VALINTAKOE 2019

Psykologia aineisto- ja tehtävävihko

Tiistai 7.5.2019 klo 9–13

Valintakoeyhteistyö:

Helsingin Yliopisto
Jyväskylän Yliopisto
Itä-Suomen Yliopisto
Tampereen Yliopisto
Turun Yliopisto

Lue huolellisesti kaikki ohjeet läpi

Kokeen alussa

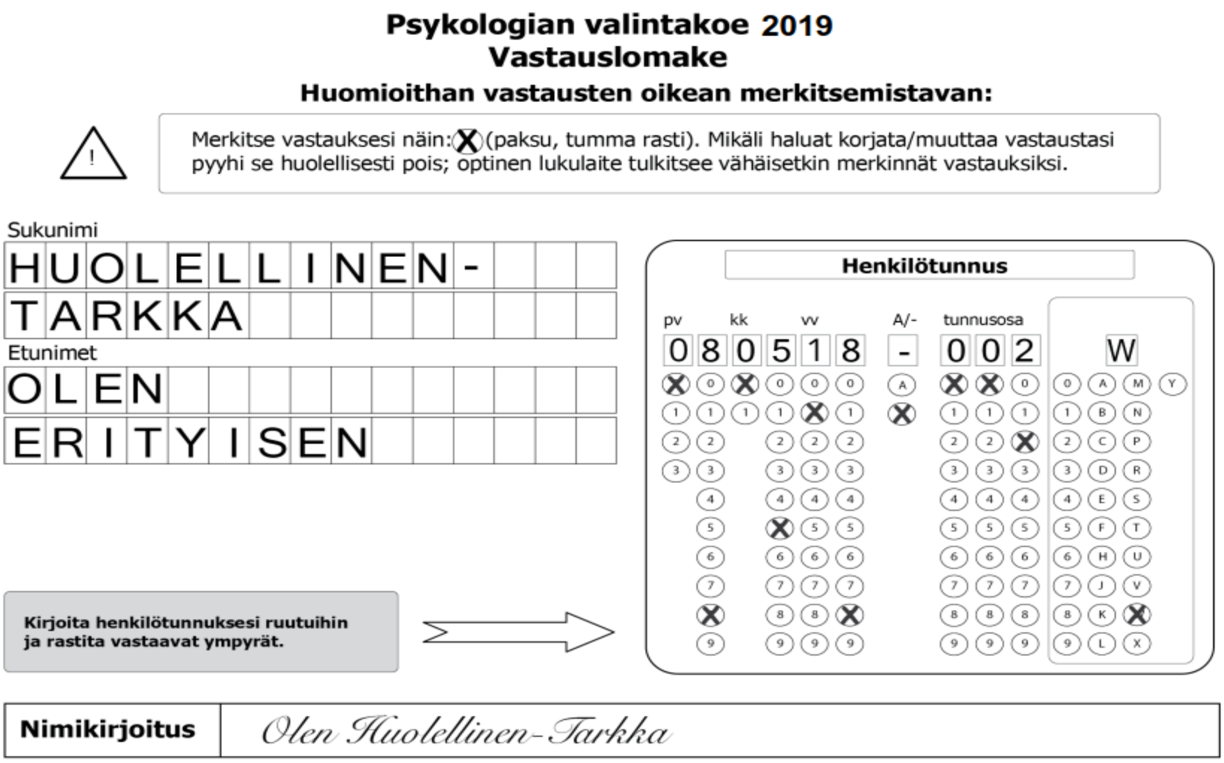
- Tarkasta, että sinulla on laskin. Tarkasta, että sinulla on aineisto- ja tehtävävihko, jossa on 35 sivua ja kaksi A4-kokoista vastauslomaketta (vastauslomakkeet 1–2). Huomioi, että molemmat vastauslomakkeet ovat yksipuolisia. Aineisto- ja tehtävävihkon sivuilla 29-35 ovat liitteenä tilastolliset taulukot ja käytettävät kaavat.

