tai jonkun yhdistyksen nimi, ei oltu kiinnostuneita.

Kyseisessä puhelinluettelon osassa on varsinaisia puhelinnumeroita sivulta 57 sivulle 1113. Nimet ovat aakkosjärjestyksessä. jokaisella sivulla on kolme palstaa, jokaisella täydellä palstalla on noin 130 puhelinnumeroa. Seuraavassa on näyte puhelinluettelosta sivulta 1080. Näytteessä on sivulta 1080 ensimmäiset 20 nimeä jokaiselta palstalta, kolme nimeä keskeltä sivua jokaiselta palstalta ja kolme nimeä sivun alalaidasta jokaiselta palstalta. osoitteet ja puhelinnumerot on jätetty pois. Alleviivattuna ovat tutkimusotokseen hyväksyttävät nimiparit.

### Vir Virtanen Aarre 1080

(palsta 1)	(palsta 2)	(palsta 3)
Virtanen Aarre	<u>Virtanen Anne ja Markku</u>	<u>Virtanen Eero ja Tarja</u>
<u>Virtanen Aarre ja Glory</u>	Virtanen Anne-Leena	Virtanen Eero
<u>Virtanen Aarre A ja Sylvia</u>	Virtanen Anneli	Virtanen Eeva
Virtanen Aatos Viljo	Virtanen Anneli	Virtanen Eeva
Virtanen Agnes	Virtanen Anneli	Virtanen Eeva
<u>Virtanen Agnes ja Kalle</u>	Virtanen Anneli	Virtanen Eeva
Virtanen Ahti	<u>Virtanen Anneli ja kalle</u>	Virtanen Eeva
Virtanen Ahti	<u>Virtanen Anneli ja Olavi</u>	<u>Virtanen Eeva ja Juhani</u>
Virtanen Ahti	Virtanen Anni	<u>Virtanen Eeva ja Olavi</u>
<u>Virtanen Ahti ja Liisa</u>	Virtanen Anni	<u>Virtanen Eeva ja Soini</u>
<u>Virtanen Ahti ja Sally</u>	<u>Virtanen Anni ja Jukka</u>	<u>Virtanen Eeva ja Veikko</u>
<u>Virtanen Ahti ja Senja</u>	<u>Virtanen Anniina ja Risto</u>	Virtanen Eeva, Jouko ja Heikki
<u>Virtanen Aija ja Tuuja</u>	Virtanen Annikki	Virtanen Eeva Kaarina
Virtanen Aila	Virtanen Annikki	Virtanen Eeva-Liisa
Virtanen Aila	Virtanen Annikki	Virtanen Eeva-Liisa
Virtanen Aila	<u>Virtanen Annikki ja Tippo</u>	<u>Virtanen Eeva-Liisa ja Matti</u>
Virtanen Aila	<u>Virtanen Annikki ja Rauno</u>	<u>Virtanen Eevi ja Kalevi</u>
Virtanen Aila	<u>Virtanen Annikki ja Tauno</u>	Virtanen Eigel
Virtanen Aila	Virtanen Annukka	Virtanen Eija
Virtanen Aila	Virtanen Ansa	Virtanen Eija
.	.	.
.	.	.
<u>Virtanen Aino ja Rainer</u>	<u>Virtanen Arto ja Anja</u>	Virtanen Elsa
<u>Virtanen Aino ja Toini</u>	<u>Virtanen Arto ja Marja-Leena</u>	Virtanen Elsa
<u>Virtanen Aino ja Vilho</u>	Virtanen Avo	Virtanen Else
.	.	.
.	.	.
<u>Virtanen Anne ja Esa</u>	Virtanen Eero	Virtanen Eva
<u>Virtanen Anne ja Juha</u>	<u>Virtanen Eero ja Pirkko</u>	<u>Virtanen Eva ja Voitto</u>
<u>Virtanen Anne ja Markku</u>	<u>Virtanen Eero ja Raila</u>	<u>Virtanen Eva-Liisa ja Kari</u>

Tutkija otti otoksen kolmella tavalla:

**Otos 1.** Ensimmäiseen otokseen valittiin kokonaisina sivut 57, 157, 257 ja 357. Jokaiselta sivulta laskettiin kaikkien otokseen hyväksyttävien parien määrät palstoittain. Otosyksikkönä on siis yhden sivun yksi palsta.

## VANHOJA VALINTAKOETEHTÄVIÄ

**Otos 2.** Virtanen on Helsingin puhelinluettelon yleisin nimi. Mukaan otettiin kaikki Virtaset. Otosyksikkönä on yhden sivun yksi palsta. (Mukaan otettiin vain ne palstat, joissa oli koko palstan mitan Virtasia.)

**Otos 3a ja 3b.** Otoksen 3a ensimmäinen otantayksikkö muodostui seuraavasti: Otettiin sivuilta 60–69 jokaisen palstan ensimmäinen pariskunta. Täten otokseen tuli 30 pariskuntaa. Toiseen otantayksikköön otettiin vastaavasti sivuilta 160–169 jokaiselta palstalta ensimmäinen pariskunta. Näin tehtiin edeten aina 100 sivua eteenpäin, niin että viimeisenä otettiin sivuilta 1060–1069 jokaisen palstan ensimmäiset parit. Otos 3b otettiin vastaavalla tavalla kuin otos 3a, mutta aloitettiin sivulta 70–79 ja edettiin aina 100 sivua eteenpäin.

**Taulukko 1.** Taulukossa on esitetty edellä mainituilla tavoilla poimitut otokset. Taulukossa on ensiksi poimintasivu (sivu) (otoksissa 3a ja 3b sivut), palstan numero sivulla (palsta), otantayksikön järjestysnumero (otantayksikön numero), Niiden parien lukumäärät, joissa naisen nimi on ensimmäisenä (nainen ensin), kaikkien otantayksikköön hyväksytyjen parien lukumäärät (kaikki parit) ja niiden parien suhteellinen osuus, joissa naisen nimi on ensin (nainen ensin/kaikki).

Otos 1						Otos 2					
sivu	palsta	otanta- yksikön numero	nainen ensin	kaikki parit	nainen ensin / kaikki	sivu	palsta	otanta- yksikön numero	nainen ensin	kaikki parit	nainen ensin / kaikki
57	1	1	3	31	0,10	1080	1	1	25	36	0,69
57	2	2	20	41	0,49	1080	2	2	18	42	0,43
57	3	3	14	40	0,35	1080	3	3	14	39	0,36
157	1	4	7	20	0,35	1081	1	4	9	31	0,29
157	2	5	9	28	0,32	1081	2	5	15	42	0,36
157	3	6	7	23	0,30	1081	3	6	5	47	0,11
257	1	7	19	43	0,44	1082	1	7	7	31	0,23
257	2	8	13	33	0,39	1082	2	8	17	25	0,68
257	3	9	13	35	0,37	1082	3	9	8	33	0,24
357	1	10	13	34	0,38	1083	1	10	5	40	0,13
357	2	11	11	37	0,30	1083	2	11	15	42	0,36
357	3	12	7	26	0,27	1083	3	12	14	36	0,39
						1084	1	13	15	34	0,44
						1084	2	14	6	23	0,26
yht:			136	391	0,35	yht:			173	501	0,35

Otos 3a					Otos 3b				
sivut	otanta- yksikön numero	nainen ensin	kaikki parit	nainen ensin / kaikki	sivut	otanta- yksikön numero	nainen ensin	kaikki parit	nainen ensin / kaikki
60 - 69	1	16	30	0,53	70 - 79	1	15	30	0,50
160 - 169	2	12	30	0,40	170 - 179	2	10	30	0,33
260 - 269	3	15	30	0,50	270 - 279	3	8	30	0,27
360 - 369	4	11	30	0,37	370 - 379	4	10	30	0,33
460 - 469	5	10	30	0,33	470 - 479	5	10	30	0,33
560 - 569	6	14	30	0,47	570 - 579	6	13	30	0,43
660 - 669	7	9	30	0,30	670 - 679	7	12	30	0,40
760 - 769	8	12	30	0,40	770 - 779	8	15	30	0,50
860 - 869	9	12	30	0,40	870 - 879	9	14	30	0,47
960 - 969	10	10	30	0,33	970 - 979	10	12	30	0,40
1060 - 1069	11	14	30	0,47	1070 - 1079	11	10	30	0,33
yht:		135	330	0,41	yht:		129	330	0,39

## VANHOJA VALINTAKOETEHTÄVIÄ

- a) Piirrä naisten suhteellista osuutta kuvaavat frekvenssikäyrät otantayksiköittäin kahteen kuvioon seuraavasti: toiseen kuvioon otos 1 yhtenäisellä viivalla ja otos 2 katkoviivalla, toiseen kuvioon otos 3a yhtenäisellä viivalla ja otos 3b katkoviivalla. X-akselille tulee otantayksikön numero, Y-akselille suhteellinen osuus.
- b) Laske otoksille 3a ja 3b 'nainen ensin' -pariskuntien suhteellisten osuuksien luottamusväli yhdelle otantayksikölle (N=30). Käytä suhteellisten osuuksien testiä, p:n arvona voit käyttää 0.4. Piirrä tämä luottamusväli molempien kuvioden reunaan.
- c) Mikä käytetyistä otantamenetelmistä vaikuttaa parhaalta? Perustele vastauksesi.
- d) Tutki voitko tehtyjen otosten perusteella 1 prosentin riskitasolla osoittaa, että joko miehen tai naisen nimi on yleisempi ensimmäisenä nimenä ja jos voit niin kumpi on yleisempi.

### **Vuosi 1992, tehtävä 3**

Tutkija osallistuu tv:n Kymppitonni-nimiseen kilpailuun. Kilpailussa on viisi osallistujaa, joista yksi aina vuorollaan tekee kysymyksen ja neljä muuta yrittävät arvata, mikä on oikea vastaus. Jos arvaajista ei kukaan tai kaikki neljä arvaavat oikein, tulee kysyjälle 500 mk:n tappio. Jos kolme vastaa oikein saa kysyjä ja jokainen oikein vastannut 100 mk, jos kaksi vastaa oikein saa kysyjä ja oikein vastanneet 300 mk ja jos vain yksi vastaa oikein saavat kysyjä ja vastaaja molemmat 1000 mk. Tutkija aikoo ilmoittaa kilpailijoille vastauksen vaihtoehdot. Hän kysyy esim. ”ajattelen jotain numeroista 1, 2, 3, 4 tai 5, arvaa mitä numeroa ajattelen.” (Täten tässä tapauksessa vaihtoehtoja on viisi).

- a) Millä vaihtoehtojen määrällä tutkija minimoi sen mahdollisuuden, että hän joutuisi maksamaan 500 mk (eli joko ei kukaan tai kaikki neljä henkilöä vastaavat oikein)?
- b) Jos vaihtoehtoja on neljä (siis esim. numerot 1, 2, 3 ja 4), niin mikä on todennäköisyys, että täsmälleen kaksi henkilöä vastaa oikein?
- c) Montako vaihtoehtoa kannattaa antaa arvattavaksi, jos haluaa maksimoida todennäköisyyden, että vain yksi henkilö arvaa oikein (eikä välitetä ollenkaan siitä mahdollisuudesta, että kaikki arvaisivat väärin tai kaikki arvaisivat oikein)?

### **Vuosi 1993, tehtävä 3**

Tutkija on pitkällä junamatkalla pääsiäisen ruuhkajunassa ravintolavaunun viereisessä vaunussa. Juna on hyvin pitkä ja väkeä kulkee jatkuvasti edestakaisin. Matkan virkistykseksi lapset haluavat pelata lehdestä löytynyttä peliä, jossa tarvitaan normaalia kuusitahoista arpakuutiota, jossa jokaisella numerolla 1-6 on yhtä suuri todennäköisyys tulla valituksi. Koko pelin läpi pelaamiseen jokainen heittäjä tarvitsee korkeintaan 50 heittoa, pelaajia on kolme. Arpakuutiota ei kuitenkaan ollut ja niinpä tutkija joutuu käyttämään muita keinoja vastaavien numeroiden arpomiseen. **Miten arpominen suoritettaisiin, jos:**

- a) Tutkijalla on käytössään vuosien 1983 - 1992 lottoarvontojen tulokset. Joka vuodelta on 52 arvontaa, joissa on arvottu 7 varsinaista ja 3 lisännumeroa. Arvotut numerot ovat väliltä 1–39. Miten näitä lukuja voitaisiin käyttää arpakuution korvikkeena niin, että saisimme lukuja väliltä 1–6, joilla olisi sama jakauma kuin kuusitahoisella arpakuutiolla saaduilla luvuilla?
- b) Tutkijalla on mukanaan vain iso joukko erilaisia kolikoita. Kerro millä Tavalla erilaisia kolikoita heittämällä voidaan arpoa lukuja väliltä 1–6 niin, että todennäköisyys saada mikä tahansa luvuista 1–6 on sama kuin arpakuution heitossakin on? Huom! Käytä vastauksessasi kolikoiden puolista nimiä: klaava (numeropuoli) ja kruuna (kuvapuoli).
- c) Lasten pelaamassa pelissä seikkailija taistelee hirviötä vastaan. Todennäköisyys, että seikkailija voittaa hirviön ensimmäisellä lyönnillä, on 0,4. Ellei hän tyrmää hirviötä, niin jokaisen lyönnin jälkeen hän menettää energiaa siten, että todennäköisyys tyrmätä hirviö on pienempi kuin edellisellä lyönnillä. Ellei seikkailija voita hirviötä ensimmäisellä lyönnillä, niin tyrmäämisen todennäköisyys on enää 0,3, sitä seuraavalla lyönnillä 0,2 ja sitä seuraavalla 0,1. Ellei hän tällöinkään ole vielä tyrmännyt hirviötä, hän on hävinnyt pelin. Mikä on todennäköisyys, että seikkailija voittaa hirviön?

## Vuosi 1994

Seuraavat tehtävät käsittelevät Pitkäveto veikkauspeliä. Ohessa on mallina 6 Super-pesiksen veikkauskohdetta 15.5.94. Taulukossa ovat seuraavat muuttujat: Nr = pelikohteen numero, Viikon ottelut = ensimmäisenä mainittu kotijoukkue, toisena vierasjoukkue. Taulukossa ovat seuraavat lopputuloksen vaihtoehdot ja Veikkaus Oy:n antamat painokertoimet kyseisille vaihtoehdoille. 1 = kotijoukkueen voitto, X = tasapeli, 2 = vierasjoukkueen voitto.

Nr	Viikon ottelut	1	X	2
1	Alajärvi-Hyvinkää	2,30	2,80	2,30
2	Riihimäki-Oulu	3,90	3,05	1,50
3	Kitee-Sotkamo	4,20	2,75	1,45
4	Muhos-Loimaa	2,65	3,10	1,95
5	Pattijoki-Imatra	3,90	3,05	1,50
6	Kankaanpää-Seinäjoki	1,90	2,80	2,75

Pitkävetoa voidaan pelata eri tavoin. Seuraavassa keskitytään pelitapaan, jota kutsutaan Trioksi. **Tässä pelitavassa yhdessä pitkävetopelissä veikataan kolmea kohdetta**, ja mikäli pelaaja veikkaa kaikki kolme kohdetta oikein, niin hän voittaa asettamaansa panoksen ja kohteiden painokertoimien tulo. Painokertoimien tuloa sanotaan kokonaispainokertoimeksi. Siis jos valitaan kohteet 1, 4 ja 5, ja kaikista valitaan vaihtoehto 2, niin kokonaispainokertoimeksi tulee  $2,30 \cdot 1,95 \cdot 1,50 = 6,7275$ . Mikäli pelaaja sijoittaa peliin 100 mk ja veikkaa oikein kaikki kolme kohdetta, niin hän saa  $100 \text{ mk} \cdot 6,7275 = 672,75 \text{ mk}$ , muussa tapauksessa hän menettää sijoittamansa panoksen.

**Vuosi 1994, tehtävä 1**

Pelaaja aikoo pelata seuraavalla strategialla: Pelaaja valitsee kolme kohdetta (ei samoja kuin yllämainitussa esimerkissä). Nimitetään niitä tässä esimerkiksi kohteiksi 1, 2 ja 3. Pelaaja olettaa etukäteisinformaation perusteella, että kaikissa kohteissa vaihtoehdon 2 todennäköisyys on 0,2. Pelaaja haluaa pelata kaikki kohteiden 1, 2 ja 3 erilaiset kombinaatiot, kun vaihtoehtoina ovat 1 ja X. Veikkaaja olettaa siis, että vaihtoehto 2 ei toteudu missään kolmessa kohteessa eikä veikkaa tätä vaihtoehtoa lainkaan. Laske seuraavat tehtävät olettaen, että yllä esitetyt oletukset pitävät paikkansa. Oletetaan, että pelaaja sijoittaa 10 mk jokaiseen peliinsä. Pelaaja joutuu täten pelaamaan 2·2·2 eli 8 peliä ja sijoittamaan peleihin yhteensä 80 mk joka viikko.

Laske seuraavat tehtävät kun pelaaja käyttää yllä esitettyä pelistrategiaa ja lähtien siitä, että vaihtoehdon 2 todennäköisyys jokaisessa kohteessa todella on 0,2.

- a) Mikä on todennäköisyys, että yksi pelaajan pelaamista peleistä on oikein, kun pelaaja on pelannut kaikki 1 ja X:n erilaiset kombinaatiot?
- b) Jos kokonaispainokertoimen odotusarvo olisi 12, niin mikä olisi voiton tai tappion odotusarvo, mikäli pelaaja pelaisi 10 viikkoa tehtävän johdannossa esitetyllä pelitavalla asettaen 10 mk jokaiseen peliinsä?
- c) mikä pitäisi keskimääräisen kokonaispainokertoimen odotusarvon olla, jotta tämän tehtävän johdannossa esitetty pelitapa olisi pitkällä tähtäimellä kannattavaa?

**Vuosi 1994, tehtävä 2**

Tässä tehtävässä tutkitaan onko pelikohteiden pienimmän painokertoimen koolla ja pelin lopputuloksella riippuvuutta keskenään. Tarkastelussa tilastoyksikkönä on yksi pelikohde. Jokainen pelikohde on luokiteltu kohteen pienimmän painokertoimen ja pelikohteen lopputuloksen mukaisesti eri luokkiin. Jokaisesta pelikohteesta etsitään ensiksi millä vaihtoehdolla 1, X tai 2 on pienin painokerroin. Pienin painokerroin vaihtelee väliltä 1,1–2,3 ja pelikohde on luokiteltu pienimmän painokertoimen mukaan joko luokkaan 1,1–1,45; 1,5–1,95 tai 2 - 2,30.

Muuttujan ”Pelikohteen lopputulos” arvot määräytyvät pelikohteen lopputuloksen mukaan seuraavasti. Jos tilastoyksikkönä olevan pelikohteen ottelu päättyy niin, että pienimmän painokertoimen omaava vaihtoehto toteutuu, niin pelikohde luokitellaan tapaukseksi: ”Pienin painokerroin toteutui” muuten tapaukseksi: ”Pienin painokerroin ei toteutunut”.

Tutki  $\chi^2$ -riippumattomuustestillä pienimmän painokertoimen koon ja pelin lopputuloksen välistä yhteyttä. Laskutoimitusten helpottamiseksi yhdistä muuttujan ”pelikohteen pienin painokerroin” luokat siten, että pidät luokan 1,1–1,45 omana luokkana ja yhdistä luokat 1,5–1,95 ja 2–2,3 yhdeksi luokaksi nimeltään ”vähintään 1,5”. Voit pyöristää laskutoimituksissa syntyvät luvut kokonaisluvuiksi.

**VANHOJA VALINTAKOETEHTÄVIÄ**

Taulukossa on 34 Super-pesis ottelun tulokset luokiteltuna pelikohteen pienimmän painokertoimen ja pelin lopputuloksen mukaisesti.

		Pelikohteen lopputulos		Yhteensä
		Pienin paino- kerroin ei toteutunut	Pienin paino- kerroin toteutui	
Pelikohteen pienin painokerroin				
1,1–1,45	F	3	8	11
	C%F	14,29	61,54	32,35
	R%F	27,27	72,73	100,00
	T%F	8,82	23,53	32,35
1,50–1,95	F	13	5	18
	C%F	61,90	38,46	52,94
	R%F	72,22	27,78	100,00
	T%F	38,24	14,71	52,94
2,00–2,20	F	5	0	5
	C%F	23,81	0	14,71
	R%F	100,00	0	100,00
	T%F	14,71	0	14,71
Yhteensä	F	21	13	34
	C%F	100,00	100,00	100,00
	R%F	61,76	38,24	100,00
	T%F	61,76	38,284	100,00

Taulukossa F=solufrekvenssi, C%F=sarakeprosentit, R%F=riviprocentit, T%F=solufrekvenssien prosenttiosuus kokonaisfrekvensseistä.

- Kirjoita tämän tehtävän  $H_0$  -hypoteesi ja  $H_1$  -hypoteesi.
- Laske  $\chi^2$  -riippumattomuustestin tulos ja testaa tuloksen merkitsevyys 1 % riskitasolla. Pane välitulokset näkyviin, merkitse vapausasteet, testin tulos, 1 % kriittinen raja  $\chi^2$  -testissä. Mikä on johtopäätöksesi testituloksen perusteella?
- Kun tarkastelet  $\chi^2$  -testin tulosta ja taulukon 2 frekvenssejä, niin millä edellytyksin voisit tämän otoksen perusteella pitää kohteen pienintä painokerrointa parhaana ennusteena lopputulokselle? Kerro, mihin taulukon lukuihin vastauksesi nojautuu.

**Vuosi 1994, tehtävä 3**

- a) Kahden muuttujan välinen korrelaatiokerroin on 0,4 ja havaintojen lukumäärä on 27. Testaa poikkeako korrelaatiokeroin nollassa 5 % riskitasolla. (Voit käyttää laskuissa yhden desimaalin tarkkuutta).
- b) Nimeä muuttujien mittaustason mukaiset neljä mitta-asteikkoa alkeellisimmasta vaativimpaan. Asteikko on sitä vaativampi, mitä useampi toimenpide on sillä sallittu. Merkitse taulukkoon 3 kohtaan ”asteikkotyyppi” jokaisen muuttujan kohdalle oikean asteikkotyypin numero yllä kirjoittamastasi asteikkotyyppien listasta.
- c) Jos noudatetaan eri korrelaatiokertoimien oletusten mukaisia mitta-asteikko vaatimuksia, niin merkitse jokaisen muuttujan kohdalle voiko muuttujaa käyttää vain Pearsonin korrelaatiokertoimen (P), vain Spearman-Brownin korrelaatiokertoimen (S), tai molempien korrelaatiokertoimien (S+P) tai ei minkäänlaisen korrelaatiokertoimen (0) laskemiseen. Merkitse kohtaan ”korr. Laskeminen” suluissa merkitty vaihtoehto, siis joko P, S, S+P tai 0.