- Tarkasta, että vastauslomakkeiden oikeassa yläkulmassa on sama numero.

- Kirjoita nimesi kaikkiin vastauslomakkeisiin ja henkilötunnuksesi vastauslomakkeeseen 1. Käytä alla olevan mallin mukaisia merkkejä!

A B C D E F G H I J
K L M N O P Q R S T
U V W X Y Z Å Ä Ö
0 1 2 3 4 5 6 7 8 9

- Merkitse vastauslomakkeeseen 1 henkilötunnuksesi myös rastimalla oikeat soikiot lomakkeiden optista lukemista varten. Katso mallia alla olevasta kuvasta.


Psykologian valintakoe 2019
Vastauslomake
Huomioithan vastausten oikean merkitsemistavan:

Merkitse vastauksesi näin: (paksu, tumma rasti). Mikäli haluat korjata/muuttaa vastaustasi pyyhi se huolellisesti pois; optinen lukulaite tulkitsee vähäisetkin merkinnät vastauksiksi.

Sukunimi
HUOLELLINEN-
TARKKA
Etunimet
OLEN
ERITYISEN

Henkilötunnus
pv kk vv A/- tunnusosa
080518 - 002 W

Kirjoita henkilötunnuksesi ruutuihin ja rastita vastaavat ympyrät.

Nimikirjoitus: Olen Huolellinen-Tarkka

- Kirjoita vastauslomakkeessa 1 olevaan laatikkoon nimikirjoituksesi merkkinä siitä, että olet tarkistanut ja suorittanut edellä mainitut asiat.

Kokeen aikana

- Lue tehtävien ohjeet **huolellisesti**. Jos et noudata ohjeita, saatat menettää pisteitä.

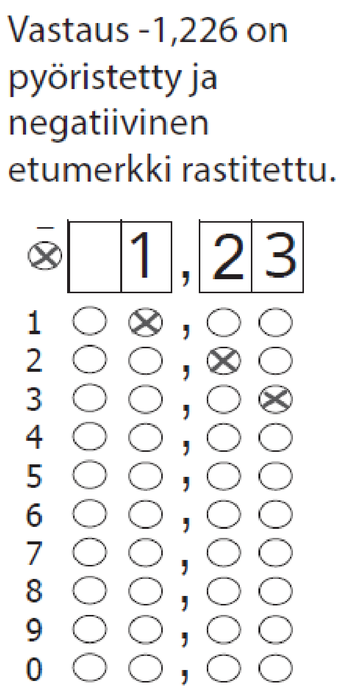
- Jokaiseen tehtävään on oma aineistonsa. Ole huolellinen, että vastaat kuhunkin tehtävään oikean aineiston pohjalta. Jos tehtävän aineisto on ristiriidassa muun tiedon kanssa, vastaa tehtävän aineiston perusteella. Muista, että arvioidessasi tehtävässä esitettyjen väitteiden totuutta, arvioit **koko väitteen** totuutta tehtävän aineiston pohjalta ja tehtävän ohjeiden mukaisesti.

- Merkitse tehtävien vastaukset huolellisesti optiseen vastausosaan. Kun vastaus on luku, merkitse luku niin että yhden suorakaiteen sisään tulee yksi numero. Luvut merkitään kymmenjärjestelmän mukaisesti. Lukujen pyöristyssääntö: Viimeinen mukaan tuleva numero korotetaan yhdellä, jos ensimmäinen pois jäävä numero on 5, 6, 7, 8 tai 9. Älä merkitse positiivista etumerkkiä. Tehtävissä, joissa negatiivinen arvo on mahdollinen, merkitse rasti negatiivista etumerkkiä tarkoittavaan soikioon, jos saamasi tulos on negatiivinen. Jos osatehtävässä vaaditaan desimaalien merkitsemistä, niin vastauslomakkeessa on osatehtävän kohdalla desimaalierotin. Jos desimaalierotinta ei ole, luku merkitään kokonaislukuna.