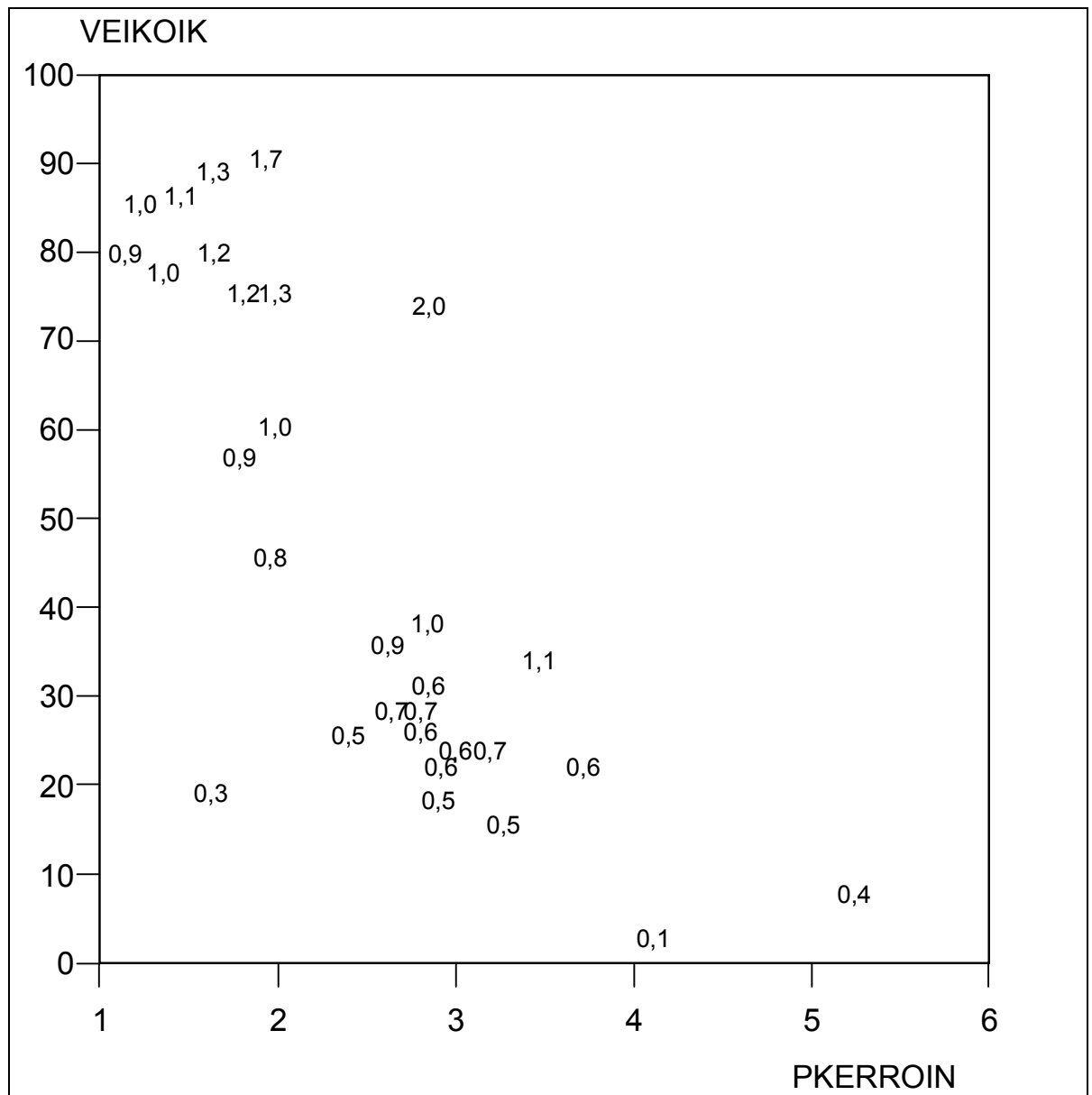
Seuraavassa taulukossa on muuttujaluettelo aineistosta, jossa tilastoyksikkönä on yksi pelikohde. Jokaisesta pelikohteesta tiedetään luettelossa ilmoitetut asiat. Jokaiselle muuttujalle on annettu nimi ja jokaiselle muuttujalle on selitysteksti, joka kuvaa millaisia arvoja muuttuja voi saada. Täytä selitystekstien tietojen perusteella kohdat ”asteikkotyyppi” ja ”korr. laskeminen” edellä esitettyjen ohjeiden mukaisesti.

asteikko- tyyppi	korr. laske- minen	Muuttujan nimi	Muuttujan selitysteksti
_____	_____	PENJÄRJ	Ennustuksen onnistuminen. Numerokoodi määräytyy pelituloksen mukaan seuraavasti. 1 jos toteutui se vaihtoehto, jolla oli pienin painokerroin 2 jos toteutui se vaihtoehto, jolla oli keskimäinen painokerroin 3 jos toteutui se vaihtoehto, jolla oli suurin painokerroin
_____	_____	TULOS	Toteutunut vaihtoehto pitkävedossa koodattuna seuraavasti: 1, X tai 2
_____	_____	PKERROIN	Toteutunut pitkävedon painokerroin: kerroin 1,10–5,2 (eli pelin lopputuloksen mukaan määräytynyt painokerroin)
_____	_____	VEIKOIK	Oikeaa vaihtoehtoa veikanneiden % -osuus kaikista niistä, jotka olivat veikanneet kyseistä kohdetta. Muuttujan minimiarvo on 0, jos kukaan heistä ei ole arvannut oikeaa vaihtoehtoa ja maksimi 100, jos he kaikki ovat arvanneet oikean vaihtoehdon.
_____	_____	X5	Muuttuja, joka on muodostettu edellisistä muuttujista seuraavalla kaavalla: $X5 = (PKERROIN^3 * (VEIKOIK/100)^3)^{1/3}$ HUOM! ^=Potenssin merkki, *=kertomerkki, ^(1/3)=kuutiojuuri
_____	_____	PVM	Päivämäärä muodossa kkpvm, esim. 22. päivä toukokuuta = 0522
_____	_____	KOTI	kotijoukkueen nimi
_____	_____	VIERAS	Vierasjoukkueen nimi

**VANHOJA VALINTAKOETEHTÄVIÄ**

Alla on esimerkin vuoksi tulostettu kolmen pelikohteen tulokset.

PENJÄRJ	TULOS	PKERROIN	VEIKOIK	X5	PVM	KOTTI	VIERAS
1	2	1,90	90	1,7	0508	Kankaanpää	Alajärvi
3	X	2,75	18	0,5	0508	Imatra	Oulu
1	1	1,75	75	1,3	0508	Kitee	Vimpeli



Kuviossa on 30 pelikohteen aineistosta tehty hajontakuviio. Vaaka-akselilla on pelin lopputuloksen mukaisen painokertoimen arvot (muuttujan nimi PKERROIN). Pystyakselilla on muuttujan VEIKOIK arvot. Tämän muuttujan arvot kuvaavat oikeaa vaihtoehtoa veikanneiden % -osuutta kaikista kyseistä kohdetta veikanneista. Havaintopisteiden tilalla kuviossa on jokaisesta tilastoyksiköstä muuttujan X5 arvo.

## VANHOJA VALINTAKOETEHTÄVIÄ

Muuttujien väliset korrelaatiot. Havaintojen lukumäärä otoksessa 30. (HUOM! Osa korrelaatioista voi olla järjettömiä.)

	PENJÄR	TULOS	PKERROIN	VEIKOIK	X5	PVM
PENJÄR	1,00					
TULOS	-0,14	1,00				
PKERROIN	0,80	-0,04	1,00			
VEIKOIK	-0,70	0,35	-0,76	1,00		
X5	-0,29	0,42	-0,42	0,80	1,00	
PVM	0,02	0,05	-0,05	0,05	0,22	1,00

Huom! Vain oikein perustelluista vastauksista saa pisteitä.

d) **VEIKOIK ja PKERROIN** muuttujien korrelaatiokeroin on otoksessa -0,76. **Onko voimakas negatiivinen korrelaatio hyvä vai huono veikkausyhtiön tuoton kannalta? Ympyröi oikea vaihtoehto:**

- 1) hyvä
- 2) huono

Perustele vastauksesi.

e) Taulukossa 3 on kerrottu kuinka muuttuja X5 on muodostettu. Mitä muuttuja X5 kuvaa? Mitä peliyhtiön kannalta tarkoittaa se, että muuttuja X5 saa suuren arvon?

f) Jos oletetaan, että kyseinen 30 pelikohteen otos kuvaisi luotettavasti muuttujien välisiä korrelaatioita, niin kumpi on yleensä peliyhtiölle edullisempaa? Sekö, että pelikohteesta toteutuu sellainen vaihtoehto, jonka painokerroin on pieni vaiko sellainen, jonka painokerroin on suuri. Ympyröi oikea vaihtoehto.

- 1) pieni
- 2) suuri

Perustele vastauksesi käyttäen apuna tässä tehtävässä annettuja tuloksia.

### Vuosi 1995, tehtävä 3

Kirkoissa on virsiä osoittavat numerotaulut. Numerotauluissa on paikat messussa veisattaville kuudelle virrelle. Oletetaan, että Suomen 400 kirkossa muutetaan virsiä osoittavat peltiset numerot muovisiksi. Rahojen säästämiseksi arkkipiispa antaa määräyksen, että mihinkään näistä kirkoista ei yhtä numerotaulua kohti saa hankkia enempää kuin 11 muovinumeroa jokaista numeroa 1–9 kohti. Siis esimerkiksi numeroa 1 osoittavia muovinumeroita hankitaan 11 jokaiseen numerotauluun, mutta ei enempää.

Oletetaan, että seuraavat kuvitteelliset ehdot ovat voimassa.

-Virsikirjassa on 632 virttä

-Jokaisessa messussa soitetaan 6 virttä

-Kaikissa kirkoissa pidetään 52 messua vuodessa.

-**Normaalimessu** (käytä myöhemmin lyhennettä NM) pidetään 51 viikkona vuodesta. Tällöin kaikissa kirkoissa virret valitaan arpomalla kaikista virsistä 1–632, kuitenkin virttä

111 ei oteta mukaan arvontaan. Yhdessä messussa ei samaa virttä valita laulettavaksi kahta tai useampaa kertaa samassa tilaisuudessa. Sen sijaan kaikissa erillisissä messuissa arvonta on riippumaton toisten messujen arvonnoista.

-**Erikoismessu** pidetään jokaisessa kirkossa kerran vuodessa (käytä myöhemmin lyhennettä EM). Tällöin valitaan virret jälleen arpomalla virsistä 1–632, kuitenkin virttä 111 ei oteta mukaan arvontaan. Tällä kertaa myös yhden messun sisällä jokaisen vireen arpominen on riippumaton edellisistä arvonnoista. Tämä tarkoittaa sitä, että yhdessä messussa sama virsi voidaan laulaa useammin kuin kerran. Muuten arvontojen riippumattomuussäännöt ovat samat kuin normaalimessussa.

**Laske todennäköisyys, että 10 vuoden aikana ainakin yhdessä messussa missä tahansa 400 kirkosta numeroa yksi osoittavat muovinumerot loppuvat kesken, kun yllä mainitut oletukset ovat voimassa.**

Seuraavassa on esimerkki yhdessä messussa laulettavista virsistä. Numerotauluun on merkitty laulettaviksi seuraavat virret: 561, 511, 501, 338, 388, 341. Tällöin olisi käytössä viisi kappaletta numeroa yksi osoittavia muovinumeroita.

Täydet pisteet saa sellaisesta ratkaisusta, jossa on esitetty kaikki tarpeelliset kaavat välivaiheineen ja luvut on sijoitettu välivaiheen kaavoihin. Numerovastausta ei tarvitse laskea (esimerkiksi  $1,77/156$  kelpaa vastaukseksi).

Vain sellainen ratkaisu hyväksytään, jossa on avattu kaikki kertoma- ja binomilausekkeet. Siis välivaiheet tulee työstää niin pitkälle, että niissä esiintyy vain plus-, miinus-, kerto-, jako-, potenssi- tai neliöjuurilausekkeitä tarpeen mukaan. Huomaa, että mikäli osa laskun välivaiheista on oikein saa osan pisteistä, vaikka lopullinen vastaus ei olisikaan oikea.

## **Vuosi 1996, tehtävä 2**

Eräässä tietokonepelissä nimeltään ”Matopeli” Veikko on huippupelaaja. Hänen ennätöksensä on 360. Pelissä täytetään ruudun pinta-alaa ja loppua kohti liikkuminen ruudulla tulee mahdottomaksi, minkä vuoksi siinä sen korkeampia pistemääriä ei juuri voi saada. Sen sijaan hyvälläkin pelaajalla voivat jotkut pelit epäonnistua täysin. Seuraavassa Veikon Matopelin pisteet 50 pelissä ja PISTEMÄÄRÄ-muuttujan tunnusluvut kyseisessä 50 pelin otoksessa.

### **Veikon matopelin pisteet 50 pelissä**

Pistemäärät lajiteltuina pienimmästä suurimpaan.

034, 048, 069, 099, 123, 140, 159, 166, 169, 197, 200, 203, 219, 219, 221, 243, 246, 248, 250, 253, 256, 257, 259, 263, 269, 270, 272, 275, 285, 285, 287, 292, 293, 301, 302, 304, 308, 310, 312, 314, 316, 320, 325, 327, 351, 351, 352

Muuttujan ”PISTEMÄÄRÄ” tunnusluvut ja desiilit tietokonetulosteena ohessa. Tunnusluvuista näkyvät mm. 50 luvun otoksesta pienin pistemäärä (min), maksimipistemäärä (max), keskiarvo (mean), keskihajonta (stddev), vinous (skewness),

## VANHOJA VALINTAKOETEHTÄVIÄ

huipukkuus (kurtosis), alakvartiili (lower\_Q), mediaani (median), yläkvartiili (upper\_Q) sekä desilit (esim. fractile(.1) = 10 % desiili).

```
Basic statistics: MATOPELI N=50
Variable: PISTEMÄÄRÄ
min=34 in obs.#47 (034)          max=352 in obs.#10 (352)
mean=245.3  stddev=77.77105      skewness=-1.025225
                                          kurtosis=0.418562
lower_Q=206.6667  median=262  upper_Q=304.4444
fractile(.1)=135  fractile(.6)=282
fractile(.2)=185  fractile(.7)=298
fractile(.3)=230  fractile(.8)=310
fractile(.4)=250  fractile(.9)=325
fractile(.5)=262
```

- a) Miten menettelisit, jos seuraavat ehdot olisivat voimassa: Veikko pelaa seuraavat 50 peliä siten, että PISTEMÄÄRÄN jakauma ja jakauman tunnusluvut pysyvät samankaltaisina kuin yllä esitetystä otoksessa.

Nyt ei ollakaan kiinnostuneita yksittäisen pelin pistemäärästä, vaan muuttujasta TULOS, joka muodostuu 50 pelin pistemäärien avulla jommallakummalla seuraavista vaihtoehtoisista tavoista.

$$\text{KAAVA 1)} \quad \text{TULOS} = \sum_{i=1}^{50} \left( \left( (pm_i - ka) / 100 \right)^5 \right)$$

$$\text{KAAVA 2)} \quad \text{TULOS} = \sum_{i=1}^{50} \left( \left( (pm_i - \text{median}) / 100 \right)^5 \right)$$

$$\text{KAAVA 3)} \quad \text{TULOS} = \sum_{i=1}^{50} \left( (pm_i - ka) / 100 \right)$$

Kaavoissa ovat seuraavat muuttujat:

**PISTEMÄÄRÄ** ( $pm_i$ ) tarkoittaa yhdessä pelissä saatua pistemäärää.  
ka tarkoittaa **kyseisten** 50 pelatun pistemäärän keskiarvoa.  
median tarkoittaa **kyseisten** 50 pelatun pistemäärän mediaania.  
Huomaa siis, että TULOS voidaan laskea vasta kun kaikki 50 peliä on laskettu ja näiden pelien keskiarvo tai mediaani on selvillä.

**Sinun tehtäväsi ei ole pelata tietokonepeliä, vaan ainoastaan panna rahasi likoon. Saat rahaa tai menetät rahaa sen mukaan, mikä on yllä kuvatun muuttujan TULOS pistemäärä, kun Veikko on pelannut seuraavat 50 peliä. Millä kaavalla laskettu tulos on taloudellisesti paras, mikäli seuraavan 50 pelin jakauma on samankaltainen kuin ylläesitetyn otoksen jakauma?**

## VANHOJA VALINTAKOETEHTÄVIÄ

Aseta kaavat paremmuusjärjestykseen: 1. paras, 2. Toiseksi paras 3. Huonoin. Kirjoita numerot alla oleville viivoille:

\_\_\_\_\_ KAAVA 1)

\_\_\_\_\_ KAAVA 2)

\_\_\_\_\_ KAAVA 3)

**Riittää, että perustelet vastauksesi PISTEMÄÄRÄN jakauman tunnuslukujen ja muun satunnaismuuttujan jakaumasta annetun informaation avulla.**

**b) Miten menettelisit, jos seuraavat ehdot olisivat voimassa:**

- 1) Muuttuja TULOS muodostuu 50 pistemäärän summan funktiona samoilla kaavoilla kuin tehtävässä 2a, ja muuttujan TULOS pistemäärä osoittaa voittonsi tai tappionsi kuten tehtävässä 2a.
- 2) Sinä saat itse pelata tietokonepeliä ja muuttuja TULOS määräytyy omien peliesi mukaan. Tavoitteenasi ei ole maksimoida yksittäisiä Matopelin pistemääriä ( $pm_i$ ), vaan yrität maksimoida lopputulosta, jota kuvaa muuttuja TULOS.
- 3) Olet yhtä taitava pelaaja kuin Veikko. Siis yrittäessäsi saada huippupistemäärät Matopelissä pistejakaumasi on samankaltainen kuin Veikolla.

**Pane kaavat paremmuusjärjestykseen, 1=paras, 2=toiseksi paras, 3=huonoin.**

\_\_\_\_\_ KAAVA 1)

\_\_\_\_\_ KAAVA 2)

\_\_\_\_\_ KAAVA 3)

**Saat pisteitä vain, jos osaat lyhyesti perustella vastauksesi kaavojen jakaumien tilastollisten ominaisuuksien avulla.**

## VANHOJA VALINTAKOETEHTÄVIÄ

Anna esimerkkinä kahden parhaan kaavan mukaisten pistemäärien (pm) jakaumien desiiilit 50 pelin sarjassa (Huom esim. Fractile(.1) tarkoittaa jakauman 10 % desiiiliä).

	Paras kaava Kaava NR_____	Toiseksi paras kaava Kaava NR_____
Fractile(.1)	_____	_____
Fractile(.2)	_____	_____
Fractile(.3)	_____	_____
Fractile(.4)	_____	_____
Fractile(.5)	_____	_____
Fractile(.6)	_____	_____
Fractile(.7)	_____	_____
Fractile(.8)	_____	_____
Fractile(.9)	_____	_____

### Vuosi 1996, tehtävä 3

Veikko on menossa Pihlajasaaren kesäjuhille, jossa hänen pitää olla kello 19.00. Kaupungin rautatieasemalta lähtee busseja 16.00 ja 16.10. Veikko nousee rautatieasemalta kello 16.00 lähtevään bussiin. Kello 16.00 lähtevä vuoro menee vain Tammikylään, johon ajoaika minuutteina noudattaa normaalijakaumaa parametrein  $N(29,2)$ .

Tammikylästä lähtee bussi Pihlajalahden rantaan kello 16.30. Mikäli rautatieasemalta lähtenyt bussi on 16.30 Tammikylässä, niin bussiin ehtii vielä mukaan. Tällä bussilla ajoaika minuutteina Tammikylästä Pihlajalahteen noudattaa  $N(50,5)$ . Rautatieasemalta kello 16.10 lähtevä bussi menee suoraan Pihlajalahteen, pysähtyen kuitenkin Tammikylässä. Myös tällä bussilla rautatieaseman ja Tammikylän välinen ajoaika minuutteina noudattaa  $N(29,2)$  ja koko ajoaika minuutteina rautatieasemalta Pihlajalahteen  $N(80,5)$ .

Pihlajalahden rannasta lähtee vene Pihlajasaareen klo 17.30. Vene lähtee täsmällisesti ajallaan tai ei lähde lainkaan. Todennäköisyys on 0,1, että venevuoro ei lähde liikkeelle. Mikäli bussi on 17.30 rannassa, veneeseen ehtii vielä mukaan.

Venematkan kesto minuutteina Pihlajalahdesta Pihlajasaareen noudattaa normaalijakaumaa parametrein  $N(40,2)$ .

**Mikä on todennäköisyys, että Veikko ehtii ajoissa kello 19.00 alkaville Pihlajasaaren kesäjuhille?**

**Vuosi 1999, tehtävä 2**

Eräissä pelissä käytetään normaalia kuutionmuotoista arpanoppaa. Jokaisen numeron 1, 2, 3, 4, 5 ja 6 esiintymistodennäköisyys on sama. Peliin pääsee mukaan, kun saa kuutosen.