- Arvostelu perustuu vastauslomakkeisiin merkittyihin valintoihin (rastitettuihin soikioihin) ja pelkästään kirjoitettu vastaus ei ole riittävä!

- Jos jokin merkintä on epäselvä, tulkitaan kohta virheellisesti täytetyksi.

- Pidä huolta siitä, että merkinnät, jotka teet vastauslomakkeisiin ovat yksiselitteisiä ja selviä. Tee vastausmerkintäsi piirtämällä lyijykynällä rasti valitsemasi vaihtoehdon mukaisen soikion sisään (katso oheinen esimerkki). Jos haluat muuttaa vaihtoehtosi tai poistaa sen, pyyhi pyyhekumilla siististi vanha merkintäsi pois ja rasti uusi soikio.



Vastauslomakkeisiin ei saa tehdä mitään muita merkintöjä.

- Pidä koemateriaalisi niin, että lähelläsi istuvat hakijat eivät pysty katsomaan vastauksiasi ja merkintöjäsi. Erityisesti pidä se vastauslomake, joita et ole täyttämässä, suojassa uteliailta katseilta.

- Tehtävistä saa pisteitä vasta, kun riittävä, tiettyä arvaamistodennäköisyyttä suurempi määrä osatehtäviä on oikein.

- Osassa laskutehtäviä lukuja kannattaa sieventää mahdollisimman pieniksi ennen kuin laskee vastauksen laskimella.

- Voit luonnostella vastauksiasi konseptipaperille. Konseptipaperille tekemiäsi merkintöjä ei huomioida arvostelussa. Olet saanut yhden arkin konseptipaperia. Voit tarvittaessa pyytää lisää konseptipaperia valvojalta.

Vastausaika

Vastausaika päättyy neljän tunnin kuluttua kokeen aloittamisesta.
Saat poistua salista aikaisintaan klo 10:00.

Pisteyttäminen

Valintakoetehtävistä saatava yhteenlaskettu raakapistemäärä muutetaan valintakoepisteiksi välille 0,000 – 120,000 seuraavasti:

• Kaikista psykologian alan valintakoeyhteistyönä järjestettävään valintakokeeseen osallistuneista hakijoista ne hakijat, jotka kuuluvat 1 % parhaiten vastanneiden joukkoon, saavat lopulliseksi valintakoepistemääräkseen 120,000 pistettä.

• Skaalauskerroin lasketaan jakamalla luku 120,000 sen hakijan raakapistemäärällä, joka sijoittuu viimeisenä 1 % parhaiten vastanneiden joukkoon.

• Muiden hakijoiden lopullinen valintakoepistemäärä lasketaan kertomalla hakijan raakapistemäärä skaalauskertoimella. Pistemäärä pyöristetään kolmen desimaalin tarkkuudelle.

Tullakseen huomioiduksi valinnassa täytyy valintakoe suorittaa hyväksytysti. Minimipisteraja (valintakokeesta edellytetty vähimmäispistemäärä) on puolet skaalatuista maksimipisteistä (= 60 pistettä).

Tehtävistä saatava pistemäärä vaihtelee tehtävän vaikeusasteen mukaan. Joissain tehtävissä vastaus voi olla osittain oikein, tällöin täysin oikeita vastauksia painotetaan enemmän kuin osittain oikeita vastauksia.

Osatehtävien suhteellinen painotus riippuu kokeen loppupisteistä. Koska ennen koetta ei ole mahdollista tietää, miten parhaiten menestynyt hakija menestyy eri osatehtävissä, ei voida sanoa jokaisen osatehtävän täsmällistä osuutta loppupisteistä. Osuudet ovat kuitenkin suuntaa-antavia ja kertovat tehtävien suhteellisen painon toisiin tehtäviin nähden. Tehtävistä saatavat maksimipistemäärät suhteessa koko kokeeseen ovat likimäärin seuraavat:

Tehtävä 1.1. 30 %
Tehtävä 1.2. 15 %
Tehtävä 2.1. 35 %
Tehtävä 2.2. 20 %

Kun aiot palauttaa koepaperit

- Tarkista ennen palautusta, että olet kirjoittanut nimesi molempiin vastauslomakkeisiin ja lisäksi henkilötunnuksesi vastauslomakkeeseen 1 sekä rastinut henkilötunnuksesi mukaiset soikiot oikein.