Sijoita oikeat luvut oikeisiin kaavoihin ja ilmoita todennäköisyys kahden desimaalin tarkkuudella.

a) Laske todennäköisyydet seuraaville kolmelle tapaukselle. Todennäköisyys sille, että:

- 1) kuutonen tulee ensimmäisen kerran ensimmäisellä heittokerralla.
- 2) kuutonen tulee ensimmäisen kerran toisella heittokerralla.
- 3) kuutonen tulee ensimmäisen kerran kolmannella heittokerralla.

b) Mikä on todennäköisyys sille, että ensimmäinen kuutonen tulee neljännellä heittokerralla tai myöhemmin.

c) Toistetaan seuraavaa koetta tuhat kertaa. Kun on saatu kuutonen, lasketaan kuinka monta heittoa pitää heittää, jotta saadaan seuraava kuutonen ja tämä luku kirjataan muuttujaksi "heitto kertojen lukumäärä yhtä kuutosta kohti". Mikä on 1) heitettävien heittokertojen mediaani (pyöristettynä lähimpään kokonaislukuun) ja 2) moodi yhtä kuutosta kohti? Anna luvut ja perustele vastauksesi.

d) Mikä seuraavista väitteistä pitää yllämainitussa 2 c. tilanteessa paikkansa (väitteet ovat aakkosjärjestyksessä). Tunnusluvuihin käytetään seuraavia lyhenteitä: ka = keskiarvo, me = mediaani, mo = moodi. Perustele vastauksesi.

1.  $ka < me < mo$     2.  $ka < me < mo$     3.  $me < ka < mo$
4.  $me < mo < ka$     5.  $mo < me < ka$     6.  $mo > ka > me$

Seuraavissa tehtävissä 2e ja 2f kuvitellaan, että hattuun on pistetty 12 samankokoista numerolappua. Jokaista numeroa 1 - 6 on 2 lappua. Hatusta nostetaan numerolappuja katsomatta, mitä numeroita otetaan ja panematta nostettuja lappuja uudestaan takaisin hattuun.

Vastaukseksi kohtiin 2e ja 2f riittää kaavoihin sijoitetut oikeat luvut.

e) Laske todennäköisyys sille, että:

- 1) ensimmäinen kuutonen tulee ensimmäisellä nostolla
- 2) ensimmäinen kuutonen tulee toisella nostolla
- 3) ensimmäinen kuutonen tulee kolmannella nostolla

f) Mikä on todennäköisyys sille, että ensimmäinen kuutonen tulee neljännellä nostolla tai myöhemmin.

### **Vuosi 1999, tehtävä 3**

Suomessa on Eurovaalit 13.6.1999, jossa valitaan 16 edustajaa Euroopan Unionin parlamenttiin. Ensi vuonna pidetään presidentinvaalit ja useat puolueet valitsevat oman ehdokkaansa oman puolueensa esivaalissa.

Kuvitellaan, että sinut palkataan markkinatutkimuslaitokselle tutkijaksi ja sinun pitää seuraavien tietojen perusteella vastata asiakkaan antamiin kysymyksiin. **Sinun tehtävänäsi on olla tilastotieteen asiantuntija, joka toimii tilastotieteen periaatteiden mukaisesti.**

Kuvitellaan, että eräs eurovaalien ehdokas aikoo tutkia eurovaalikampanjansa ohessa myös sitä, kuka olisi hänen puolueensa (jota tässä kutsutaan nimellä PUO) paras ehdokas ensi vuoden presidentinvaaleissa. Eurovaaliehdokas pitää 100 puhelilaisuutta, johon kaikilla ihmisillä on vapaa pääsy. tilaisuuksia pidetään arkipäivisin, iltaisin ja viikonloppuisin.

Puolueen PUO esivaali pidetään kyseisen edustajan tilaisuuksista riippumattomana tilaisuutena eräänä kesäkuun viikonloppuna. Äänioikeutettuja ovat kaikki puolueeseen kuuluvat 50 000 Suomen kansalaista ja äänestys tapahtuu puoluekokouksen yhteydessä.

- a) Euroehdokkaan pitämissä 100 puhelilaisuudessa käy keskimäärin 20 henkeä tilaisuutta kohti. Jokainen saa vastata nimettömänä kyselyyn, jossa kysytään puolueen PUO parasta ehdokasta presidentinvaaliin. Oletetaan, että 90 % vastaa kyselyyn ja vastanneista 70 % kannattaa puolueen ehdokkaaksi presidentinvaaliin vaalipuheita pitänyttä eurovaaliehdokasta.

**Eurovaaliehdokas antaa tehtäväksesi laskea, voidaanko kyseisen aineiston perusteella 1 %:n riskitasolla päätellä, että puolueen PUO presidenttiehdokkaan esivaalissa kyseinen eurovaaliehdokas saisi yli puolet esivaalin äänistä.**

- b) Tutkija menee puolueen PUO esivaalitalaisuuteen mukaan. Puolueessa toimitetaan esivaali, jossa puolueen jäsenet saavat salaisessa äänestyksessä kirjoittaa haluamansa ehdokkaan nimen äänestyslippuun. Tämän vaalin perusteella valitaan puolueen presidenttiehdokas. Esivaalin osallistuu 2000 puolueen jäsentä.

Lopullinen äänestys toimitetaan suljettuna lippuäänestyksenä ja liput pudotetaan äänestyslaatikkoon. Äänestyksen tapahduttua tutkija saa 100 äänestyslipun otoksen laatikosta, jossa äänestysliput ovat satunnaisessa järjestyksessä. Kuvitellaan, että kyseisessä 100 äänestyslipun otoksessa 60 on äänestänyt henkilöä A presidenttiehdokkaaksi. **Saat tehtäväksesi laskea, voitko tämän perusteella päätellä 5 %:n riskitasolla, että kyseinen henkilö A valitaan puolueen PUO presidenttiehdokkaaksi.**

**Vuosi 2000, tehtävä 3**

Tutkijan asunnon luona on kahden suuren taloyhtiön biologiset roskasäiliöt. Roskasäiliöt ovat katetussa tilassa ja säiliöitä on kahdeksan. Tutkija halusi selvittää ihmisten roskiskäyttäytymistä.

Tutkija numeroi säiliöt oven viereltä lähtien koodeilla R1-R8. Säiliön täyttöasteen hän arvioi iltaisin silmävaraisesti suhdeasteikkoa 0-10 käyttäen. 0 tarkoittaa, että säiliö on tyhjä ja 10, että se on täysi. Lisäksi on käytössä 11, joka tarkoittaa, että säiliö on niin täysi, ettei kansi pysy enää kiinni.

Säiliöt tyhjennetään joka keskiviikko. Huoltomies saattaa painella roskia silloin, kun säiliön kansi ei mene enää kiinni. Jos tutkija näkee, että jokin roskasäiliö on täysi, hän panee numeron 10. Kuitenkin tämä roskasäiliö on voinut olla tilassa 11 saman päivän aikana, mutta talonmies on survonut roskia niin, että säiliön kansi menee kiinni. Oletetaan, että talonmies painelee säiliöitä vain silloin, kun kansi ei mene kiinni.

Tässä oletetaan, että roskien viejät jättävät roskansa aina siihen säiliöön, jonka sattuvat ensimmäisenä avaamaan, jos säiliö ei ole vielä täysi. Mikäli säiliö on täysi, niin jotkut roskien viejät jättävät roskansa ensin valitsemaansa säiliöön, jolloin kansi ei pysy kiinni. Tässä tilanteessa toiset etsivät sellaisen säiliön, joka ei ole vielä täysi. Lisäksi oletetaan, että sen jälkeen kun kansi ei enää pysy kiinni, kaikki roskien viejät panevat roskansa johonkin muuhun säiliöön.

**HYPOTEESI: Tutkija on kiinnostunut siitä, onko säiliöiden täytyminen satunnaista silloin, kun kaikki roskien viejät voivat pudottaa roskansa siihen säiliöön, jonka ensimmäisenä valitsevat.**

Tutkija poimii aineistosta sellaiset havaintonäytteet, jotka täyttävät kohdassa 3a. esitetyt ehdot. Sen jälkeen hän muodostaa indeksin, jonka hän saa laskemalla valitsemistaan havaintonäytteistä pistemäärien summan jokaiselle roskasäiliölle. Sen jälkeen tutkija testaa sitä, voidaanko olettaa, että ihmiset pistävät roskansa täysin satunnaisesti mihin tahansa roskasäiliöistä R1-R8.

Seuraavassa yhtenä havaintonäytteenä pidetään tutkijan tiettyä iltana tekemää tilastoa, jossa on 8 lukua kertoen jokaisen säiliön täyttöasteen kyseisenä iltana. Jokaisella havaintonäytteellä on oma järjestysnumerosa (muuttuja NR).

Ohessa on taulukko, josta näkyy tarkastuspäivämäärä ja säiliöiden tila. Huomaa! Tutkija ei ole päässyt joka päivä tarkastamaan roskatynnyreitä.

**VANHOJA VALINTAKOETEHTÄVIÄ**

**Roskisten täyttöasteet**

pvm	viikon päivä	NR	R1	R2	R3	R4	R5	R6	R7	R8	perustelut:
2.3.	to	<b>1</b>	3	5	4	2	3	0	0	2	
3.3.	pe	<b>2</b>	4	7	7	3	6	3	0	2	
6.3.	ma	<b>3</b>	6	9	11	8	10	9	1	1	
10.3.	pe	<b>4</b>	1	1	7	1	3	3	0	2	
11.3.	la	<b>5</b>	1	1	9	2	4	6	1	3	
12.3.	su	<b>6</b>	3	3	10	5	9	9	1	3	
13.3.	ma	<b>7</b>	5	8	10	5	9	9	1	3	
14.3.	ti	<b>8</b>	6	11	10	6	9	10	3	3	
15.3.	ke	<b>9</b>	1	1	2	1	1	1	0	0	
16.3.	to	<b>10</b>	1	3	6	1	1	3	0	0	
17.3.	pe	<b>11</b>	2	6	6	1	2	9	0	0	
18.3.	la	<b>12</b>	2	8	9	3	3	5	1	1	
19.3.	su	<b>13</b>	3	11	10	4	5	7	1	1	
20.3.	ma	<b>14</b>	4	10	11	6	6	10	1	1	
21.3.	ti	<b>15</b>	4	11	11	6	9	10	3	3	
22.3.	ke	<b>16</b>	0	1	2	0	1	0	0	1	
23.3.	to	<b>17</b>	1	1	4	0	2	1	0	1	
24.3.	pe	<b>18</b>	1	6	5	2	3	2	1	1	
25.3.	la	<b>19</b>	6	9	8	4	4	4	1	1	
26.3.	su	<b>20</b>	7	11	9	6	5	6	2	1	
27.3.	ma	<b>21</b>	7	11	9	6	5	6	2	1	
28.3.	ti	<b>22</b>	7	11	9	7	7	7	3	3	
9.4.	su	<b>23</b>	5	8	3	8	10	6	2	0	
10.4.	ma	<b>24</b>	6	11	5	8	10	6	2	0	
11.4.	ti	<b>25</b>	9	9	7	10	10	8	3	1	
12.4.	ke	<b>26</b>	2	1	0	0	0	1	0	0	
25.4.	ti	<b>27</b>	7	9	11	8	8	8	5	1	

- a) Valitse taulukosta kaikki sellaiset havaintonäytteet, jotka tutkija voi valita muodostaessaan indeksinsä. Havaintonäytteiden tulee kuitenkin täyttää seuraavat kolme ehtoa: 1) havaintonäytteet ovat toisistaan riippumattomia. 2) Havaintonäytteet ovat sellaisia, joista tutkija voi olla varma, että ne täyttävät ne edellytykset, joita edellisellä sivulla vahvennetulla olevan hypoteesin tutkiminen edellyttää. 3) Kyseiseen indeksiin poimittavien havaintonäytteiden tulee olla lisäksi sellaisia, että ne kuvaavat mahdollisimman edustavasti eri säiliöiden käyttöä.

**Merkitse valitsemasi havaintonäytteet ympyröimällä kyseisen näytteen järjestysnumero (muuttuja NR). Perustele valintasi. Kirjoita perustelut taulukon 2 viereen.**

- b) Oletetaan, että alla taulukossa on eräs indeksi, jonka tutkija on muodostanut yllä olevan aineiston perusteella (Indeksissä lukuja on muokattu niin, että laskeminen olisi helpompaa). Testaa, voidaanko olettaa ihmisten jättävän roskansa täysin satunnaisesti mihin tahansa roskasäiliöistä R1-R8 kyseiseen indeksiin poimitun otoksen perusteella. Kerro myös johtopäätöksesi kyseisen tuloksen perusteella.

Indeksi eri roskasäiliöiden käytölle:

R1	R2	R3	R4	R5	R6	R7	R8	Yht
14	24	30	11	15	17	3	6	120

### Vuosi 2001, tehtävä 1

Tutkijat olivat kiinnostuneita ihmisten käyttäytymisen muutoksista. He kävivät YLE:n nauha-arkistossa tutkimassa presidentinlinnan itsenäisyyspäivän kutsujen tervehtimis-seremonioita eri vuosilta. Tutkijoilla oli kiinnostuksen kohteena se, onko naisten niaaminen tervehtiessä ollut yhtä yleistä eri vuosikymmenillä ja eri presidenttien aikoina. Ohessa on tutkijoiden tekemien otosten tulokset. Käyttöön saadut nauhoitukset olivat eripituisia ja itse kättelytilannetta oli kuvattu eri vuosina eri tavoin (mm. kuvakulmat, etäisyys). Nauhoja ei ollut saatavissa kaikilta vuosilta ja joiltakin vuosilta nauhat eivät olleet tarkoitukseemme sopivia. Tavoitteena oli saada keskenään vertailukelpoiset otokset (noin 10 minuutin kättelyjakso jokaiselta otokseen valitulta vuodelta). Tutkittavana muuttujana on naisten käyttäytyminen kättelytilanteessa. Kaikki naispuoliset kättelijät luokiteltiin kahteen luokkaan: niaajat ja ne jotka eivät niaanneet.

Presidenttinä toimi linnan kutsuilla 1956–1980 Urho Kekkonen, otokset vuosilta 1967, 1968 ja 1980, Mauno Koivisto 1981–1993, otokset vuosilta 1986, 1992, 1993, Martti Ahtisaari 1994–1999, otokset vuosilta 1994, 1999 ja Tarja Halonen vuonna 2000.

- a) Aineistosta on tehty uusi taulukko summaamalla kaikki otokseen kuuluneet kättelyt presidentin mukaan.

PRESIDENTTI	Niaajat	Kaikki kättelijät	Niaajien %-osuus kaikista
KEKKONEN	92	323	28
KOIVISTO	72	497	14
AHTISAARI	52	394	13
HALONEN	14	99	14

Testaa voidaanko oheisen aineiston perusteella sanoa, että niiaaminen on ollut yhtä yleistä kaikkien presidenttien aikana.

- b) Testaa kahden suhteellisen osuuden testaamiseen tarkoitettulla testillä, onko presidentti Kekkonen viimeisenä vuonna 1980 ja otoksessa olevana ensimmäisenä Koiviston vuonna 1986 niiaamisessa eroa.

Vuosi	Niiaajat	Kaikki kättelijät	Niiaajien%-osuus kaikista
1980	13	54	24
1986	18	127	14

### Vuosi 2001, tehtävä 2

Tutkijat havaitsivat, että nuoret naiset niiasivat eniten. Näytti siltä kuin edellä kulkevan nuoren käyttäytymisellä olisi ollut vaikutusta jäljessä tuleviin. Jos useita nuoria naisia meni peräkkäin ja ensimmäinen niiasi, niin myös jäljessä olevat niiasivat. Seuraavassa tästä aihepiiristä johdettu kuviteltu tehtävä (luvut eivät vastaa todellisuutta). Kuvitellaan, että tarkkaillaan presidentinlinnan kättelytilanteessa käyttäytyviä nuoria naisia vuoden 1967 presidentinjuhlissa.

Oletetaan seuraavien sääntöjen pitävän paikkansa: Jos edellä oleva ei ole niiannut, niin niiaamisen todennäköisyys on 0.5. Jos edellä oleva on niiannut, niin perässätulijan todennäköisyys niiaata on 0.8. Kuvitellaan, että kolme nuorta naista tulee peräkkäin kättelytilanteeseen ja ennen heitä on tullut useita miehiä.

- Mikä on todennäköisyys sille, että vähintään yksi näistä kolmesta niiaa?
- Mikä on todennäköisyys sille, että keskimääräinen niiaa?
- Mikä on todennäköisyys sille, että ensimmäisenä tullut on niiannut, jos tiedetään että kolmantena tullut on niiannut.

### Vuosi 2002, tehtävä 1

Useat Euroopan unionin maat siirtyivät yhteisen rahayksikön, euron käyttöön 1.1.2002. Yhden euron kolikossa on toinen puoli (klaava) kaikissa maissa samanlainen, mutta toisella puolella (kruuna) on maakohtainen tunnus.

Euroopan unionissa on noussut huoli, että erilaisesta kuvioinnista johtuen eri maiden yhden euron kolikoilla kruunan ja klaavan saamisen todennäköisyys ei olisi sama.

**1.1.** Unioni on kerännyt kahden maan (Kreikka ja Saksa) yhden euron kolikoilla suoritettujen rahanheittojen tulokset. Kunkin maan kolikolla tehtyjen heittojen kokonaismäärät eroavat. Tehtäväsi on arvioida ”heittokokeita” tilastollisten periaatteiden mukaan ja tehdä niistä johtopäätökset ja perustella ne laskemalla. **Tutki ja testaa siis sitä, onko ko. maan eurolla kruunan ja klaavan saamisen todennäköisyys yhtä suuri. Käytä 5 % merkitsevyytasoa.**

## VANHOJA VALINTAKOETEHTÄVIÄ

- a) Kreikan euroa heitettiin 6 kertaa. Tulos oli 2 kruunaa, 4 klaavaa.
- b) Saksan euroa heitettiin 2 500 kertaa. Tulos oli 1 300 kruunaa, 1 200 klaavaa.

**1.2.** EU:n selvityksissä ilmeni, että belgialaisella yhden euron kolikolla kruunan saamisen todennäköisyys on 0,52 ja hollantilaisella 0,48. **Muodosta todennäköisyysjakauma kruunien lukumäärälle satunnaiskokeessa, jossa heitetään samanaikaisesti kaksi kappaletta belgialaisia yhden euron kolikkoja ja kaksi kappaletta hollantilaisia yhden euron kolikkoja. Esitä jakauma myös graafisesti.**

**1.3.** Euroopan unioni jatkoi selvitystään keräämällä aineiston Ranskasta ja Espanjasta. Ranskalaista yhden euron kolikkoa heitettiin 5 000 kertaa ja saatiin 2 530 kruunaa. Espanjalaista yhden euron kolikkoa heitettiin myös 5 000 kertaa ja saatiin 2 480 kruunaa. **Testaa 10 % merkitsevyystasolla hypoteesi, jonka mukaan kruunien osuus on sama.**

## Vuosi 2002, tehtävä 2

Pienimmän neliösumman menetelmällä on laadittu regressiomalli, jossa y-muuttujaa ennustettiin kahdellakymmenellä x-muuttujalla ( $x_1, x_2, \dots, x_{20}$ ) eli malli on muotoa:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_{20}x_{20}$$

Alla olevassa taulukossa on mallin regressiokertoimet ja kunkin regressiokertoimen t-arvo. Malli laskettiin satunnaisotoksesta, jossa oli 121 havaintoa. Jokaisen x-muuttujan keskiarvo on 0 ja keskihajonta 1. Pienin kahden x-muuttujan välinen korrelaatio on -0,20 ja suurin on 0,30. Y-muuttujan keskiarvo on 5 ja hajonta 15.

	$b_0$	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$	$b_7$	$b_8$	$b_9$	$b_{10}$
	5,23	2,96	-3,13	2,01	2,19	-2,27	2,24	2,01	-1,74	-4,53	-3,69
t-arvo	5,00	2,31	-2,81	1,91	1,97	-1,97	2,11	1,83	-1,77	-4,05	-3,92
	$b_{11}$	$b_{12}$	$b_{13}$	$b_{14}$	$b_{15}$	$b_{16}$	$b_{17}$	$b_{18}$	$b_{19}$	$b_{20}$	
	3,72	-2,20	-1,97	2,25	-2,56	2,56	2,36	-3,07	-1,97	3,06	
t-arvo	3,70	-2,05	-1,84	2,10	-2,50	2,46	2,19	-2,93	-1,92	3,02	

**2.1. Mitkä muuttujat voisit jättää pois mallista, jos käytät poisjättämisen kriteerinä p-arvoa 0,1? Perustele.**

**2.2.** Edellä laskettua regressiomallia käytetään ennustamaan kahden uuden havainnon (havainnot H1 ja H2) y-muuttujan arvoa ( $\hat{y}$ ). Havainto H1 saa muuttujalla  $x_1$  arvon 1, muuttujalla  $x_2$  arvon 2 ja muuttujalla  $x_3$  arvon 3. Havainnolla H2 arvot ovat  $x_1 = 3$ ,  $x_2 = 1$  ja  $x_3 = 2$ .

## VANHOJA VALINTAKOETEHTÄVIÄ

Muiden  $x$ -muuttujien arvoista havainnoilla  $H1$  ja  $H2$  sinulla on seuraavat tiedot:

- ne ovat joko  $-3, -2, -1, 0, 1, 2$  tai  $3$
- kunkin  $x$ -muuttujan ( $x_4, x_5, \dots, x_{20}$ ) arvo havainnolla  $H1$  on yhtä suuri kuin havainnolla  $H2$ . Siis jos esim. muuttujan  $x_4$  arvo havainnolla  $H1$  on  $2$ , niin muuttujan  $x_4$  arvo havainnolla  $H2$  on myös  $2$ .

Lisäksi tiedät, että:

- kaikkien  $x$ -muuttujien keskiarvo on  $0$  ja keskihajonta  $1$
- pienin kahden  $x$ -muuttujan välinen korrelaatio on  $-0,20$  ja suurin on  $0,30$
- $y$ -muuttujan keskiarvo on  $5$  ja hajonta  $15$

**Pystytkö näiden tietojen perusteella selvittämään, kumpi havainnoista,  $H1$  vai  $H2$ , saa suuremman arvon ennustemuuttujalla  $\hat{y}$ ? (Merkitse rasti oikean vaihtoehdon kohdalle ja perustele vastauksesi.)**

Kyllä: \_\_\_\_\_ Ei: \_\_\_\_\_

**Perustelut:**

### **Vuosi 2003, tehtävä 1**

Tämä tehtävä on saanut virikkeen Kiinassa käytetystä syntyvyyden säätelymallista. Mallia on muokattu niin, että se sopisi koekysymykseksi. Oletetaan, että seuraavat ehdot pätevät.