- Järjestä vastauslomakkeet numerojärjestykseen (1, 2). Järjestä niiden perään palautuksen tarkistusta varten aineisto- ja tehtävävihko ja konseptipaperit. Muista palauttaa myös laskin.

- Palauta molemmat vastauslomakkeet, vaikket olisikaan tehnyt joitakin tehtäviä tai mitään tehtäviä.

- Palauttaessasi kokeen esitä ensimmäiseksi henkilöllisyystodistuksesi.

- Kokeeseen osallistuminen ja vastauslomakkeiden palautus merkitään palautuksen yhteydessä osallistujalistaan kokeen valvojan toimesta. Tarvittaessa saat kokeen valvojalta erillisen todistuksen valintakokeeseen osallistumisesta.

- Aineisto- ja tehtävävihkon saat halutessasi ottaa mukaan, kun palautusta valvova valintakoevalvoja on suorittanut henkilöllisyyden tarkistuksen ja tehnyt osallistumismerkinnän sekä vastauslomakkeiden palautusmerkinnät.

Onnea kokeeseen!

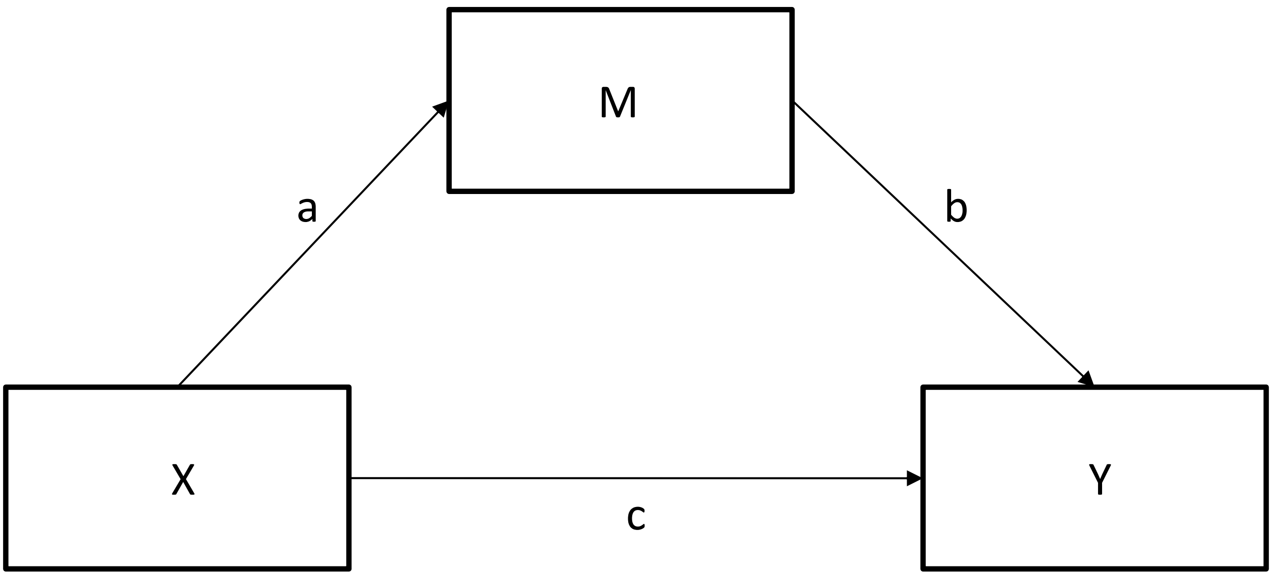
Tehtävä 1

Vastaa ennakkomateriaalin ja kokeessa jaetun aineiston perusteella. Vastaa tehtävän 1.1 osatehtäviin 1.1.1 – 1.1.2 vastauslomakkeisiin 1 ja 2 ja tehtävään 1.2 vastauslomakkeeseen 2, osatehtävien ohjeiden mukaisesti. Vaikka kaikki aineistot eivät perustu todelliseen tutkimukseen, sinun tulee olettaa esitettyjen aineistojen olevan totta. Tehtävän 1.2 osatehtävissä voi olla useampi kuin yksi vastausvaihtoehto oikein, mutta jokaisessa kohdassa on ainakin yksi vastausvaihtoehto oikein. Osatehtävissä on löydettävä kaikki ja vain kaikki oikeat vastausvaihtoehdot, jotta saisi täydet pisteet. Rasti valitsemiasi vastausvaihtoehtoja vastaavat soikiot vastauslomakkeisiin 1 – 2. Ellet ole vastannut osatehtävään mitään, tulkitaan vastaus vääräksi. Merkitse luvut ohjeiden mukaisesti optiseen vastausosaan.