Jossakin eristyneessä kylässä on  $120$  perhettä, joiden vanhemmat ovat suunnilleen saman ikäisiä ja ovat sellaisessa iässä, että voivat hankkia lapsia. Pojan ja tytön saamisen todennäköisyys on yhtä suuri. Kaikki perheet saavat lapsia ja hankkivat lapsia sallitun maksimimäärän seuraavien ehtojen mukaisesti:

- 1) Mikäli perheen ensimmäinen lapsi on poika, toista lasta ei saa hankkia.
- 2) Mikäli perheen ensimmäinen lapsi on tyttö, perhe hankkii toisen lapsen.
- 3) Kolmatta lasta ei saa hankkia kukaan.

Oletetaan, että seuraavien  $30$  vuoden aikana kukaan syntyneistä lapsista ei kuole ja kylään ei tule uusia lapsia ja kukaan ei muuta kylästä pois. Ajatellaan, että kyseessä on hierarkkinen yhteiskunta, jossa aina mies valitsee vaimon eikä toisinpäin. Kaikki yllämainituin ehdoin syntyneet pojat, jotka löytävät itselleen vaimon menevät naimisiin (eivätkä eroa). Vaimon he valitsevat vain näissä yllämainituissa  $120$  perheessä yllämainituin ehdoin syntyneiden tyttöjen joukosta. Sisaren tai veljen kanssa avioliiton solmiminen on kiellettyä, eivätkä miehet, joilla on sisar, valitse vaimokseen naista, jolla on veli. Ensimmäisenä valitsevat vaimon sellaisten perheiden pojat, jotka ovat syntyneet ensimmäisenä lapsena. Vasta tämän jälkeen voivat vaimon valita ne pojat, jotka ovat syntyneet toisena lapsena. Muuten kylän miehet valitsevat

puolisokseen kylän naisen satunnaisesti ikäeroista riippumatta, ja satunnaisessa järjestyksessä ikäerosta riippumatta.

- a) Mikä on naimattomaksi jäävien miesten teoreettinen maksimimäärä ja todennäköisyys sille, että maksimimäärä miehiä jää naimattomaksi? (Tulosta ei tarvitse laskea, riittää kun esität lausekkeen, jossa on oikeat luvut sijoitettuna ja perustelet vastauksesi.)
- b) Mikä on naimattomaksi jäävien naisten teoreettinen maksimimäärä ja todennäköisyys sille, että maksimimäärä naisia jää naimattomaksi? (Tulosta ei tarvitse laskea, riittää kun esität lausekkeen, jossa on oikeat luvut sijoitettuna ja perustelet vastauksesi.)
- c) Oletetaan, että poikia ja tyttöjä on syntynyt sekä 1. että 2. lapseksi tarkalleen odotusarvojen suuruiset määrät. Mikä on tällöin naimattomaksi jäävien miesten ja naisten lukumäärän odotusarvo? (Esitä sekä oikea lauseke että tulos.)

## Vuosi 2004, tehtävä 2

Tutkijan harrastuksena oli pelata Kimbleä vaimonsa kanssa. Kimble -pelissä on kuusi-tahkoinen noppa muovikuomun sisässä. Noppa lepää peltisellä alustalla. Kuomua painamalla saadaan nopan tulos. Peli oli kulunut ja vanha, ja tutkija alkoi epäillä nopan tulosten harhattomuutta. Tutkija alkoi epäillä, että vanha Kimble -peli pyöräyttää näkyviin nopan vastakkaisen puolen todennäköisemmin kuin jonkun muun luvun.

Arpakuutiossa silmäluvut 1-6 on merkitty niin, että vastakkaisilla puolilla noppaa olevien lukujen summa on seitsemän.

Tutkija esitti hypoteesin: Todennäköisyys, että kahden toisiaan seuraavan luvun summa on 7, on suurempi kuin teoreettinen arvo edellyttäisi.

Tutkiakseen tätä hypoteesia, tutkija paineli noppaa 360 kertaa ja tilastoi järjestyksessään jokaisen nopan luvun.

Alla oleva taulukko on tehty seuraavasti. **Ensimmäinen luku** tarkoittaa lukua, joka on kuomussa esillä ennen kuin sitä painetaan ja **seuraava luku** kuomun painamisen jälkeen saatu luku.

**VANHOJA VALINTAKOETEHTÄVIÄ**

*Vanhan Kimble -pelin nopan tulosten frekvenssit. Taulukossa on ristiintaulukoitu nopanheiton tulosten jakauma sen mukaan, mikä oli ensimmäinen luku ja mikä sitä seuraava luku.*

		Seuraava luku						Yhteensä
		1	2	3	4	5	6	
Ensimmäinen luku	1	13	10	6	6	10	17	62
	2	12	6	10	13	14	9	64
	3	5	9	8	18	10	6	56
	4	7	7	24	5	9	8	60
	5	11	17	6	10	15	5	64
	6	14	15	2	8	6	9	54
Yhteensä		62	64	56	60	64	54	360

- a) Tutki, ovatko tapaukset, joissa kahden toisiaan seuraavan luvun summa on 7, jakautuneet kuten harhattomalla nopalla. Testaa 1% merkitsevyystasolla.

Tutkijan poika huomasi, että vanhan pelin muovikuomu oli niin kulunut, ettei lukuja tahtonut nähdä kuomun läpi. Niinpä hän osti vanhemmilleen joululahjaksi uuden ”Nalle Puh Kimble” -pelin. Tutkijaa kiinnosti nähdä oliko uuden pelin ja vanhan pelin välillä eroa. Seuraavassa on ”Nalle Puh Kimble” -pelin vastaava 360 napin painalluksen sarja kuin Taulukossa 3.

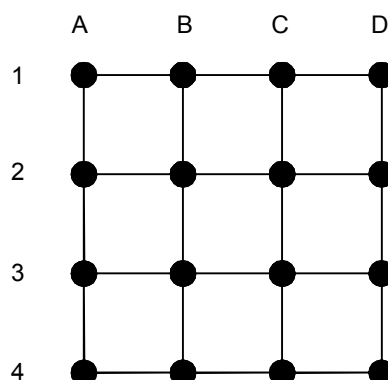
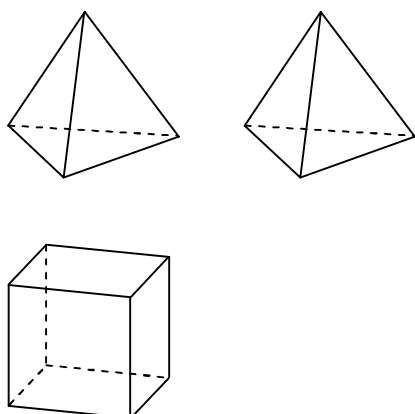
*”Nalle Puh Kimble” -pelin nopan tulosten frekvenssit. Taulukossa on ristiintaulukoitu nopanheiton tulosten jakauma sen mukaan, mikä oli ensimmäinen luku ja mikä sitä seuraava luku.*

		Seuraava luku						Yhteensä
		1	2	3	4	5	6	
Ensimmäinen luku	1	10	14	16	11	7	9	67
	2	13	6	5	7	13	10	54
	3	10	13	5	13	11	9	61
	4	13	8	13	6	9	9	58
	5	6	5	19	11	13	13	67
	6	14	8	4	10	14	3	53
Yhteensä		66	54	62	58	67	53	360

- b) Onko tapauksen ”kahden perättäisen nopanpainalluksen summa on 7” suhteellinen osuus molemmissa peleissä sama? Testaa 1% merkitsevyystasolla.

### Vuosi 2004, tehtävä 3

Aartenetsintä-pelissä käytetään kahta tetraedrin muotoista (4 tahkoista) noppaa ja yhtä kuutionoppaa (6 tahkoa). Toisen tetraedrinopan tahkoihin on merkitty kirjaimet A, B, C ja D - yksi kuhunkin tahkoon. Heitettäessä tätä noppaa kunkin kirjaimen saamisen todennäköisyys on  $1/4$ . Toisen tetraedrinopan tahkoihin on merkitty numerot 1, 2, 3 ja 4 - yksi kuhunkin tahkoon. Heitettäessä tätä noppaa kunkin numeron 1-4 saamisen todennäköisyys on  $1/4$ . Kuutionopan tahkoihin on merkitty numerot 0, 1, 2, 3, 4 ja 5 - yksi kuhunkin tahkoon. Heitettäessä tätä noppaa kunkin numeron 0-5 saamisen todennäköisyys on  $1/6$ .



Pelilautana käytetään ruudukkoa, joka koostuu pysty- ja vaakasuoraan 4 pisteestä ja näitä pisteitä yhdistävistä poluista.

Yksi pelikierros koostuu seuraavista vaiheista:

Heitetään molempia tetraedrinoppia. Heiton tulos määrittää 'aarteen' paikan (pisteen koordinaatit) pelilaudalla.

Heitetään molempia tetraedrinoppia uudelleen. Heiton tulos määrittää 'etsijän' paikan pelilaudalla.

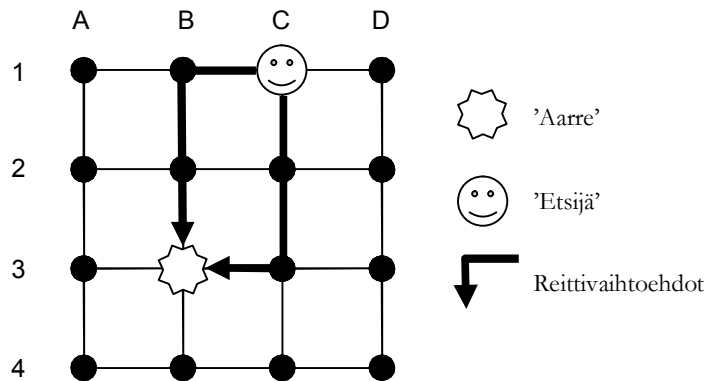
Heitetään kuutionoppaa. Heiton tulos määrittää sen, kuinka monta polkua (pisteiden väliä) etsijä voi enintään kulkea kohti 'aarretta'.

'Etsijä' kulkee 'aarretta' kohti aina lyhyintä mahdollista reittiä. Reitillään 'etsijä' saa käyttää jokaista polkua vain yhden kerran ja kääntyä vain yhden kerran. Pelissä on myös mahdollista, että 'aarteen' ja 'etsijän' paikka on tetraedrinoppien heittojen jälkeen täsmälleen sama ja 'etsijä' löytää 'aarteen' kulkematta lainkaan.

**Esimerkki:**

Ensimmäisellä tetraedrinoppien heitolla saadaan B ja 3. Toisella tetraedrinoppien heitolla saadaan C ja 1. Viereiseen kuvaan on merkitty 'aarteen' ja 'etsijän' paikat ja reitit, joita etsijä voi kulkea. Etsijän reitin pituus on esimerkeissä 3 polkua.

Jos kuutionopan heitolla saadaan 3 tai enemmän, niin 'etsijä' pääsee 'aarteen' luo.



Tarkastellaan yhtä ainoata pelikierrosta.

- Laske todennäköisyys sille, että 'etsijä' pääsee 'aarteen' luo, vaikka kuutionopalla heitettäisiin luku 0? Merkitse näkyviin laskun vaiheet.
- Laske todennäköisyys sille, että 'etsijä' ei pääse 'aarteen' luo, vaikka kuutionopalla heitettäisiin luku 5? Merkitse näkyviin laskun vaiheet.
- Selvitä 'etsijän' kaikkien eripituisten reittien, jotka pelissä ovat mahdollisia, todennäköisyysjakauma ja esitä se taulukkona. Merkitse näkyviin, miten olet laskenut todennäköisyydet.
- Laske todennäköisyys sille, että 'etsijä' pääsee 'aarteen' luo. Merkitse näkyviin laskun vaiheet.

**Vuosi 2005, tehtävä 3**

Tämä tehtävä perustuu todelliseen havaintoaineistoon. Asioiden hahmottamisen helpottamiseksi eri aineistot on nimetty esittelyn yhteydessä. Taulukossa 2 on kaikkien aineistojen tulokset yhdistettynä. Taulukon 2 muuttujat on määritelty aineistojen esittelyn yhteydessä.

**AINEISTO A**

Tutkijalla ei ollut majoituspaikassaan Tartossa muuta lukemista kuin Etelä-Viron puhelinluettelo. Niinpä hän päätti puhelinluettelon avulla selvittää, mitkä ovat Etelä-Viron yleisimmät sukunimet. Yhdellä puhelinluettelon palstalla on noin 100 nimeä. Tutkija kirjasi puhelinluettelosta kaikki sellaiset sukunimet, joita oli vähintään yksi palsta eli 100 nimeä. Nämä sukunimet ja sukunimien lukumäärät löytyvät taulukosta 2. Tällaisia nimiä löytyi 31 eri nimeä ja yhteensä näiden nimien haltijoita oli 4360.

Lisäksi taulukkoon 2 on lisätty katkoviivan alapuolelle ne sukunimet, jotka tulivat esille aineistoissa B ja C. Myös näistä sukunimistä tutkija laski saman puhelinluettelon avulla sukunimien lukumäärät.

Taulukon 2 muuttujat:

LKM: löydettyjen nimien lukumäärä Etelä-Viron puhelinluettelossa.

JNR: on nimen yleisyyttä kuvaava järjestysluku kaikissa aineistoissa A, B ja C Etelä-Viron puhelinluettelon mukaan. Lukuun 31 saakka (vähintään 100 nimeä) se kuvaa nimen yleisyyttä myös koko puhelinluettelossa. Sen sijaan siitä alaspäin mukaan on otettu vain sellaiset nimet, jotka ovat tulleet mukaan muiden aineistojen keruun yhteydessä.

### **AINEISTO B**

Tämän jälkeen tutkijaa kiinnosti se, kuinka ihmiset osaavat arvata mitkä ovat yleisimmät sukunimet. Hän teki Tarton yliopiston psykologian laitoksen opiskelijoille kaksivaiheisen kyselyn. Kyselyyn vastasi 14 henkilöä.

AINEISTO B 1: Avoin kysely

Lomakkeen ensimmäisellä puolella oli 10 tyhjää riviä, ja opiskelijoiden piti kirjoittaa riveille heidän mielestään Etelä-Viron puhelinluettelon 10 yleisintä nimeä järjestyksessä yleisimmästä nimestä kymmenenneksi yleisempään nimeen. Jatkossa tästä kyselytavasta käytetään nimeä avoin kysely, Kaikki opiskelijoiden ehdottamat nimet on lisätty taulukkoon lisättyinä tiedolla siitä, kuinka monta kyseistä nimeä löytyi Etelä-Viron puhelinluettelosta.

Taulukon 2 muuttujat:

B1PIST: Pistemääräindeksi

Muodostettiin B1PIST -indeksi niin, että jokaisessa 14 vastauspaperissa ensimmäiseksi ehdotettu nimi sai 10 pistettä, toiseksi 9 pistettä ja niin edelleen niin, että kymmenenneksi ehdotettu nimi sai yhden pisteen. Jokaisen ehdotetun nimen pisteet summattiin yhteen.

Esimerkkinä yleisimmäksi arvioitu nimi Tamm pistemäärä muodostuu seuraavasti  $10+10+10+10+10+10+10+10+9+9+7+1=106$  Eli 12 henkilöä 14:sta on arvioinut nimen Tamm kymmenen yleisimmän nimen joukkoon ja näistä 8 on arvioinut sen yleisimmäksi nimeksi. Näin saatiin pistemääräindeksi jota kuvataan taulukossa 2 nimellä B1PIST.

B1JNR: Pistemääräindeksin järjestyslukumuuttuja

B1LKM: Nimeä ehdottaneiden lukumäärä

B1LJNR:B1LKM:n järjestyslukumuuttuja

AINEISTO B 2: Valmis nimilista

Tämän jälkeen opiskelijat saivat kääntää sivua. Toisella puolella paperia oli lueteltu 30 nimeä aakkosjärjestyksessä. Lisäksi sivulla oli lukuja 420:sta lukuun 10 saakka suuruusjärjestyksessä. Opiskelijoiden tehtävänä oli yhdistää oikeat nimet ja oikeat luvut. Luvut kuvasivat sitä kuinka monta kutakin sukunimeä oli luettelosta löytynyt. Näiden nimien joukossa oli 10 yleisintä nimeä ja lisäksi hämäyksen vuoksi harvinaisempia sukunimiä.

## VANHOJA VALINTAKOETEHTÄVIÄ

Alla olevassa taulukossa 1 ovat kyseiset luettelossa olleet nimet ja niihin liittyneet luvut. (Opiskelijoille nimet oli kirjoitettu allekkain ja viereen oli jätetty tyhjä tila numerolle. Numerot olivat suurimmasta pienempään sivun oikeassa reunassa. Tässä tilan säästämisen takia nimet ja luvut on kirjoitettu peräkkäin).

Taulukko 1.

Kyselyssä (muodostaa AINEISTO 2 B:n) mukana olleet 30 nimeä ja niihin liittyneet numerot. Tehtävänä oli yhdistää nimet ja numerot niin, että ne vastaavat Etelä-Viron puhelinluettelossa olevia sukunimien lukumääriä. Joukossa on puhelinluettelon 10 yleisintä nimeä ja lisäksi tehtävän vaikeuttamiseksi muita nimiä.

Allik, Aru, Ilves, Ivanov/Ivanova, Juhkam, Kaasik, Kask, Kroon, Kukk, Kuusik, Kuznetsov/Kutznetsova, Käis, Köiv, Leis, Liivat, Lill, Luik, Puusepp, Pärn, Raudsepp, Realo, Rebane, Roos, Röömus, Saag, Saar, Sikk, Tamm, Uibo, Valk.

Nimiin yhdistettävät nimien yleisyyttä kuvaavat nimien lukumäärät:

420, 280, 220, 200, 200, 200, 190, 170, 170, 160, 140, 140, 120, 120, 100, 100, 80, 80, 60, 60, 40, 40, 30, 30, 30, 20, 20, 10, 10, 10

Taulukon 2 muuttujat:

B2KA: Jokaisen Taulukkoon 1 kuuluneeseen nimeen liitettyjen lukumäärien keskiarvo 14 vastaajalta.

B2JNR: B2KA muuttujan järjestyslukumuuttuja.

B2RYHMÄ: Ryhmä 1: Etelä-Viron puhelinluettelon yleisimmät nimet 1 – 10  
Ryhmä 2: Etelä-Viron puhelinluettelon mukaan melko yleiset nimet  
Ryhmä 3: Etelä-Viron puhelinluettelon mukaan harvinaiset nimet

## AINEISTO C

Tutkija sai käsiinsä internetistä tämän vuoden maaliskuussa tehdyn lehtikirjoituksen, jossa oli esitetty koko Viron 25 yleisintä sukunimeä ja nimien määrät Virossa. Nämä nimet on lisätty taulukkoon lisättyinä tiedolla siitä kuinka monta kyseistä nimeä löytyi Etelä-Viron puhelinluettelosta.

Taulukon 2 muuttujat:

CLKM: Tilaston mukaan Virossa kyseisen nimen haltijoiden lukumäärä

CJNR: CLKM:n järjestyslukumuuttuja.

Lisämuuttuja KANSAL: Kansallisuus (1 = virolainen, 2 = venäläinen)

VANHOJA VALINTAKOETEHTÄVIÄ

Taulukko 2. (jatkuu myös seuraavalle sivulle)  
Tulokset Etelä-Viron yleisimmistä sukunimistä.