Tehtävä 1.1

Tehtävä 1.1.1

Välittävien eli medioivien vaikutusten tarkastelu on keskeisessä roolissa psykologian tutkimuksessa tutkittaessa tiettyjen ilmiöiden taustalla vaikuttavia mekanismeja. Välittävällä vaikutuksella tarkoitetaan sitä, että jokin kolmas tekijä, jota kutsutaan mediaattoriksi, välittää tietyn ennustavan tekijän vaikutusta ennustettavaan tekijään. Välittävää vaikutusta voidaan kuvata kuvassa 1 esitetyllä polkukaaviolla.



Kuva 1: Välittävien vaikutusten polkukaavio

Kuvassa perusajatuksena on, että ennustavan tekijän (X) vaikutus välittyy välittävän tekijän eli mediaattorin (M) kautta polkua a ja b pitkin ennustettavaan tekijään (Y). On mahdollista, että ainoastaan osa ennustavan tekijän vaikutuksesta välittyy ja osa vaikutuksesta on suoraa efektiä, joka kulkee suoraan polun c kautta. Mikäli polku c on 0, eli kaikki vaihtelu X:n ja Y:n välillä kulkee mediaattorin kautta, puhutaan täydellisestä mediaatiosta ja mikäli c ≠ 0, eli osa vaihtelusta on suoraa, puhutaan osittaisesta mediaatiosta.

Mediaatiota voidaan testata ns. Baronin ja Kennyn askelilla, jotka muodostuvat sarjasta korrelaatio- ja regressioanalyysejä.

Askel 1: On osoitettava, että ennustava muuttuja (X) on tilastollisesti merkitsevässä yhteydessä ennustettavaan muuttujaan (Y). Tämä voidaan osoittaa joko yhden selittäjän regressioanalyysillä, jossa Y-muuttujaa ennustetaan X-muuttujalta tai yksinkertaisesti tarkastelemalla muuttujien välisiä korrelaatioita. Tämä askel varmistaa, että muuttujien välillä on yhteys, jonka on mahdollista välittyä välittävän tekijän kautta.

Askel 2: On osoitettava, että ennustavan muuttujan ja mediaattorin välillä on tilastollisesti merkitsevä yhteys. Tämä osoitetaan tyypillisesti regressioanalyysillä, jossa mediaattoria (M) ennustetaan ennustavalla muuttujalla (X)

Askel 3: On osoitettava, että mediaattori (M) on yhteydessä ennustettavaan muuttujaan (Y). Tämä osoitetaan kahden selittäjän regressioanalyysillä, jossa ennustetaan samanaikaisesti ennustettavaa muuttujaa sekä ennustavalla muuttujalla (X), että mediaattorilla (M). On tärkeää huomata, että tämän askeleen tarkastelemiseen ei riitä ainoastaan tutkia mediaattorin (M) ja ennustettavan (Y) muuttujan välistä yhteyttä esimerkiksi korrelaatiolla, sillä välittävien vaikutusten mallissa sekä mediaattorin, että ennustettavan muuttujan (Y) molempien vaihtelun ajatellaan selittyvän varsinaisella ennustajalla (X). Tämän takia tarkasteltaessa mediaattorin ja ennustettavan muuttujan välistä yhteyttä on ennustavan tekijän selittämä vaihtelu vakioitava.

Askel 4: Mikäli mediaattori (M) välittää täydellisesti ennustavan ja ennustettavan muuttujan välistä yhteyttä, pitää ennustavan ja ennustettavan muuttujan välisen yhteyden, joka on aikaisemmin havaittu askeleessa 1, muuttua tilastollisesti ei-merkitseväksi, kun tarkastellaan yhteyttä mallissa, jossa mediaattorin (M) aiheuttama vaihtelu on vakioitu. Mikäli mediaatioefekti on osittainen, pitäisi yhteyden olla pienempi askeleessa 1 havaittuun yhteyteen verrattuna. Tämän pienentymisen voi nähdä esimerkiksi p-arvon kasvusta, verrattaessa muuttujien välisiin yhteyksiin liittyvien tunnuslukujen p-arvoja. Tässä tehtävässä eroa ei tarvitse testata, vaan riittää tarkastella absoluuttista muutosta sen suhteen muuttuuko yhteys tilastollisesti vähemmän merkitseväksi. On syytä huomata, että askeleet 3 ja 4 voidaan tutkia samalla kahden selittävän tekijän regressioanalyysillä.

Tilastollisesti merkitsevällä yhteydellä tarkoitetaan tässä tehtävässä enintään 5 % riskitasoa, eli $p < 0,05$.

Seuraavilla sivuilla esitetyissä taulukoissa on esitetty regressioanalyysien tuloksia, joilla pyritään selvittämään, onko muuttujien välillä välittävää vaikutusta. Taulukoissa on esitetty viisi erillistä mediaatioanalyysiä (Taulukot A-E), joiden perusteella tehtävänä on päätellä, onko kyseessä välittävä vaikutus muuttujien välillä.

Lyhenteiden selitteet:

B = standardoimaton regressiokerroin

SE = standardoimattoman regressiokertoimen keskivirhe

t = regressiokertoimen hypoteesin testaukseen liittyvä testisuureen arvo

p = regressiokertoimen hypoteesin testaukseen liittyvä p-arvo

r = ennustavan muuttujan ja y-muuttujan välinen perinteinen Pearsonin tulomomentti korrelaatiokerroin. Tilastollisesti merkitsevät korrelaatiot on merkitty tähdillä seuraavasti:

$p < 0,05$ = *, $p < 0,01$ = ** ja $p < 0,001$ = ***

Tulostaulukko tehtävään 1.1.1.

Taulukko A

Regressiomallin selitys		B	SE	t	p	r
Ennustaja x, ennustettava m	Vakio	4,0152	0,2028	19,8037	0,0000	
	x	0,6421	0,0950	6,7617	0,0000	0,578***
		B	SE	t	p	r
Ennustajat x ja m	Vakio	3,0203	0,4513	6,6924	0,0000	
Ennustettava y	x	0,0240	0,1144	0,2096	0,8344	0,410***
	m	0,3264	0,1005	3,2468	0,0016	0,499***

Taulukko B

Regressiomallin selitys		B	SE	t	p	r
Ennustaja x, ennustettava m	Vakio	4,0635	0,2225	18,2640	0,0000	
	x	-0,0332	0,1216	-0,2727	0,7857	0,019
		B	SE	t	p	r
Ennustajat x ja m	Vakio	2,8057	0,3927	7,1449	0,0000	
Ennustettava y	x	0,4793	0,1023	4,6828	0,0000	0,260**
	m	0,0119	0,0850	0,1406	0,8885	-0,017