Sukunimi	A		B1				B2		RYHMÄT	C		Kansal- lisuus
	LKM	JNR	B1PIST	B1JNR	B1LKM	B1LJNR	B2KA	B2JNR	B2RYHMÄ	CLKM	CJNR	
Tamm	420	1	106	1	12	1	329	1	1	5259	2	1
Saar	280	2	59	2	8	2	237	3	1	4320	3	1
Raudsepp	220	3	3	50	1	28	132	8	1	1790	18	1
Kukk	200	4	5	37	1	28	139	7	1	2730	9	1
Köiv	200	4	15	11	2	13	99	14	1	-	-	1
Ilves	200	4	6	33	1	28	87	17	1	2155	12	1
Rebane	190	7	9	19	2	13	131	9	1	2272	11	1
Kask	170	8	58	3	8	2	256	2	1	2834	7	1
Puusepp	170	8	0	74	0	74	128	10	1	-	-	1
Luik	160	10	0	74	0	74	71	20	1	1809	15	1
Pöder	150	11	11	14	2	13	-	-	0	1639	22	1
Lohmus	150	11	0	74	0	74	-	-	0	-	-	1
Uibo	140	13	10	15	2	13	112	13	2	-	-	1
Mägi	140	13	45	4	7	4	-	-	0	3630	4	1
Sepp	130	15	38	6	5	6	-	-	0	3624	5	1
Oja	130	15	9	19	3	7	-	-	0	1785	19	1
Kütt	130	15	3	50	1	28	-	-	0	1638	23	1
Kull	130	15	0	74	0	74	-	-	-	-	-	1
Mölder	130	15	0	74	0	74	-	-	-	-	-	1
Kuusik	120	20	7	27	2	13	187	4	2	-	-	1
Kaasik	120	20	7	27	2	13	182	5	2	1798	17	1
Lepik	120	20	15	11	2	13	-	-	0	1805	16	1
Leppik	110	23	0	74	0	74	-	-	0	-	-	1
Pärn	110	23	9	19	3	7	170	6	2	1930	13	1
Kivi	110	23	5	37	2	13	-	-	0	-	-	1
Ivanov	100	26	0	74	0	74	94	16	2	6624	1	2
Kuznetsov	100	26	0	74	0	74	57	27	2	1625	24	2
Kuusk	100	26	17	9	3	7	-	-	0	1769	20	1
Vaher	100	26	5	37	1	28	-	-	0	-	-	1
Jürgenson	100	26	0	74	0	74	-	-	0	-	-	1
Kallas	100	26	0	74	0	74	-	-	0	-	-	1
Raud	90	32	6	33	2	13	-	-	0	-	-	1
Rätsep	90	32	5	37	1	28	-	-	0	-	-	1
Koppel	90	32	0	74	0	74	-	-	0	1891	14	1
Roos	80	35	0	74	0	74	80	18	2	-	-	1
Sikk	80	35	0	74	0	74	52	28	2	-	-	1
Karu	80	35	20	8	3	7	-	-	0	1711	21	1
Järv	80	35	11	14	2	13	-	-	0	-	-	1
Lill	70	39	0	74	0	74	98	15	2	2579	10	1
Mets	70	39	26	7	3	7	-	-	0	-	-	1
Laas	70	39	1	68	1	28	-	-	0	-	-	1
Juhkam	60	42	0	74	0	74	59	26	2	-	-	1
Unt	60	42	10	15	3	7	-	-	0	-	-	1
Lepp	60	42	16	10	2	13	-	-	0	1569	25	1
Must	60	42	8	23	2	13	-	-	0	-	-	1
Pöld	50	46	13	13	2	13	-	-	0	-	-	1
Järve	50	46	5	37	1	28	-	-	0	-	-	1
Suits	40	48	5	37	1	28	122	11	3	-	-	1
Allik	40	48	41	5	6	5	114	12	3	-	-	1
Laar	40	48	5	37	1	28	-	-	0	-	-	1
Mark	40	48	5	37	1	28	-	-	0	-	-	1
Padar	40	48	4	48	1	28	-	-	0	-	-	1
Valk	30	53	7	27	2	13	75	19	3	-	-	1
Tammeorg	30	53	3	50	1	28	65	24	3	-	-	1
Leis	30	53	0	74	0	74	59	25	3	-	-	1
Hunt	30	53	5	37	1	28	-	-	0	-	-	1
Lillo	30	53	2	60	1	28	-	-	0	-	-	1
Soo	30	53	2	60	1	28	-	-	0	-	-	1
Mänd	30	53	1	68	1	28	-	-	0	-	-	1
Smirnov	30	53	0	74	0	74	-	-	0	3260	6	2
Petrov	30	53	0	74	0	74	-	-	0	2811	8	2
Kroon	20	62	0	74	0	74	68	22	3	-	-	1

**VANHOJA VALINTAKOETEHTÄVIÄ**

Viin	20	62	2	60	2	13	31	30	3	-	-	1
Nõmme	20	62	8	23	1	28	-	-	0	-	-	1
Pai	20	62	6	33	1	28	-	-	0	-	-	1
Musting	20	62	3	50	1	28	-	-	0	-	-	1
Mäe	20	62	3	50	1	28	-	-	0	-	-	1
Palo	20	62	1	68	1	28	-	-	0	-	-	1
Tammik	20	62	1	68	1	28	-	-	0	-	-	1
Sauk	10	70	7	27	1	28	71	21	3	-	-	1
Röömus	10	70	0	74	0	74	67	23	3	-	-	1
Realo	10	70	2	60	1	28	35	29	3	-	-	1
Oru	10	70	10	15	1	28	-	-	0	-	-	1
Rannu	10	70	9	19	1	28	-	-	0	-	-	1
Varul	10	70	7	27	1	28	-	-	0	-	-	1
Rannik	10	70	6	33	1	28	-	-	0	-	-	1
Keskküla	10	70	3	50	1	28	-	-	0	-	-	1
Kreek	10	70	3	50	1	28	-	-	0	-	-	1
Heinla	10	70	2	60	1	28	-	-	0	-	-	1
Laasik	10	70	2	60	1	28	-	-	0	-	-	1
Nurm	10	70	2	60	1	28	-	-	0	-	-	1
Puskar	10	70	2	60	1	28	-	-	0	-	-	1
Orgusaar	10	70	1	68	1	28	-	-	0	-	-	1
Meri	4	84	3	50	1	28	-	-	0	-	-	1
Konstabel	3	85	5	37	1	28	-	-	0	-	-	1
Öpik	3	86	3	50	1	28	-	-	0	-	-	1
Puust	2	87	8	23	1	28	-	-	0	-	-	1
Pihla	2	87	1	68	1	28	-	-	0	-	-	1
Tõnissaar	1	89	7	27	1	28	-	-	0	-	-	1
Haab	0	90	8	23	1	28	-	-	0	-	-	1
Täht	0	90	5	37	1	28	-	-	0	-	-	1
Meel	0	90	4	48	1	28	-	-	0	-	-	1
Likas	0	90	3	50	1	28	-	-	0	-	-	1
YHTEENSÄ	6285		770		140		3407			64857		

Taulukon 2 loppu

**Taulukko 3.**

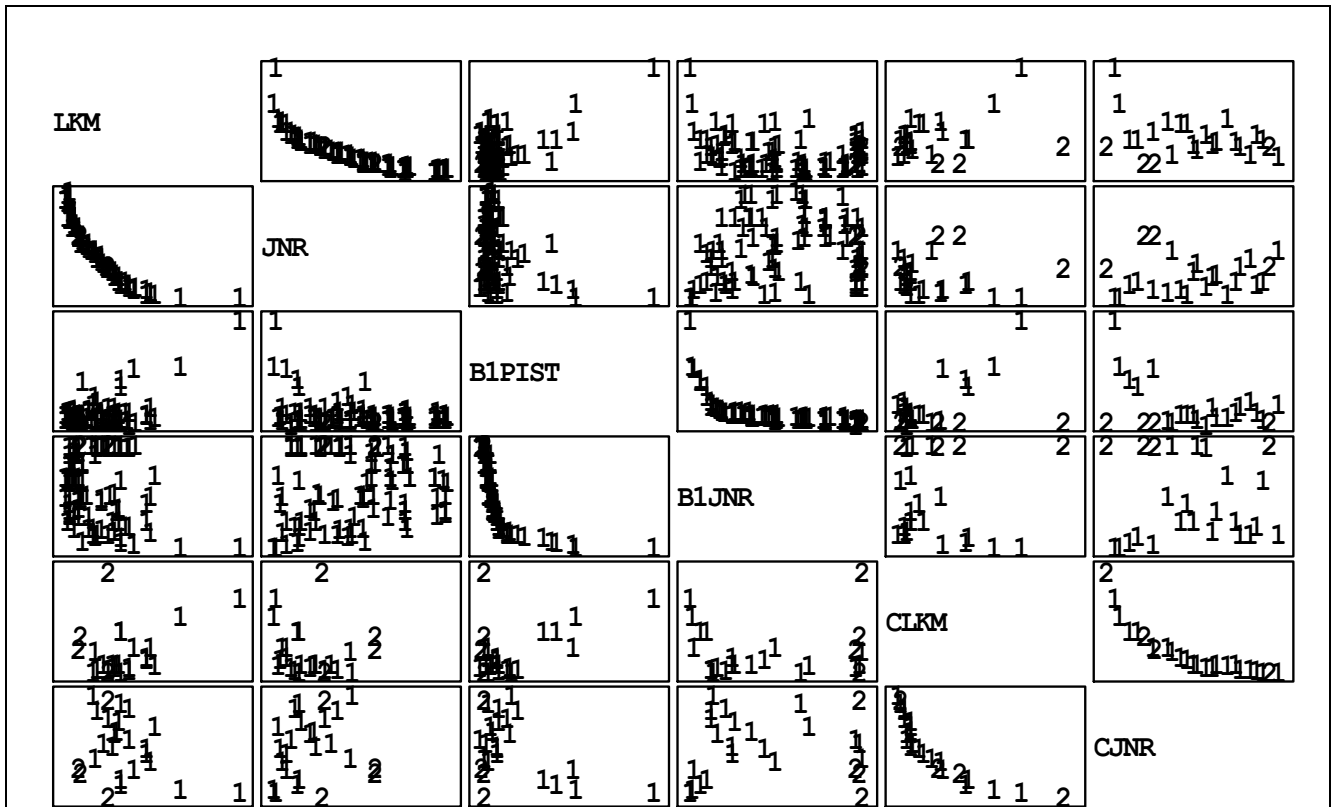
Muuttujien välistet korrelaatiot. Havaintoina ovat olleet ne 30 nimeä, jotka olivat valmiilla listalla.

	LKM	JNR	B1PIST	B1JNR	B1LKM	B1LJNR	B2KA	B2JNR
LKM	1.00	-0.88	0.67	-0.49	0.62	-0.37	0.77	-0.69
JNR	-0.88	1.00	-0.38	0.41	-0.36	0.28	-0.63	0.70
B1PIST	0.67	-0.38	1.00	-0.68	0.98	-0.55	0.82	-0.58
B1JNR	-0.49	0.41	-0.68	1.00	-0.77	0.92	-0.69	0.70
B1LKM	0.62	-0.36	0.98	-0.77	1.00	-0.69	0.83	-0.62
B1LJNR	-0.37	0.28	-0.55	0.92	-0.69	1.00	-0.57	0.57
B2KA	0.77	-0.63	0.82	-0.69	0.83	-0.57	1.00	-0.90
B2JNR	-0.69	0.70	-0.58	0.70	-0.62	0.57	-0.90	1.00

**Taulukko 4.**

Muuttujan B2KA keskiarvot, hajonnat ja frekvenssit muuttujan B2RYHMÄ mukaan luokiteltuna

	1	2	3	Total
Keskiarvo	160.90	109.10	70.70	113.57
Keskihajonta	84.17	52.48	29.00	68.75
Frekvenssi	10	10	10	30



Kuva 1.

Muuttujien väliset hajontakuviot. Kuvissa 25 Viron yleisintä nimeä. Kansalaisuus on merkitty siten että 1 = virolainen, 2 = venäläinen.

Huom! Havaintopisteitä on kuvassa täristetty. Toisin sanoen täsmälleen päällekkäisiä pisteitä on siirretty satunnaisesti johonkin suuntaan, että eri havainnot erottuvat paremmin toisistaan.

Taulukko 5.

Muuttujien välisiä korrelaatioita. Havaintoina ovat olleet ne 25 nimeä, jotka olivat mukana Viron 25 yleisimmän nimen joukossa.

	LKM	JNR	B1PIST	B1JNR	CLKM	CJNR
LKM	1.00	-0.82	0.72	-0.43	0.36	-0.34
JNR	-0.82	1.00	-0.43	0.46	-0.11	0.14
B1PIST	0.72	-0.43	1.00	-0.66	0.49	-0.47
B1JNR	-0.43	0.46	-0.66	1.00	-0.00	0.01
CLKM	0.36	-0.11	0.49	-0.00	1.00	-0.85
CJNR	-0.34	0.14	-0.47	0.01	-0.85	1.00

Taulukko 6.

Muuttujien välisiä korrelaatioita. Havaintoina ovat olleet ne 21 virolaista nimeä, jotka olivat mukana Viron 25 yleisimmän nimen joukossa.

	LKM	JNR	B1PIST	B1JNR	CLKM	CJNR
LKM	1.00	-0.79	0.69	-0.24	0.73	-0.58
JNR	-0.79	1.00	-0.37	0.18	-0.46	0.48
B1PIST	0.69	-0.37	1.00	-0.63	0.87	-0.65
B1JNR	-0.24	0.18	-0.63	1.00	-0.39	0.20
CLKM	0.73	-0.46	0.87	-0.39	1.00	-0.88
CJNR	-0.58	0.48	-0.65	0.20	-0.88	1.00

Tutki seuraavia tutkimusongelmia.

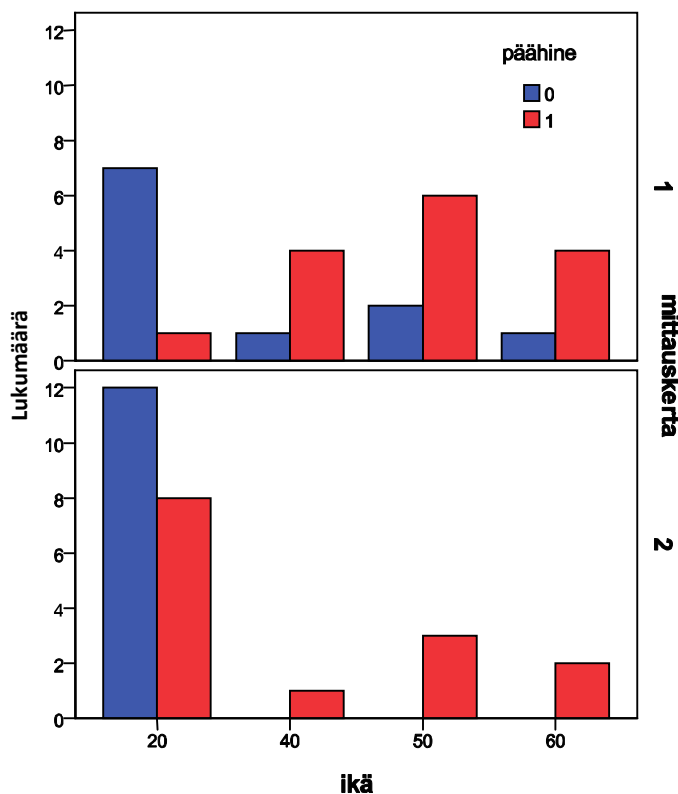
- Auttaako valmiin nimilistan esittäminen nimien yleisyyden arviointia verrattuna avoimen listan käyttöön? Mikä on tulos Taulukon 3 lukujen perusteella? Esitä kaikki ne taulukon 3 tulokset, jotka ovat tarpeelliset hypoteesin tutkimiseksi.** Käytä taulukon muuttujalyhenteitä. (Huom! Tarpeettomien tulosten ilmoittamisesta voi saada miinus pisteitä.)
- Laske suhteellisen osuuden testin avulla voiko aineiston perusteella sanoa, että Etelä-Virossa on vähemmän venäläisiä kuin koko Virossa.** Tässä oletetaan, että venäläisten nimien osuus yleisimmistä nimistä koko Viron tilastossa (Aineisto C) on suhteessa sama kuin kaikista nimistä. Oletetaan myös, että sama pätee Etelä-Viron puhelinluettelosta havaituista yleisimmistä nimistä (katso aineiston määrittelystä kohdasta A).
- Vertaa valmiin nimilistan Etelä-Viron puhelinluettelon 10 yleisimmän nimen (eli nimien JNR 1- 10, eli B2RYHMÄ=1) ja valmiin nimilistan 10 harvinaisimman nimen (B2RYHMÄ=3) välistä eroa. Ovatko tutkittavat osanneet yhdistää nimet ja luvut niin, että ero on oletetun suuntainen. Käytä tulosten testaamiseen mielestäsi sopivinta testiä.** Keskiarvot, hajonnat ja frekvenssit näet Taulukosta 4.
- Koko Viron nimitilasto oli julkisuudessa maaliskuussa. Tutkijalla heräsi epäily, että tämä olisi vaikuttanut opiskelijoiden avoimen kyselyn tuloksiin. Voisitko hyväksyä kaiken käytettävissä olevan aineiston perusteella oletuksen, että opiskelijat muodostivat avoimen kyselyn nimien yleisyyden arvionsa tietoon koko Viron nimien yleisyydestä huomioon ottaen venäläisten ja virolaisten osuuden väestöstä Etelä-Virossa. Esitä mikä materiaali on oleellista hypoteesin hyväksymiseen tai kumoamiseen.** Viittaa taulukkojen numeroihin ja muuttujalyhenteisiin tai muihin tässä tutkimuksessa esiin tulleisiin seikkoihin, jotka tukevat tai kumoavat esitetyn hypoteesin. (Huom! Tarpeettomien tulosten ilmoittamisesta voi saada miinus pisteitä.)

# Vanhojen valintakoetehtävien ratkaisut

*Jep, tässä osassa ovat vuoden 2012 valintakoekirjan mukaiset ratkaisut edellisen osan vanhoihin valintakoetehtäviin.*

## Vuosi 1989, tehtävä 4

- a) Mittauskerralla 1 päähinettä käytti viisitoista Sankarin havaitsemista jalankulkijoista ja mittauskerralla 2 neljätoista jalankulkijaa. Ero on siis hyvin pieni. Kun tarkastellaan sukupuolen ja iän jakaumia mittauskerroilla, havaitaan että mittauskerralla 2 on jalankulkijoista suuri osa ollut noin 20-vuotiaita. Mittauskerran 1 tuloksista havaitaan, että kaikissa muissa ikäryhmissä päähineen käyttäjiä on enemmän kuin niitä, jotka eivät käytä päähinettä. Näyttäisi siis siltä, että monet 20-vuotiaat eivät käytä päähinettä, vaikka olisi kylmä ja koska tämä ikäryhmä on yliedustettuna mittauskerralla 2, tulokset ovat melko varmasti harhaiset.



- b) Tulisi huomioida, että eri ikäryhmistä tulee mukaan tutkimusyksiköjä (jalankulkijoita) samassa suhteessa kummallakin mittauskerralla. Saattaa olla, että myös sukupuoli olisi syytä huomioida samaan tapaan, vaikka kovin vahvaa näyttöä sen vaikutuksesta ei olekaan. Tarkastellaan sukupuolta ja ikää samanaikaisesti taulukoimalla mittauskerta, sukupuoli, päähineen käyttö ja ikä. Mittauskerralla 1 saattaa olla niin, että 20-vuotiaat naiset käyttävät päähinettä vähiten ja 60-vuotiaat naiset eniten. Havaintojen määrä on pieni asian osoittamiseen kovin varmasti. Koe kannattaisi suorittaa kiintiötantaa käyttäen iän ja sukupuolen perusteella.

	mittauskerta							
	1				2			
	sukupuoli				sukupuoli			
	M		N		M		N	
ikä	päähine		päähine		päähine		päähine	
	0	1	0	1	0	1	0	1
20	3	1	4	-	6	4	6	4
40	1	2	-	2	-	-	-	1
50	1	3	1	3	-	1	-	2
60	1	1	-	3	-	1	-	1

**Vuosi 1990, tehtävä 2**

- a)  $P(X < 5) = P\left(Z < \frac{5-4}{1}\right) = P(Z < 1) \approx 0.8413$
- b)  $P(X < 5) = P\left(Z < \frac{5-6}{1,5}\right) = P(Z < -0,67) \approx 0.2514$
- c)  $P(X < 5) = \frac{5-3,5}{6-3,5} = 0.6$
- d)  $P(X < 5) = \frac{9}{12} \cdot 0.8413 + \frac{1}{12} \cdot 0.2514 + \frac{2}{12} \cdot 0.6 \approx 0.7519$

**Vuosi 1990, tehtävä 3**

- a)  $H_0$ : Kypärän käyttö ei ole lisääntynyt,  $p = p_0$   
 $H_1$ : Kypärän käyttö on lisääntynyt,  $p > p_0$  – huom. yksisuuntainen testi

b) 
$$Z = \frac{P - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.2 - 0.1}{\sqrt{\frac{0,1 \cdot 0,9}{100}}} = \frac{0.1}{\sqrt{0.0009}} = \frac{0.1}{0.03} \approx 3,33$$

Koska  $3,33 > 1,645$  (kriittinen arvo yksisuuntaisessa testauksessa 5% merkitsevyystasolla), voidaan  $H_0$  hylätä. Päätellään, että kypärän käyttö on lisääntynyt.

- c) I lajin virhe: tosi nollahypoteesi hylätään eli tässä pääteltäisiin, että kypärän käyttö on lisääntynyt vaikka se oikeasti ei olisi.
- d) II lajin virhe: epätosi nollahypoteesi jää voimaan eli tässä pääteltäisiin, että kypärän käyttö ei ole lisääntynyt, vaikka se oikeasti olisi.

**Vuosi 1991, tehtävä 1**

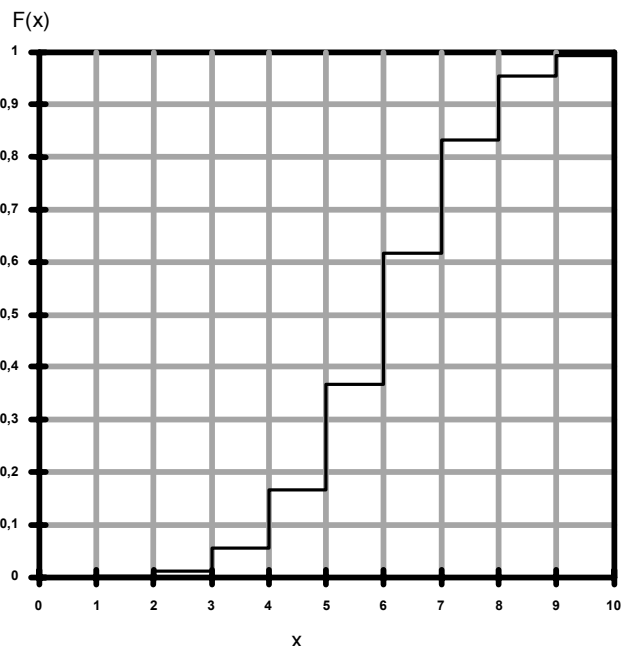
- a) Kuviosta nähdään, että muuttuja on diskreetti ja koska kaikki pistemäärät ovat näkyvissä  $F(10)=1$ , on toistojen lkm (10) pieni. Kyseessä on binomijakauma, jossa  $n = 10$ . Ratkaistaan  $p$  odotusarvon avulla

$$EX = \sum_i p_i x_i = np$$

$p_i$	$x_i$	$p_i x_i$	$F(x)$
0,0001	0	0	0,00010
0,00157	1	0,00157	0,00167
0,01061	2	0,02122	0,01228
0,04246	3	0,12738	0,05474
0,11147	4	0,44588	0,16621
0,20065	5	1,00325	0,36686
0,25082	6	1,50492	0,61768
0,21499	7	1,50493	0,83267
0,12093	8	0,96744	0,95360
0,04031	9	0,36279	0,99391
0,00604	10	0,0604	0,99995
$\Sigma = 0,9995$		$\Sigma = 5,99978$	

$$EX \approx 6 \rightarrow p = \frac{6}{10} = 0,6$$

HUOM! Se, että pistetodennäköisyyksien summa ei ole tasan 1 johtuu pyöristämisestä viiteen desimaaliin.



b) Kuvioista nähdään, että jakauma on jatkuva tasajakauma, jonka

$$f(x) = \begin{cases} \frac{1}{10-2}, & \text{kun } 2 \leq x \leq 10 \\ 0, & \text{kun } x < 2 \text{ tai } x > 10 \end{cases}$$

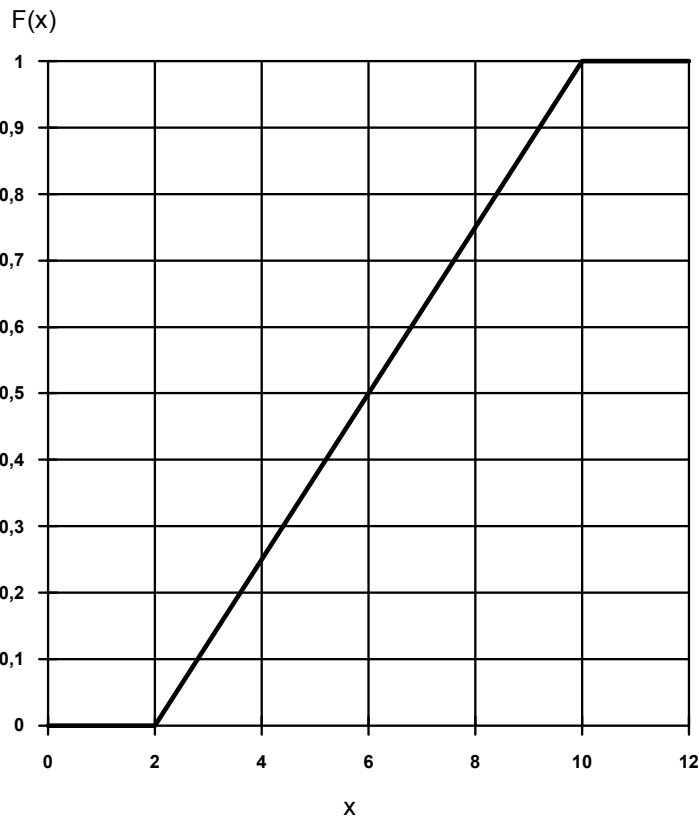
Seuraavaa integraalilaskennan tapaa ei edellytetä:

$$E(x) = \int_2^{10} xf(x)dx = \int_2^{10} \frac{1}{8} x dx = \frac{1}{8} \left[ \frac{1}{2} x^2 \right]_2^{10} = \frac{1}{8} \left( \frac{1}{2} \cdot 10^2 - \frac{1}{2} \cdot 2^2 \right) = 6$$

$$\begin{aligned} D(x) &= \sqrt{E(x^2) - [E(x)]^2} = \sqrt{\int_2^{10} x^2 f(x) dx - 6^2} = \sqrt{\int_2^{10} \frac{1}{8} x^2 dx - 6^2} = \sqrt{\frac{1}{8} \left[ \frac{1}{3} x^3 \right]_2^{10} - 6^2} \\ &= \sqrt{\frac{1}{8} \left( \frac{1}{3} \cdot 10^3 - \frac{1}{3} \cdot 2^3 \right) - 6^2} \approx \sqrt{41,33 - 36} \approx 2,31 \end{aligned}$$

Tasajakauma odotusarvon ja keskihajonnan voi laskea helpomminkin tavalla jota valintakoe kirjassa ei käydä läpi

$$E(x) = \frac{2+10}{2} = 6 \quad D(x) = \sqrt{\frac{(10-2)^2}{12}} \approx 2,31$$



- c) Jakauma on normaalijakauma ja symmetrinen arvon 100 molemmin puolin (tämä nähdään sekä kuvaajasta että taulukosta); siis  $\mu=100$ . Keskihajonta voidaan laskea normitetun normaalijakauman avulla kaavasta

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

Taulukosta ei nähdä, kuinka suuri osuus pinta-alasta on pistemäärillä alle 10 tai yli 190, mutta nämä voidaan laskea. Koska välillä 10–190 oleva osuus on 0,9999 ja jakauman koko pinta-ala on 1, on välin ulkopuolinen osuus  $1-0,9999=0,0001$  ja tämä jaettuna kahdella on 0,00005. Nyt voidaan laskea taulukkoon kertymäfunktion arvot.

x	F(x)	x	F(x)	x	F(x)
-10	0,00005	80	0,15866	150	0,99377
20	0,00007	90	0,30854	160	0,99862
30	0,00027	100	0,50000	170	0,99973
40	0,00138	110	0,69146	180	0,99993
50	0,00623	120	0,84134	190	0,99995
60	0,02277	130	0,93318	190-	1
70	0,06682	140	0,97723		

$$0,97723 \approx \Phi(2)$$

$$0,84134 \approx \Phi(1)$$

$$0,06682 \approx \Phi(-1,5)$$

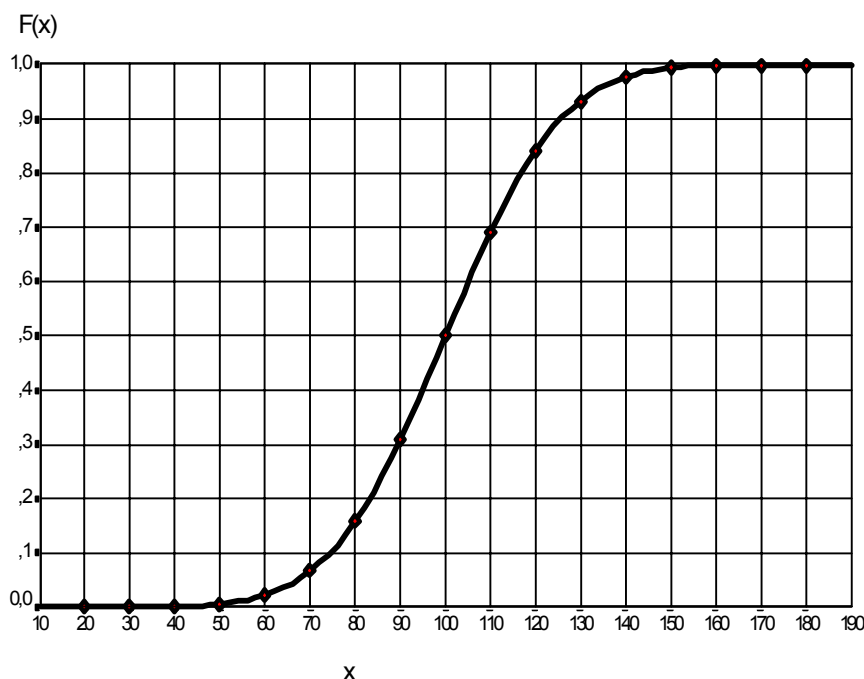
$$\sigma = \frac{140 - 100}{2} = 20 \quad \text{tai}$$

$$\sigma = \frac{120 - 100}{1} = 20 \quad \text{tai}$$

$$\sigma = \frac{70 - 100}{-1,5} = 20$$

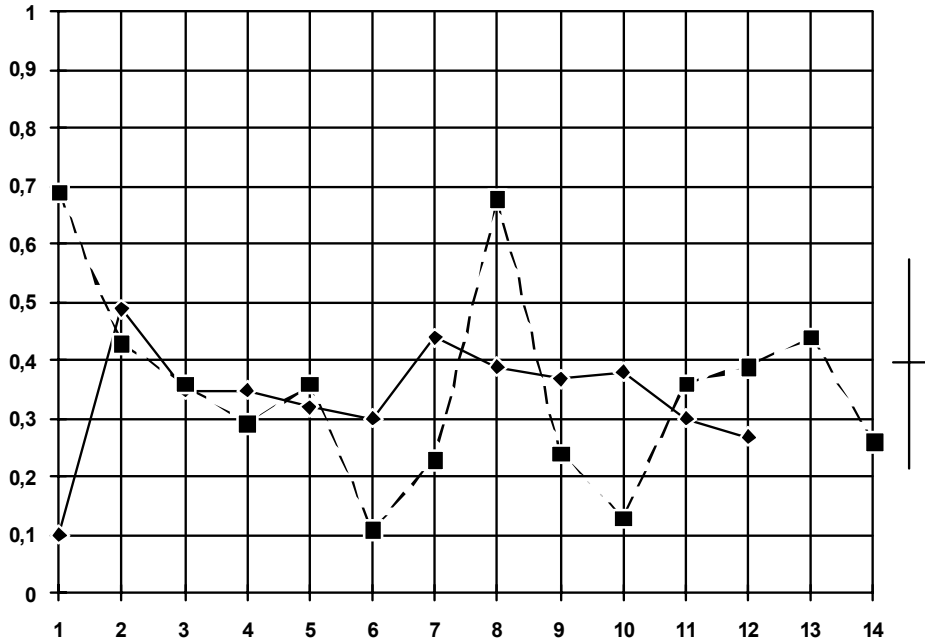
jakauma on siis  $N(100,20)$

Kertymäfunktion piirtämiseksi käytetään edellä taulukkoon laskettuja arvoja ja piirretään niiden mukaan S-käyrä.

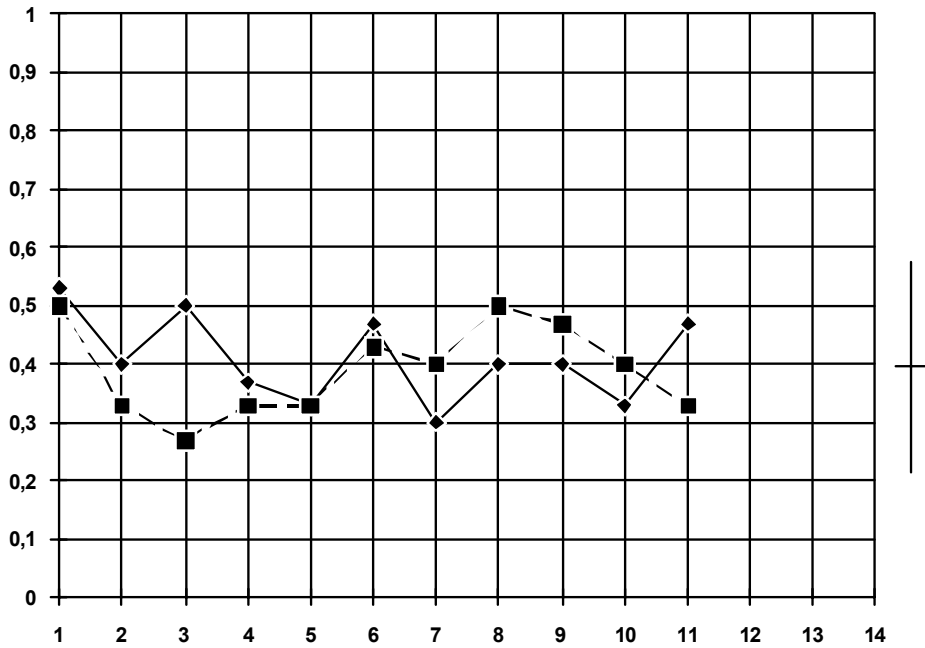


**Vuosi 1991, tehtävä 2**

a) Naisten (niiden parien, joissa nainen on mainittu ensin) suhteellinen osuus näkyy taulukoiden sarakkeissa ”nainen ensin / kaikki”.



Otokset 1 ja 2



Otokset 3a ja 3b

b) Käytetään  $\alpha=0,05$ , jolloin saadaan siis 95 % luottamusväli.  $q = (1-p)$

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\hat{p}\hat{q}/n} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sqrt{\hat{p}\hat{q}/n}$$

$$0,4 - 1,96 \cdot \sqrt{0,4 \cdot 0,6/30} \leq p \leq 0,4 + 1,96 \cdot \sqrt{0,4 \cdot 0,6/30}$$

$$0,22 \leq p \leq 0,58$$

c) Jos kriteerinä käytetään laskettua luottamusväliä, otokset 3a ja 3b vaikuttavat parhailta. Myös periaatteellisesti ne ovat parhaita, sillä kahteen muuhun otokseen voi sisältyä harhaa. Harha saattaa aiheutua siitä, että otoksissa 1 ja 2 on mukana enemmän saman sukunimisiä pareja, joista ainakin osa on keskenään sukulaisia ja suvussa saattaa olla systemaattisia käytäntöjä se mukaan, kumman nimi kuuluu tulla ensin. Mistään otoksesta ei ole riittävästi kuvattu, minkälaista satunnaistamismenettelyä on käytetty. Otoksissa 1, 3a ja 3b voidaan kuitenkin ajatella, että kyseessä on systemaattinen otanta, jossa ensimmäinen otantayksikkö on valittu satunnaisesti. Otoksen 2 kuvailusta on vaikea hyvällä tahdollakaan löytää mitään satunnaista ja se vaikuttaakin enemmän näytteeltä.

d) Olkoon  $p =$  niiden parien suhteellinen osuus, joissa naisen nimi on ensimmäisenä

$H_0$ : miehen ja naisen nimi ensimmäisenä on yhtä yleinen;  $p=0,5$

$H_1$ : miehen ja naisen nimi ensimmäisenä ei ole yhtä yleinen;  $p \neq 0,5$

Yhdistetään otokset 1, 3a ja 3b, koska niissä kaikissa on eri havainnot. Näytettä (otos 2) ei oteta mukaan.

$$n = 391+330+330=1051, X = 136+135+129=400, P = 400/1051$$

$$Z = \frac{P - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{400/1051 - 0,5}{\sqrt{\frac{0,5 \cdot 0,5}{1051}}} \approx \frac{0,38 - 0,5}{\frac{0,5}{\sqrt{1051}}} \approx \frac{-0,12 \cdot \sqrt{1051}}{0,5}$$

$$\approx \frac{-0,12 \cdot 32,42}{0,5} \approx -7,78$$

Koska  $-7,78 < -2,576$   $H_0$  voidaan hylätä. Miehen nimi ensimmäisenä on yleisempi.

(Kriittinen arvo 2,576 löytyy parhaiten t-jakauman taulukosta, kun  $f = \infty$  ja kaksisuuntaisen testin merkitsevyytaso on 1 %, mutta yhtä hyvin voisi käyttää pyöristettyä arvoa 2,58, joka kerrotaan sivulla 181.)

**Vuosi 1992, tehtävä 3**

a) Tässä on tutkittava muutamia binomijakaumia, joissa tapahtuma A=henkilö vastaa oikein, n = henkilöiden lkm eli 4 ja p = 1/vaihtoehtojen lkm (koska 1 vaihtoehto on oikea).

Merkinnöissä HUOM!  $C_n^r = \binom{n}{r}$  ja  $P(X = r) = p(r) = f(r)$

P(A tapahtuu 0 kertaa tai 4 kertaa)

5 vaihtoehtoa  $X \sim \text{Bin}(4;0,2)$

$$f(0) + f(4) = \binom{4}{0} 0,2^0 \cdot 0,8^4 + \binom{4}{4} 0,2^4 \cdot 0,8^0 = 0,4112$$

4 vaihtoehtoa  $X \sim \text{Bin}(4;0,25)$

$$f(0) + f(4) = \binom{4}{0} 0,25^0 \cdot 0,75^4 + \binom{4}{4} 0,25^4 \cdot 0,75^0 = 0,3203125$$

3 vaihtoehtoa  $X \sim \text{Bin}(4;1/3)$

$$f(0) + f(4) = \binom{4}{0} (1/3)^0 \cdot (2/3)^4 + \binom{4}{4} (1/3)^4 \cdot (2/3)^0 \approx 0,209876$$

2 vaihtoehtoa  $X \sim \text{Bin}(4;0,5)$

$$f(0) + f(4) = \binom{4}{0} 0,5^0 \cdot 0,5^4 + \binom{4}{4} 0,5^4 \cdot 0,5^0 = 0,125$$

Vastaus: tarjoamalla 2 vaihtoehtoa.

b) 4 vaihtoehtoa  $X \sim \text{Bin}(4;0,25)$

$$f(2) = \binom{4}{2} 0,25^2 \cdot 0,75^2 = \frac{4!}{2! \cdot 2!} \cdot 0,0625 \cdot 0,5625 = 0,2109375$$

c) Jälleen on tutkittava erilaisia vaihtoehtoja.

5 vaihtoehtoa  $X \sim \text{Bin}(4; 0,2)$

$$f(1) = \binom{4}{1} 0,2^1 \cdot 0,8^3 = 0,4096$$

4 vaihtoehtoa  $X \sim \text{Bin}(4; 0,25)$

$$f(1) = \binom{4}{1} 0,25^1 \cdot 0,75^3 = 0,421875$$

3 vaihtoehtoa  $X \sim \text{Bin}(4; 1/3)$

$$f(1) = \binom{4}{1} (1/3)^1 \cdot (2/3)^3 \approx 0,39506$$

2 vaihtoehtoa  $X \sim \text{Bin}(4; 0,5)$

$$f(1) = \binom{4}{1} 0,5^1 \cdot 0,5^3 = 0,25$$

Vastaus: 4 vaihtoehtoa

### **Vuosi 1993, tehtävä 3**

a) Esimerkiksi:

1. Koska luvut noudattavat tasajakaumaa välillä 1 - 39 riittää, että otamme järjestyksessä lukuja, joissa viimeinen numero on väliltä 1 - 6 ja jos luku on suurempi, nostamme niin kauan kunnes luku on välillä 1 - 6.

2. Emme huomioi lukuja 37, 38 ja 39 ja jaamme saadun luvun (1 - 36) kuudella. On oletettu, että paperilla lottonumerot ovat arpomisjärjestyksessä. Eli ei ole tarvinnut kiinnittää huomiota siihen, että lottonumerot olisi esitetty joka kierroksella nousevassa järjestyksessä.

b) Esimerkiksi:

Otetaan kolme kolikkoa esim. 10 p., 50 p. ja 1 mk kolikot. Erilaisia kolikon kombinaatioita on 8 kappaletta. Numeroidaan kombinaatiot ja jätetään kaksi vaihtoehtoa numeroimatta.

	10 p.	50 p.	1 mk	nopan luku	
1	K	L	L	1	K=kruuna L=klaava
2	K	K	L	2	
3	K	L	K	3	
4	L	K	K	4	
5	L	L	K	5	
6	L	K	L	6	
7	K	K	K	-	
8	L	L	L	-	

Heitetään kolikot pöydälle ja katsotaan tulos taulukosta. Jos tulos on K, K, K tai L, L, L heitetään uudestaan kunnes tulos on joku muu.

c) Todennäköisyys, että seikkailija voittaa hirviön on  $1-(0,6*0,7*0,8*0,9)=0,6976$  tai toisella tavalla:  $0,4+0,6*0,3+0,6*0,7*0,2+0,6*0,7*0,8*0,1=0,6976$ .

### Vuosi 1994, tehtävä 1

a) Vain yksi kahdeksasta pelistä voi olla oikein, koska kaikki pelattavat pelit ovat erilaisia samojen kolmen kohteen pelejä. Todennäköisyys, että ensimmäinen kohde on oikein:  $1-0,2=0,8$ , samoin tn, että toinen ja kolmas kohde ovat oikein on  $0,8$ . Täten todennäköisyys, että jokin peleistä on oikein on, että kaikki kolme kohdetta ovat yhtä aikaa oikein, eli  $0,8*0,8*0,8=0,512$ .

b) Todennäköisyys voittaa yhdellä viikolla on  $0,512$ . Keskimääräinen painokerroin on  $12$ , pelipanos on  $10$  mk. Kymmenen viikon odotusarvo on tällöin  $0,512*12*10*10=614,4$ . Pelaaja joutuu maksamaan  $80$  mk joka viikko, eli  $10$ :ssä viikossa  $800$  mk ja häviää  $800-614,4=185,60$  mk.

Tulos: Pelaajan tappion odotusarvo on  $185,60$  mk.

c) todennäköisyys, että jokin peleistä on oikein, on edellisen mukaan  $0,512$ .

Odotusarvon pitäisi olla enemmän kuin sijoitettu panos

$$\text{eli } 0,512*X*10 > 8*10, X > \frac{8}{0,512} = 15,625$$

Vastaus: Painokertoimen tulee olla suurempi kuin  $15,625$ .

**Vuosi 1994, tehtävä 2**

a)

$H_0$  - hypoteesi: Pelikohteen pienimmän painokertoimen koko ja pienimmän painokertoimen toteutuminen riippumattomia toisistaan.

$H_1$  - hypoteesi: Pelikohteen pienimmän painokertoimen koko ja pienimmän painokertoimen toteutuminen riippuvat toisistaan.

b)

pienimmän kertoimen koko	pienimmän kertoimen toteutuminen		$\Sigma$
	ei-toteutunut	toteutui	
1,1-1,45	3	8	11
1,5	18	5	23
$\Sigma$	21	13	34

Odotetut frekvenssit

pienimmän kertoimen koko	pienimmän kertoimen toteutuminen		$\Sigma$
	ei-toteutunut	toteutui	
1,1-1,45	$\frac{21 \cdot 11}{34} \approx 7$	$\frac{13 \cdot 11}{34} \approx 4$	11
1,5	$\frac{21 \cdot 13}{34} \approx 14$	$\frac{13 \cdot 23}{34} \approx 9$	23
$\Sigma$	21	13	34

$$\chi^2 = \frac{(3-7)^2}{7} + \frac{(8-4)^2}{4} + \frac{(18-14)^2}{14} + \frac{(5-9)^2}{9} \approx 9,2$$

Jos pyöristää lausekkeen tekijät kokonaisluvuiksi niin  $2+4+1+2=9$

Tarkka arvo (kun odotetut frekvenssit desimaalilukuja):  $\chi^2 = 8,19172$ ;  $df=(2-1) \cdot (2-1)=1$ .

Kriittinen raja  $\chi^2$  - testissä yhdellä vapausasteella 1 % merkitsevyystasolla on 6,635.

$$9,2 > 6,635$$

Johtopäätös: Pelikohteen lopputulos on riippuvainen pelikohteen pienimmän painokertoimen suuruudesta.

c) Pienin painokerroin on paras ennuste, kun painokerroin on alle 1,5. Johtopäätös perustuu ensiksikin  $\chi^2$ -testin lopputulokseen, jossa todetaan, että muuttujat eivät ole riippumattomia ja siihen, että kun painokerroin on 1,5 - 1,95, niin pienin painokerroin toteutui enää alle 30 % ja kun painokerroin oli väh. 2 niin pienimmän painokertoimen mukainen tulos ei toteutunut lainkaan.

Eli kun painokerroin on alle 1,5 niin kohteen pienintä painokerrointa voidaan pitää ennusteena. Perustuu riviprosentteihin.

### Vuosi 1994, tehtävä 3

a)

$H_0$  : korrelaatiokerroin ei poikkea 0:sta.