Taulukko C

Regressiomallin selitys		B	SE	t	p	r
Ennustaja x, ennustettava m	Vakio	3,6785	0,1934	19,0233	0,0000	
	x	0,5172	0,1067	4,8486	0,0000	0,643***
		B	SE	t	p	r
Ennustajat x ja m	Vakio	2,3677	0,3963	5,9742	0,0000	
Ennustettava y	x	0,0142	0,1124	0,1268	0,8994	0,169
	m	0,0832	0,0956	0,8701	0,3864	0,108

Taulukko D

Regressiomallin selitys		B	SE	t	p	r
Ennustaja x, ennustettava m	Vakio	4,0076	0,1903	21,0584	0,0000	
	x	0,6509	0,0925	7,0364	0,0000	0,506**
		B	SE	t	p	r
Ennustajat x ja m	Vakio	3,0195	0,4922	6,1344	0,0000	
Ennustettava y	x	0,2938	0,1249	2,3529	0,0206	0,549***
	m	0,4138	0,1112	3,7226	0,0003	0,668***

Taulukko E seuraavalla sivulla!

Taulukko E

Regressiomallin selitys		B	SE	t	p	r
Ennustaja x, ennustettava m	Vakio	4,1926	0,1689	24,8234	0,0000	
	x	0,7795	0,0836	9,3291	0,0000	0,514**
		B	SE	t	p	r
Ennustajat x ja m	Vakio	2,4264	0,5240	4,6307	0,0000	
Ennustettava y	x	-0,1148	0,1319	-0,8704	0,3862	0,382**
	m	0,7456	0,1161	6,4232	0,0000	0,629***

Tehtävä 1.1.1

Vastaa vastauslomakkeiseen 1. Vastauslomakkeessa osatehtävät ovat riveillä ja eri vastausvaihtoehdot sarakkeissa.

a) Valitse kaikki tulokset (A-E), joissa Baronin ja Kennyn ensimmäinen askel (Askel 1) toteutuu

b) Valitse kaikki tulokset (A-E), joissa Baronin ja Kennyn toinen askel (Askel 2) toteutuu

c) Valitse kaikki tulokset (A-E), joissa Baronin ja Kennyn kolmas askel (Askel 3) toteutuu

d) Valitse kaikki tulokset (A-E), joissa Baronin ja Kennyn neljäs askel (Askel 4) ja kaikki edelliset askeleet toteutuvat siten, että tulos viittaa täydelliseen mediaatioon

e) Valitse kaikki tulokset (A-E), joissa Baronin ja Kennyn neljäs askel (Askel 4) ja kaikki edelliset askeleet toteutuvat siten, että tulos viittaa osittaiseen mediaatioon

Tehtävä 1.1.2

Tutkija haluaa tutkia välittääkö työtyytyväisyys työstressin ja työsuorituksen välistä yhteyttä. Kaikkia kolmea muuttujaa on mitattu kyselylomakkeella, jota voidaan pitää luotettavana mittarina näiden ilmiöiden arvioimiseksi. Aineisto, joka on viidentoista (15) havainnon satunnaisotos tietystä populaatiosta, on esitetty taulukossa 1.1.2a. Analyysien kaikkien oletusten voi olettaa olevan voimassa.