$H_1$  : Korrelaatiokerroin poikkeaa 0:sta.

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} \rightarrow t = \frac{0,4 \cdot \sqrt{27-2}}{\sqrt{1-0,4^2}} \rightarrow t = \frac{0,4 \cdot 5}{0,9} \rightarrow t = 2,22$$

t:n kriittinen arvo vapausasteilla  $f-2=25$  on 5 % riskitasolla kaksisuuntaisessa testissä 2,060. Koska testin tulos on suurempi kuin 5 % kriittinen arvo, niin korrelaatiokerroin poikkeaa 0:sta 5 % merkitsevyystasolla.

b) Asteikkotyypit: 1. Nominaaliasteikko

2. Järjestysasteikko

3. Intervalliasteikko

4. Suhdeasteikko

Huom. Pelkästä asteikkotyyppien tuntemisesta ei ole saanut pisteitä. On myös pitänyt osata määritellä alla olevien muuttujien asteikkotyypit oikein.

c)

Osatehtävä b.

↓

asteikko tyyppi	korr. laskeminen	muuttujan nimi
2	S	PENJÄR
1	0	TULOS
4	S+P	PKERROIN
4	S+P	VEIKOIK
4	S+P	X5
2	S	PVM
1	0	KOTI
1	0	VIERAS

d) Veikkausyhtiön kannalta on **hyvä**, että korrelaatio on voimakkaasti negatiivinen. Jos korrelaatio olisi positiivinen, niin peli tuottaisi yhtiölle tappiota, koska tällöin tapahtuisi niin, että mitä suurempi painokerroin, niin sen parempi tuotto pelaajille.

Voimakas negatiivinen korrelaatio tarkoittaa, että suuria voittoja on vähän. (Tuhoisia yhtiölle olisivat suuret poikkeamat kuvion regressiosuoralta ylöspäin, jolloin oikein veikanneita olisi paljon suurilla kertoimilla.)

(Kuvioiden perusteella ei voida kuitenkaan laskea absoluuttista arviota tuotosta, koska siinä ei ole kerrottu kuinka moni yleensä kyseistä kohdetta on veikannut. Pelaajahan saa valita 40 kohteesta 3 ja tilastossa näkyy vain kuinka hyvin tiettyä kohdetta veikanneet ovat onnistuneet veikkaamaan sen oikein, ei sitä kuinka moni sitä on veikannut ja millaisilla panoksilla.)

e) X5 on muodostettu edellisistä muuttujista seuraavalla kaavalla:

$$X5 = (\text{PKERROIN}^3 * (\text{VEIKOIK} / 100)^3)^{\frac{1}{3}} = \sqrt[3]{\text{PKERROIN}^3 * (\text{VEIKOIK} / 100)^3}$$

Muuttuja kuvaa pelikohteen keskimääräistä tuottoa. Mikäli kerroin on yli 1, niin voidaan arvioida, että tämä pelikohde tuottaa tappiota peliyhtiön kannalta. Koska jokainen peli koostuu kolmesta pelikohteesta, ei yksi pelikohde luonnollisesti vaikuta yksin pelitulokseen. VEIKOIK -muuttuja kuvaa vain %-osuutta, siinä ei näy kokonaisfrekvenssi, joka vasta kertoisi tuloksen suuruuden. Joka tapauksessa suuri arvo tarkoittaa tappiota peliyhtiölle.

f) **Suuri** painokerroin on edullisempaa. X5 kuvaa pelin tuottoa siten, että X5:n suuri arvo on epäedullinen yhtiölle. X5:n ja painokertoimen negatiivinen korrelaatio merkitsee sitä, että painokertoimen kasvaessa X5 pienenee, eli tulee edullisemmaksi yhtiölle.

### **Vuosi 1995, tehtävä 3**

#### Normaalipyhä.

Jotta kaikki 12 ykköstä tarvittaisiin, jokaisessa 6 virressä täytyy olla 2 ykköstä. Tällaisia virsiä ovat ainoastaan seuraavat:

51 viikon aikana

11, 101, 110, 112, 113, 114, 115, 116, 117, 118, 119,  
121, 131, 141, 151, 161, 171, 181, 191, 211, 311, 411, 511, 611.

Eli yhteensä 24 virttä, kaikkiaan 631:stä virrestä.

Täten todennäköisyys, että yksi virsi kuuluu näiden 24:n joukkoon on  $24/631 = 0,03803486529319$ .

A. Ensiksi lasketaan todennäköisyys, että yhdessä messussa ykköset loppuvat kesken:

Käyttäen ehdollista todennäköisyssääntöä saadaan, että yhdessä messussa todennäköisyys, että tarvitaan kaikki 12 numero ykkösen laattaa on:

Normaalipyhä

$$p_{NM} = (24/631) \cdot (23/630) \cdot (22/629) \cdot (21/628) \cdot (20/627) \cdot (19/626)$$

$$p_{NM} = 0,00000000157233$$

Todennäköisyys, että ei tarvita:  $1 - p_{NM} = 0,99999999842767$

Erikoispyhä

Koska erikoispyhänä sama virsi voi toistua useasti, todennäköisyys, että jossakin messussa tarvitaan 12 ykköstä on:

$$p_{EM} = (24/631)^6 = 0,00000000302755$$

Todennäköisyys, että ei tarvita:  $1 - p_{EM} = 0,99999999697245$

B. Lasketaan todennäköisyys, että jossain kirkossa jonain pyhänä ykköset loppuvat kesken.

Lasketaan ensiksi todennäköisyys, että missään kirkossa ei normaalipyhänä missään messussa tarvita 12 kappaletta numero ykköstä.

Tässä käytetään riippumattomien tapahtumien kertolaskusääntöä. Riippumattomana todennäköisyytenä on, että ykköset eivät lopu yhtenä pyhänä yhdessä kirkossa. Lasketaan riippumattomaa t:nä käyttäen todennäköisyys, että ykköset eivät lopu missään kirkossa 10 vuoden aikana.

$$p(\text{eivät lopu}) = (1 - p_{NM})^{(400 \cdot 51 \cdot 10)} \cdot (1 - p_{EM})^{(400 \cdot 10)}$$

$$p(\text{loppuvat}) = 1 - p(\text{eivät lopu})$$

$$p(\text{loppuvat}) = 1 - ((1 - p_{NM})^{(400 \cdot 51 \cdot 10)}) \cdot ((1 - p_{EM})^{(400 \cdot 10)})$$

$$p(\text{loppuvat}) = 0,000332809297$$

**Vuosi 1996, tehtävä 2**

a)

järjestys

  2   KAAVA 1)  3   KAAVA 2)  1   KAAVA 3)

Kaava 3 ei tuota voittoa, mutta ei tappiotakaan, kaksi muuta kaavaa tuottavat tappiota (TULOS on negatiivinen). Kaava 1) antaa paremman tuloksen. Tämä johtuu siitä, että jakauma on vasemmalle (negatiivisesti) vino. Tällöin kaava 1)

$$\text{TULOS} = \sum_{i=1}^{50} \left( \left( (pm_i - ka) / 100 \right)^5 \right)$$

painottaa pieniä poikkeavia arvoja vähemmän kuin kaava 2)

$$\text{TULOS} = \sum_{i=1}^{50} \left( \left( (pm_i - \text{median}) / 100 \right)^5 \right)$$

koska jakauman keskiarvo on pienempi kuin mediaani. Koska eniten poikkeavat arvot ovat negatiiviseen suuntaan ja pariton eksponentti säilyttää lausekkeen etumerkin, suuremmat luvut eivät voi korvata negatiivista tulosta.

b)

  2   KAAVA 1)  1   KAAVA 2)  3   KAAVA 3)

Jos pelataan peliä siten, että jakaumasta tulee oikealle vino, kaavojen järjestys vaihtuu käänteiseksi ja suurimman tappion edellä antanut kaava antaa nyt suurimman voiton. Kaava 3:n tulos on edelleen 0. Mediaani saa siis pienemmän arvon kuin keskiarvo ja tunnuslukuja suuremmat havaintoarvot painottavat lopputulosta enemmän kuin niitä pienemmät arvot.

Kaavalla 2 kannattaa pelata puolet peleistä (tai 26 peliä, riippuen miten laskee mediaanin) siten, että niiden pistemääräksi tulee nolla ja loput mahdollisimman hyvin. Tällöin Mediaani on nolla eikä tulokseen tule lainkaan sitä laskevia negatiivisia arvoja.

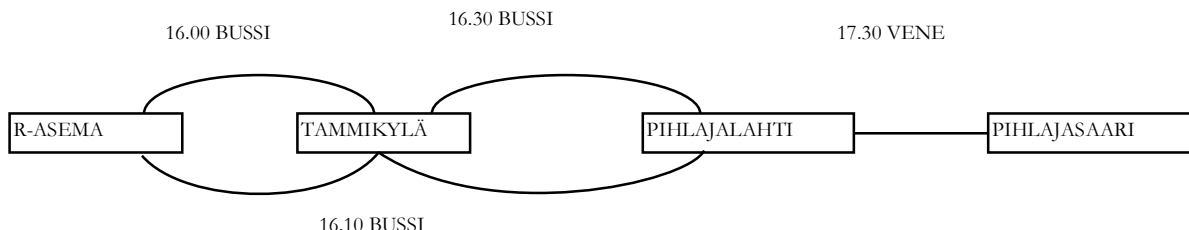
Kaavalla 1 kannattaa pelata esim. kaksi peliä mahdollisimman korkeilla pisteillä ja loput 0-pisteillä, jolloin keskiarvo jää pieneksi, mutta positiiviset ääriarvot painottavat tulosta eksponentiaalisesti enemmän kuin negatiiviset arvot (0-ka).

VANHOJEN VALINTAKOETEHTÄVIEN RATKAISUT

	Paras kaava Kaava NR_2	Toiseksi paras kaava Kaava NR_1
Fractile(.1)	<u>0</u>	<u>0</u>
Fractile(.2)	<u>0</u>	<u>0</u>
Fractile(.3)	<u>0</u>	<u>0</u>
Fractile(.4)	<u>0</u>	<u>0</u>
Fractile(.5)	<u>0</u>	<u>0</u>
Fractile(.6)	<u>≈160</u>	<u>0</u>
Fractile(.7)	<u>≈250</u>	<u>0</u>
Fractile(.8)	<u>≈280</u>	<u>0</u>
Fractile(.9)	<u>≈310</u>	<u>0</u>

### Vuosi 1996, tehtävä 3

Veikon mahdolliset tavat kulkea Pihlajasaareen ovat.



Lasketaan ensin todennäköisyys, että Veikko ehtii 16.00 bussilla ajoissa Tammikylään siis ennen kello 16.30.

$$z = \frac{30 - 29}{2} = 0,5 \quad \Phi(0,5) = 0,6915$$

Todennäköisyys, että hän ehtii perille ajoissa Pihlajalahden rantaan siis ennen kello 17.30 on

$$z = \frac{60 - 50}{5} = 2 \quad \Phi(2) = 0,9772$$

Todennäköisyys, että hän on ajoissa Pihlajalahdessa 16.00 lähtevällä bussilla on  $0,6915 \cdot 0,9772$ .

Lasketaan todennäköisyys, ehtii Pihlajalahteen 16.10 bussilla. Tn, että Veikko ei ehdi Tammikylästä lähtevään bussiin on  $1 - 0,6915 = 0,3085$ . Todennäköisyys, että hän ehtii 16.10 bussiin on 1 (Voidaan olettaa, että 16.10 bussi ei ohita matkalla Tammikylään 16.00 lähtenyttä bussia, jolloin Veikko pääsee joka tapauksessa Pihlajalahteen. Matkan jatkaminen Tammikylästä bussilla, joka lähti asemalta 16.10, vastaa tilannetta, jossa alunperin olisi noustu jo asemalla tähän bussiin.) Todennäköisyys, että hän ehtii tällä bussilla perille Pihlajalahden rantaan ennen kello 17.30 on

$$z = \frac{80 - 80}{5} = 0 \quad \Phi(0) = 0,5$$

Todennäköisyys, että hän on ajoissa Pihlajalahdessa 16.10 lähtevällä bussilla on  $0,3085 \cdot 0,5$ .

Näin todennäköisyys, että hän ehtii ajoissa Pihlajalahteen on  $0,6915 \cdot 0,9772 + 0,3085 \cdot 0,5$ .

Jos vene lähtee, on sen todennäköisyys, olla ajoissa saarella

$$z = \frac{90 - 40}{2} = 25 \quad \Phi(25) \approx 1$$

Koska vene lähtee todennäköisyydellä 0,9, on todennäköisyys olla ajoissa saarella

$$(0,6915 \cdot 0,9772 + 0,3085 \cdot 0,5) \cdot 0,9 \cdot 1 \approx 0,747$$

**Vuosi 1999, tehtävä 2**

a)

1)  $P(\text{kutonen tulee ensimmäisen kerran ensimmäisellä heittokerralla}) = 1/6 \approx 0,167$

2)  $P(\text{kutonen tulee ensimmäisen kerran toisella heittokerralla}) = (5/6) \cdot (1/6) \approx 0,139$

3)  $P(\text{kutonen tulee ensimmäisen kerran kolmannella heittokerralla}) = (5/6) \cdot (5/6) \cdot (1/6) \approx 0,116$

b)

$P(\text{ensimmäinen kutonen tulee neljännellä heittokerralla tai myöhemmin}) =$   
 $1 - P(\text{kutonen tulee ensimmäisen kerran ensimmäisellä, toisella tai kolmannella heittokerralla}) =$   
 $1 - (1/6 + (5/6) \cdot (1/6) + (5/6) \cdot (5/6) \cdot (1/6)) = 1 - 91/216 = 125/216 \approx 0,579$

c)

Mediaani on 50. prosenttipiste, joka selviää muuttujan kertymäfunktion ( vastaa havaintojakauman suhteellista summafrekvenssiä) avulla.

x (heittokert a, jolla saadaan ensimmäi nen kutonen)	P(kutonen tulee ensimmäisen kerran x:nellä heittokerralla)	P:n Kertymäfunktio
1	$1/6 \approx 0,167$	$1/6 \approx 0,167$
2	$(5/6) \cdot (1/6) \approx 0,139$	$1/6 + (5/6) \cdot (1/6) = 11/36 \approx 0,306$
3	$(5/6) \cdot (5/6) \cdot (1/6) \approx 0,116$	$1/6 + (5/6) \cdot (1/6) + (5/6) \cdot (5/6) \cdot (1/6) = 91/216 \approx 0,421$
4	$(5/6) \cdot (5/6) \cdot (5/6) \cdot (1/6) \approx 0,096$	$1/6 + (5/6) \cdot (1/6) + (5/6) \cdot (5/6) \cdot (1/6) + (5/6) \cdot (5/6) \cdot (5/6) \cdot (1/6) = 671/1296 \approx 0,518$

Kertymäfunktion arvo ylittää 50. Prosenttipisteen (=0,5) neljännellä heittokerralla. Mediaani on siis 4.

Moodi on havaintoarvo, jonka frekvenssi on suurin. Todennäköisyys vastaa suhteellista frekvenssiä ja koska todennäköisyys saada ensimmäinen kutonen tietyllä heittokerralla on suurin 1:n heittokerran kohdalla on moodi = 1.

d)

Jakauma on oikealle vino, minkä voi päätellä c-kohdan taulukosta, on moodi pieni ja keskiarvo suurin. Siis väite 5 on tosi. Huomaa, että väitteet 1 ja 2 ovat samoja, samoin 3 ja 6.

e)

$$1) P(\text{ensimmäisen kutonen tulee ensimmäisellä nostolla}) = 2/12 \approx 0,167$$

$$2) P(\text{ensimmäisen kutonen tulee toisella nostolla}) = (10/12) \cdot (2/11) \approx 0,152$$

$$3) P(\text{ensimmäisen kutonen tulee kolmannella nostolla}) = (10/12) \cdot (9/11) \cdot (2/10) \approx 0,136$$

f)

$$P(\text{ensimmäinen kutonen tulee neljännellä nostolla tai myöhemmin}) =$$

$$1 - P(\text{kutonen tulee ensimmäisen kerran ensimmäisellä, toisella tai kolmannella nostolla}) =$$

$$1 - (2/12 + (10/12) \cdot (2/11) + (10/12) \cdot (9/11) \cdot (2/10)) = 1 - 600/1320 = 720/1320 \approx 0,545$$

### Vuosi 1999, tehtävä 3

a)

Tilaisuuksissa kerätty otos ei ole satunnaisotos puolueen PUO jäsenistä. Tilaisuuksissa kyselyyn osallistuu myös muita kuin jäseniä. Lisäksi ne puolueen jäsenet, jotka menevät puoluekokoukseen eivät välttämättä käy euroehdokkaan puhetilaisuuksissa. Otos on siis valikoitunut ja epäedustava. Tämä johtaa poimintaharhaan tunnusluvuissa, joista ei tässä tapauksessa tule päätellä mitään.

b)

$H_0$ : Henkilö A ei saa yli 50 %:a äänistä.

$H_1$ : Henkilö A saa yli 50 %:a äänistä.

Tässä tilanteessa on perusteltua käyttää yksisuuntaista testausta, koska A valitaan puolueen PUO presidenttiehdokkaaksi vain, jos hän saa yli 50 %:a äänistä.

$$\hat{p} = \frac{60}{100} = 0.6 \quad n = 100 \quad p_0 = 0.5$$

$$H_0: p_0 = 0.5$$

$$H_1: p_0 > 0.5$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.6 - 0.5}{\sqrt{0.5(1-0.5)/100}} = \frac{0.1}{0.05} = 2$$

P-arvo =  $1 - \Phi(2) = 1 - 0,9772 = 0,0228$ . P-arvo < 5 %, joten  $H_0$  voidaan hylätä 5 %:n riskitasolla ja päätellä, että henkilö A valitaan puolueen PUO presidenttiehdokkaaksi.

### Vuosi 2000, tehtävä 3

a) Ehdon 1 perusteella kultakin viikolta (alkaa keskiviikosta ja loppuu tiistaihin, koska keskiviikkona tapahtuva tyhjennys ”nollaa” tilanteen) valitaan vain yksi päivä. Viikon sisällähän havainnot eivät ole riippumattomia, sillä esim. säiliön maanantain täyttöasteeseen vaikuttaa sunnuntainen täyttöaste. Näin aineisto jakautuu seitsemään jaksoon (NR:t 1-3, 4-8, 9-15, 16-22, 23-25, 26 ja 27), joista kustakin vain yksi havainto voidaan valita.

Ehdon 2 perusteella ei missään tapauksessa tule valita havainnoiksi sellaisia päiviä, joina yksikin säiliö saa arvon 10 tai 11, koska tällöin kaikki roskien viejät eivät käytä ensimmäisenä valitsemaansa säiliötä.

Ehdon 3 perusteella tulee valita päivä, joka antaa eniten informaatiota eli jakson viimeisin päivä, joka täyttää ehdon 2.

HUOM! 18.3. lauantaina joko tutkijalle on tullut tallennusvirhe tai sitten joku on dyykannut säiliötä R6, koska säiliön täyttöaste oli edellisenä päivänä 9 ja ko. päivänä vain 5. Tästä jaksosta on varmempaa valita havainnoksi tuo 17.3. Jos kokeessa valitsi 18.3., niin sekin hyväksyttiin.

Valitut päivät ovat siis:

pvm	viikon päivä	NR	R1	R2	R3	R4	R5	R6	R7	R8
3.3.	pe	<b>2</b>	4	7	7	3	6	3	0	2
11.3.	la	<b>5</b>	1	1	9	2	4	6	1	3
17.3.	pe	<b>11</b>	2	6	6	1	2	9	0	0
25.3.	la	<b>19</b>	6	9	8	4	4	4	1	1
12.4.	ke	<b>26</b>	2	1	0	0	0	1	0	0

b) Jos roskien jättäminen olisi täysin satunnaista, eri säiliöiden täyttöasteindeksit olisivat suunnilleen yhtäsuuria. Indeksit voidaan ajatella roskafrekvensseinä, jolloin niiden jakauma eri säiliöittäin olisi tasainen, jos roskien jättäminen olisi satunnaista. Tämän tutkimiseen sopii  $\chi^2$ -testi.

$H_0$ : Säiliöiden frekvenssien jakauma noudattaa tasajakaumaa.

$H_1$ : Säiliöiden frekvenssien jakauma ei noudata tasajakaumaa.

Odotettu frekvenssi kullekin säiliölle on  $120/8=15$

	R1	R2	R3	R4	R5	R6	R7	R8	Yht
$o_i$	14	24	30	11	15	17	3	6	120
$e_i$	15	15	15	15	15	15	15	15	120

$$\begin{aligned} \chi^2 &= \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \\ &= \frac{(14-15)^2}{15} + \frac{(24-15)^2}{15} + \frac{(30-15)^2}{15} + \frac{(11-15)^2}{15} + \frac{(15-15)^2}{15} + \frac{(17-15)^2}{15} + \frac{(3-15)^2}{15} + \frac{(6-15)^2}{15} \\ &= \frac{1+81+225+16+0+4+144+81}{15} = \frac{552}{15} = 36,8 \end{aligned}$$

Vapausasteet:  $f=8-1-0=7$ , jolloin merkitsevyytasolla 0,001 kriittinen arvo on  $24,321 < 36,8$  eli nollahypoteesi voidaan hylätä 0,1 % tasolla. Säiliöiden frekvenssi ei noudata tasajakaumaa, eli ihmiset eivät pane roskiaan säiliöihin satunnaisesti. Näyttäisi, että oven vierellä olevat säiliöt ovat suositumpia. Tulos on tilastollisesti erittäin merkitsevä.

### Vuosi 2001, tehtävä 1

a)

Asian voi selvittää  $\chi^2$ -yhteensopivuustestin avulla. Lasketaan ensin odotetut frekvenssit tilanteessa, jossa niaajien suhteellinen osuus olisi ollut yhtä suuri jokaisen presidentin aikana.

Niaajien kokonaismäärä:  $92+72+52+14 = 230$

Kaikkien kättelijöiden kokonaismäärä:  $323+497+394+99 = 1313$

Näin ollen niaajien odotettu suhteellinen osuus on  $230/1313 \approx 0,175$

PRESIDENTTI	Odotetut frekvenssit
KEKKONEN	$0,175 \cdot 323 \approx 57$
KOIVISTO	$0,175 \cdot 497 \approx 87$
AHTISAARI	$0,175 \cdot 394 \approx 69$
HALONEN	$0,175 \cdot 99 \approx 17$

Asetetaan hypoteesit:

$H_0$ : Niaaminen on ollut yhtä yleistä kaikkien neljän presidentin aikana.

$H_1$ : Niaaminen ei ole ollut yhtä yleistä kaikkien neljän presidentin aikana.

$\chi^2$ -testisuure on

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} = \frac{(92 - 57)^2}{57} + \frac{(72 - 87)^2}{87} + \frac{(52 - 69)^2}{69} + \frac{(14 - 17)^2}{17} \approx 28.8$$

Testisuure noudattaa  $\chi^2$ -jakaumaa vapausasteilla  $k - 1 = 4 - 1 = 3$ .

0,1 % merkitsevyytason kriittinen arvo on  $\chi^2$ -jakauman kertymäfunktio- taulukon perusteella 16.266. Koska havaittu testisuureen arvo on suurempi kuin 0,1 % merkitsevyytason kriittinen arvo, voidaan nollahypoteesi hylätä. Niiaaminen ei ole ollut yhtä yleistä kaikkien neljän presidentin aikana.

b)

Asetetaan hypoteesit:

$$H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2$$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{54 \cdot 0,24 + 127 \cdot 0,14}{54 + 127} \approx 0,17$$

$$Z = \frac{P_1 - P_2}{\sqrt{P(1 - P) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0,24 - 0,14}{\sqrt{0,17(1 - 0,17) \left( \frac{1}{54} + \frac{1}{127} \right)}} \approx 1,64$$

Testi on kaksisuuntainen. 5 % merkitsevyytastolla kriittinen arvo on 1,96. Koska testisuure on pienempi, nollahypoteesi jää voimaan. Aineiston perusteella ei voi väittää, että Kekkonen viimeisenä vuonna 1980 ja ensimmäisenä aineistossa olevana Koiviston vuonna 1986 niiaamisessa olisi ollut eroa.

## Vuosi 2001, tehtävä 2

a)

$$P(\text{vähintään yksi niiaa}) = 1 - P(\text{kukaan ei niiaa})$$

$$1 - 0,5 \cdot 0,5 \cdot 0,5 = 0,875$$

b)

$$P(N1 \text{ niiaa ja } N2 \text{ niiaa}) + P(N1 \text{ ei niiaa ja } N2 \text{ niiaa})$$

$$0,5 \cdot 0,8 + 0,5 \cdot 0,5 = 0,65$$

c)

Tässä vaihtoehdot kannattaa taulukoida

N1	N2	N3	P1	P2	P3	P1 · P2 · P3
niiaa	niiaa	niiaa	0,5	0,8	0,8	$0,5 \cdot 0,8 \cdot 0,8 = 0,32$
niiaa	ei niiaa	niiaa	0,5	0,2	0,5	$0,5 \cdot 0,2 \cdot 0,5 = 0,05$
ei niiaa	niiaa	niiaa	0,5	0,5	0,8	$0,5 \cdot 0,5 \cdot 0,8 = 0,2$
ei niiaa	ei niiaa	niiaa	0,5	0,5	0,5	$0,5 \cdot 0,5 \cdot 0,5 = 0,125$

Kokonaistodennäköisyys sille, että kolmantena tullut on niannut on  $0,32+0,05+0,2+0,125 = 0,695$ .

Todennäköisyys sille, että ensimmäisenä tullut on niannut, kun tiedetään että kolmantena tullut on niannut on  $(0,32+0,05)/0,695 \approx 0,532$ .

## Vuosi 2002, tehtävä 1

### 1.1.a

Koska heittojen lkm on pieni, tulee käyttää binomijakaumasta saatavia todennäköisyyksiä.

$H_0$ : Kruunan ja klaavan tn yhtäsuuri,  $P(\text{kruuna})=0,5$

$H_1$ : Kruunan ja klaavan tn erisuuri,  $P(\text{kruuna}) \neq 0,5$

$n=6$ ;  $k=0,1,2$ ;  $p=0,5$

$$P(X = k) = p(k) = \binom{n}{k} p^k q^{n-k} \quad P(X = 0) = p(0) = \binom{6}{0} 0,5^0 0,5^6 = \frac{6!}{0!6!} \cdot 0,5^6 \approx 0.016$$

$$P(X = 1) = p(1) = \binom{6}{1} 0,5^1 0,5^5 = \frac{6!}{1!5!} \cdot 0,5^6 \approx 0.094$$

$$P(X = 2) = p(2) = \binom{6}{2} 0,5^2 0,5^4 = \frac{6!}{2!4!} \cdot 0,5^6 \approx 0.234$$

$$p = 0,016+0,094+0,234=0,344 > 0,05$$

$H_0$  jää voimaan. Kruunan ja klaavan tn:n ei voi uskoa olevan erisuuri.

### 1.1.b

Nyt heittojen lukumäärä on suuri ja siksi käytetään normaalijakaumaan perustuvaa testiä.

$H_0$ : Kruunan ja klaavan tn yhtäsuuri,  $P(\text{kruuna})=0,5$

$H_1$ : Kruunan ja klaavan tn erisuuri,  $P(\text{kruuna}) \neq 0,5$

$n=2500$ ;  $X=1300$ ;  $P=0,52$ ;  $p=0,5$

$$Z = \frac{P - p}{\sqrt{pq/n}} = \frac{0,52 - 0,5}{\sqrt{0,5 \cdot 0,5 / 2500}} = \frac{0,02}{0,01} = 2$$

$$Z = \frac{X - np_0}{\sqrt{np_0q_0}} = \frac{1300 - 2500 \cdot 0,5}{\sqrt{2500 \cdot 0,5 \cdot 0,5}} = \frac{50}{25} = 2$$

$$2 \cdot \Phi(2) = 2 \cdot 0,0228 = 0,0456$$

$Z = 2 > 1,96$ , 5% kriittinen arvo, 2-suuntainen testaus.

$H_0$  Hylätään. Saksan eurolla kruunia tulee enemmän kuin klaavoja.

1.2.

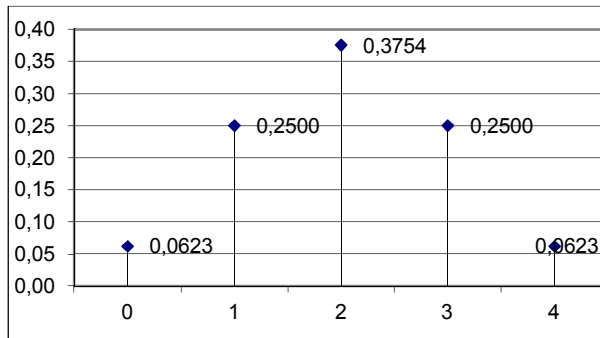
Lasketaan ensin kummankin maan binomitodennäköisyydet ja sitten todennäköisyyksien laskusäännöillä yhdistetään tapauksia.

$$P(X = k) = p(k) = \binom{n}{k} p^k q^{n-k}$$

				Hollanti				
				n	2	2	2	
Belgia				p	0.48	0.48	0.48	
				k	0	1	2	
				P(k)	0.2704	0.4992	0.2304	1
n	p	k	P(k)					
2	0.52	0	0.2304	0.0623	0.115016	0.053084		
2	0.52	1	0.4992	0.134984	0.249201	0.115016		
2	0.52	2	0.2704	0.073116	0.134984	0.0623		
				1				1

Yllä taulukossa laskettu binomitapauksien tulot. Esim.  $0.2704 \cdot 0.2304 = 0.0623$ . Alla ao. tulot laskettu yhteen.

kr.		
lkm		
0		0.06230
1	$0.134984 + 0.115016$	$= 0.25000$
2	$0.073116 + 0.249201 + 0.053084$	$= 0.37540$
3	$0.134984 + 0.115016$	$= 0.25000$
4		$= 0.06230$
		1.00000



1.3.

Kyseessä on kahden suhteellisen osuuden testi.

$H_0$ : Kruunien osuus on sama ransk. ja esp. kolikolla,  $P_1=P_2$        $n_1=5000$ ;  
 $P_1=2530/5000=0,506$

$H_1$ : Kruunien osuus ei ole sama ransk. ja esp. kolikolla,  $P_1 \neq P_2$        $n_2=5000$ ;  
 $P_2=2480/5000=0,496$

$$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{2530 + 2480}{5000 + 5000} = 0,501$$

$$Z = \frac{P_1 - P_2}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} = \frac{0,506 - 0,496}{\sqrt{0,501 \cdot 0,499 \cdot (1/5000 + 1/5000)}} \approx \frac{0,01}{0,01} = 1$$

$$2 \cdot \Phi(1) = 2 \cdot 0,1587 = 0,3174$$

$Z = 1 < 1,64$ , 10% kriittinen arvo, 2-suuntainen testaus.

$H_0$  jää voimaan. Ranskan ja Espanjan euroilla ei voi sanoa olevan eroa.

**Vuosi 2002, tehtävä 2**

2.1.

Mitään muuttujaa ei tule jättää pois, sillä t-arvo (vapausasteilla  $121-21=100$ ) on 1,66 (2-suuntainen testaus). Koska kaikkien regressiokertoimien t-arvojen itseisarvot ovat suurempia kuin 1,66, kaikki muuttujat ovat mallissa tilastollisesti merkitseviä.

2.2.

Tietojen perusteella pystyy selvittämään, kumpi havainnoista,  $H_1$  vai  $H_2$ , saa suuremman arvon ennustemuuttujalla. Koska muuttujilla  $x_4, x_5, \dots, x_{20}$  havainnoilla on samat arvot, näiden

muuttujien vaikutus ennustemuuttujaan on molemmilla havainnoilla sama. Näiden kahden havainnon ennustemuuttujan arvojen ero voidaan siis laskea muuttujien  $x_1$ ,  $x_2$  ja  $x_3$  arvojen perusteella.

$$H1: 2,96 \cdot 1 - 3,13 \cdot 2 + 2,01 \cdot 3 = 2,73$$

$$H2: 2,96 \cdot 3 - 3,13 \cdot 1 + 2,01 \cdot 2 = 9,77$$

Havainto H2 saa siis suuremman arvon ennustemuuttujalla.

### Vuosi 2003, tehtävä 1

a)

Naimattomien miesten maksimimäärä on 120, mikä on mahdollista vain kahdella tavalla: A) Kaikkiin perheisiin syntyy ensimmäisenä poika tai B) kaikkiin perheisiin syntyy ensin tyttö ja sitten poika. Todennäköisyys, että jompikumpi tapahtuu on

$$P(A) + P(B) = 0,5^{120} + 0,5^{120} \cdot 0,5^{120} = 0,5^{120} + 0,5^{240}$$

b)

Naimattomien naisten maksimimäärä on 240, mikä on mahdollista vain jos kaikkiin perheisiin syntyy kaksi tyttöä.

$$P(C) = 0,5^{120} \cdot 0,5^{120} = 0,5^{240} \quad \text{tai} \quad P(C) = (0,5 \cdot 0,5)^{120} = 0,25^{120}$$

c)

Ensin on selvitettävä 1. ja 2. lapseksi syntyneiden poikien ja tyttöjen lukumäärien odotusarvot. Koska perheitä on 120, niin ensimmäinen lapsi on poika  $0,5 \cdot 120 = 60$  perheessä eikä näihin synny toista lasta. Ensimmäinen lapsi on tyttö  $0,5 \cdot 120 = 60$  perheessä ja näistä  $0,5 \cdot 60 = 30$  toinen lapsi on poika ja  $0,5 \cdot 60 = 30$  toinen lapsi on tyttö. Sekä poikia että tyttöjä syntyy siis 90. Merkitään eri lapsiperheitä seuraavasti: P= 1. lapsi poika, TP= 1. lapsi tyttö ja 2. poika, TT= molemmat lapset tyttöjä. P-poikia on siis 60, TP-poikia 30, TP-tyttöjä 30 ja TT-tyttöjä 60. P-pojat valitsevat ensin vaimonsa. Koska he valitsevat vaimonsa satunnaisesti TP- ja TT-tyttöjen joukosta, valitsevat he TP-tytön todennäköisyydellä  $1/3$  ja TT-tytön todennäköisyydellä  $2/3$  eli  $2/3 \cdot 60 = 40$  P-poikaa valitsee vaimokseen TT-tytön ja  $1/3 \cdot 60 = 20$  P-poikaa valitsee vaimokseen TP-tytön. Jäljelle jää siis 20 TT-tyttöä ja 10 TP-tyttöä. Koska TP-pojat eivät voi mennä naimisiin TP-tyttöjen kanssa, jää 10 TP-poikaa ja 10 TP-tyttöä naimattomaksi. Siis yhteensä 20 henkilöä.

**Vuosi 2004, tehtävä 2**

a)

$H_0$ : Tapaukset, joissa kahden toisiaan seuraavan luvun summa on 7, ovat jakautuneet kuten harhattomalla nopalla,  $e_i=360/36=10$

$H_1$ : Tapaukset, joissa kahden toisiaan seuraavan luvun summa on 7, eivät ole jakautuneet kuten harhattomalla nopalla,  $e_i \neq 10$

$$\chi^2 = \sum_i \frac{(o_i - e_i)^2}{e_i} \quad f = k - 1$$

$$\chi^2 = \frac{(14-10)^2}{10} + \frac{(17-10)^2}{10} + \frac{(24-10)^2}{10} + \frac{(18-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(17-10)^2}{10} = 39$$

$$f = 6 - 1 = 5$$

1 % merkitsevyystasolla kriittinen arvo, kun  $f=5$ , on  $15,086 < 39$ , jolloin nollahypoteesi hylätään. Tapaukset, joissa kahden toisiaan seuraavan luvun summa on 7, eivät ole jakautuneet kuten harhattomalla nopalla

b)

$H_0$ : Tapauksen ”kahden perättäisen nopanpainalluksen summa on 7” suhteellinen osuus on molemmissa peleissä sama,  $p_1 = p_2$

$H_1$ : Tapauksen ”kahden perättäisen nopanpainalluksen summa on 7” suhteellinen osuus ei ole molemmissa peleissä sama,  $p_1 \neq p_2$

$$Z = \frac{P_1 - P_2}{\sqrt{P(1-P)(1/n_1 + 1/n_2)}} \quad P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$P_1 = \frac{14 + 17 + 24 + 18 + 14 + 17}{360} = \frac{104}{360} \approx 0,289 \quad P_2 = \frac{14 + 5 + 13 + 13 + 13 + 9}{360} = \frac{67}{360} \approx 0,186$$

$$P = \frac{104 + 67}{720} = \frac{171}{720} = 0,2375$$

$$Z = \frac{0,289 - 0,186}{\sqrt{0,2375 \cdot 0,7625 \cdot (1/360 + 1/360)}} \approx 3,247$$

1 % merkitsevyystasolla kaksisuuntaisessa testauksessa kriittinen on 2,576. Koska testisuureen itseisarvo  $3,247 > 2,576$ , hylätään nollahypoteesi. Tapauksen ”kahden perättäisen nopanpainalluksen summa on 7” suhteellinen osuus ei ole molemmissa peleissä sama. (Testisuureen p-arvo on n. 0,0012 eli 0,12 %, joka on pienempi kuin 1 %.)

**Vuosi 2004, tehtävä 3**

a)

'Etsijä' pääsee 'aarteen' luo, vaikka kuutionopalla heitettäisiin luku 0, vain siinä tapauksessa, että 'etsijän' ja 'aarteen' paikka on täsmälleen sama.

$$P(A) = \frac{16}{16} \cdot \frac{1}{16} = \frac{1}{16} = 0,0625$$

b)

'Etsijä' ei pääse 'aarteen' luo, vaikka kuutionopalla heitettäisiin luku 5, vain siinä tapauksessa, että 'aarteen' paikka on jossakin kulmassa ja 'etsijän' paikka on vinosti vastakkaisessa kulmassa.

$$P(B) = \frac{4}{16} \cdot \frac{1}{16} = \frac{4}{256} = \frac{1}{64} = 0,015625$$

c)

Erilaisia 'etsijä' – 'aarre' tilanteita eli mahdollisia reittejä on yhteensä  $16 \cdot 16 = 256$  kpl. Mahdolliset reitin pituudet ovat 0, 1, 2, 3, 4, 5 ja 6 polkua. Näistä itse asiassa 0 ja 6 pituisen polun t:n:t laskettiin kohdissa 3a ja 3b. T:n:ien selvittämistä helpottaa se, että pelilaudalla nurkkapisteet A1, A4, D1 ja D4 (4 kpl) ovat identtisiä etäisyyksien kannalta, samoin keskuspisteet B2, B3, C2 ja C3 (4 kpl) sekä sivupisteet A2, A3, B1, B4, C1, C4, D2 ja D3 (8 kpl).

reitin pituus ( $x_i$ )	nurkka (4 kpl)	keskus (4 kpl)	sivu (8 kpl)	yhteensä	$P(X = x_i)$
0	1	1	1	$4 \cdot 1 + 4 \cdot 1 + 8 \cdot 1 = 16$	$16/256 = 1/16$
1	2	4	3	$4 \cdot 2 + 4 \cdot 4 + 8 \cdot 3 = 48$	$48/256 = 3/16$
2	3	6	4	$4 \cdot 3 + 4 \cdot 6 + 8 \cdot 4 = 68$	$68/256 = 17/64$
3	4	4	4	$4 \cdot 4 + 4 \cdot 4 + 8 \cdot 4 = 64$	$64/256 = 1/4$
4	3	1	3	$4 \cdot 3 + 4 \cdot 1 + 8 \cdot 3 = 40$	$40/256 = 5/32$
5	2	0	1	$4 \cdot 2 + 4 \cdot 0 + 8 \cdot 1 = 16$	$16/256 = 1/16$
6	1	0	0	$4 \cdot 1 + 4 \cdot 0 + 8 \cdot 0 = 4$	$4/256 = 1/64$
				256	

d)

Kysytty todennäköisyys saadaan reitin pituuksien t:n:ien ja kuutionopan t:n:ien yhdistelmänä.

$$P(D) = \frac{1}{16} \cdot \frac{6}{6} + \frac{3}{16} \cdot \frac{5}{6} + \frac{17}{64} \cdot \frac{4}{6} + \frac{1}{4} \cdot \frac{3}{6} + \frac{5}{32} \cdot \frac{2}{6} + \frac{1}{16} \cdot \frac{1}{6} + \frac{1}{64} \cdot \frac{0}{6} = \frac{224}{384} = \frac{7}{12} \approx 0,5833$$

**Vuosi 2005, tehtävä 3**

a)

Taulukon 3 perusteella valmiin nimilistan esittäminen auttaa nimien yleisyyden arvioinnissa. Muuttuja LKM kuvaa nimien todellista yleisyyttä ja sen korrelaatio muuttujaan B2KA on 0,77 ( $r^2 \approx 0,59$ ) ja muuttujaan B1PIST 0,67 ( $r^2 \approx 0,45$ ). Valmiin listan kohdalla yhteys todelliseen yleisyyteen on suurempi kuin avoimen kyselyn kohdalla. Tilanne on sama, kun katsotaan muuttujien JNR, B1JNR ja B2JNR korrelaatioita.

b)

Venäläisten osuus koko Virossa on

$$\frac{6624 + 1625 + 3260 + 2811}{64857} = \frac{14320}{64857} \approx 0,221$$

$H_0$ : Venäläisten osuus Etelä-Virossa on 22,1 %,  $p = 0,221$

$H_1$ : Venäläisten osuus Etelä-Virossa on alle 22,1 %,  $p < 0,221$

$$P = \frac{200}{4730} \approx 0,042$$

$$Z = \frac{\frac{P - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}{\sqrt{\frac{0,221 \times 0,779}{4730}}} = \frac{0,042 - 0,221}{\sqrt{\frac{0,221 \times 0,779}{4730}}} \approx -29,67$$

Yksisuuntaisen testin kriittinen arvo 0,1 % merkitsevyystasolla on 3,09. Koska testisuureen itseisarvo  $|-29,67| > 3,09$ , hylätään  $H_0$  ja päätellään, että Etelä-Virossa on vähemmän venäläisiä kuin koko Virossa.

c)

$H_0$ : 10 yleisimmän nimen keskiarvo on yhtä suuri kuin 10 harvinaisimman nimen,

$$\mu_1 = \mu_3$$

$H_1$ : 10 yleisimmän nimen keskiarvo on suurempi kuin 10 harvinaisimman nimen,

$$\mu_1 > \mu_3$$

$$s^2 = \frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(10_1 - 1) \times 84,17^2 + (10 - 1) \times 29^2}{(10 - 1) + (10 - 1)} = 3962,79445$$

$$s \approx 62,951$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{160,9 - 70,7}{62,951 \times \sqrt{\frac{1}{10} + \frac{1}{10}}} \approx 3,204$$

Vapausasteet:  $n_1 + n_2 - 2 = 10 + 10 - 2 = 18$ , Yksisuuntaisessa testissä 0,5% merkitsevyystasolla kriittinen arvo on 2,878 ja 0,1% merkitsevyystasolla 3,611. Koska testisuureen arvo  $2,878 < 3,204 < 3,611$ ,  $H_0$  voidaan hylätä 0,5% merkitsevyystasolla, mutta ei 0,1% merkitsevyystasolla.

d)

Virolaisten nimien yleisyys koko Virossa on vaikuttanut opiskelijoiden arvioon. Taulukossa 6 muuttujien CLKM ja B1PIST välinen korrelaatio (0,87) on korkeampi kuin muuttujien LKM ja B1PIST välinen korrelaatio (0,69). Venäläisiä nimiä opiskelijat eivät kirjanneet avoimessa kyselyssä lainkaan. Koska taulukon 5 korrelaatiot on laskettu kaikista nimistä, venäläisten nimien mukanaolo vaikuttaa korrelaatioihin. Kun tarkastelee kuvaa 1 erikseen virolaisten ja venäläisten nimien osalta muuttujien CLKM ja B1PIST sekä LKM ja B1PIST kohdalla, havaitsee saman.

## **Lopuksi**

Rutkasti intoa kokeeseen valmistautumiseen!