Taulukko 1.1.2a: Aineisto

	Työstressi	Työtyytyväisyys	Työsuoriutuminen
1	5,00	6,00	6,00
2	5,00	7,00	7,00
3	4,00	5,00	3,00
4	3,00	2,00	,00
5	5,00	6,00	4,00
6	6,00	5,00	5,00
7	7,00	10,00	9,00
8	10,00	9,00	10,00
9	10,00	6,00	7,00
10	5,00	10,00	8,00
11	6,00	9,00	6,00
12	8,00	10,00	9,00
13	7,00	9,00	5,00
14	4,00	8,00	7,00
15	5,00	3,00	4,00
Keskiarvo	6,00	7,00	6,00

Tutkija oli saanut analyysit valmiiksi, mutta aivan viime metreillä hajamielinen professori tuhosi vahingossa taulukon ainoasta versiosta osan tuloksista. Lisäksi tilasto-ohjelman lisenssi umpeutui ja atk-tuki pystyy uusimaan lisenssin vasta kuukauden päästä. Puuttuvat tulokset joudutaan siis nyt kiireen takia laskemaan käsin. Taulukossa 1.1.2b on esitetty tulokset, jotka ovat säilyneet aikaisemmista analyyseistä.

Taulukossa 1.1.2b sarake B on standardoimaton regressiokerroin,

SE on regressiokertoimeen liittyvä keskivirhe,

t-arvo on regressiokertoimen hypoteesin testaukseen liittyvä testisuureen arvo ja viimeisessä sarakkeessa on p-arvo.

Taulukossa on esitetty kolmen eri regressioanalyysin tulokset ja tiettyyn regressioanalyysiin liittyvän osataulukon alapuolella on kerrottu ennustettava muuttuja.

Taulukosta puuttuvia tietoja ovat:

1) Baronin ja Kennyn askeleeseen 2 liittyvä regressiokerroin ja sen tilastolliseen merkitsevyyden testaukseen liittyvä testisuureen arvo ja p-arvo

2) Baronin ja Kennyn askeleisiin 3 ja 4 liittyvät regressiokertoimet ja niiden tilastolliseen merkitsevyyden testaukseen liittyvien testisuureiden arvot ja p-arvot

Taulukko 1.1.2b: Puutteelliset tulokset mediaatioefektin testauksesta

Ennustajat	B	SE	t-arvo	p-arvo
Vakio	0,900	1,643	0,548	0,593
Työstressi	0,850	0,260	3,272	0,006

Ennustettava muuttuja: työsuoriutuminen

Ennustajat	B	SE	t-arvo	p-arvo
Vakio	??	1,916	??	??
Työstressi	??	0,303	??	??

Ennustettava muuttuja: Työtyytyväisyys

Ennustajat	B	SE	t-arvo	p-arvo
Vakio	??	1,185	??	??
Työstressi	??	0,190	??	??
Työtyytyväisyys	??	0,153	??	??

Ennustettava muuttuja: Työsuoriutuminen

Tehtävä 1.1.2

Seuraavista kaavoista voi olla apua tehtävän 1.1.2 tekemisessä. Kaavat on esitetty tilastomateriaalissa, mutta ensimmäisessä versiossa oli pieni virhe, joka korjattiin 12.4.2019 korjauspäivityksessä. Kaavat eivät kuitenkaan ole tehtävän ratkaisemiseksi välttämättömiä.

$$\sum(x_1y) = \sum(X_1 - \bar{X}_1)(Y - \bar{Y}) = \sum X_1Y - \frac{\sum X_1 \sum Y}{N}$$

$$\sum(x_2y) = \sum(X_2 - \bar{X}_2)(Y - \bar{Y}) = \sum X_2Y - \frac{\sum X_2 \sum Y}{N}$$

$$\sum(x_1x_2) = \sum(X_1 - \bar{X}_1)(X_2 - \bar{X}_2) = \sum X_1X_2 - \frac{\sum X_1 \sum X_2}{N}$$

Kirjoita vastauslomakkeisiin 1 ja 2

a) Baronin ja Kennyn askeleeseen 2 liittyvän regressiokertoimen laskemiseen tarvittavan kaavan numero sivun 35 kaavakokoelmasta

b) regressiokertoimen arvo

c) regressiokertoimeen liittyvä testisuureen arvo

d) regressiokertoimen testaukseen liittyvä vapausasteluku

e) valitse vastauslomakkeen vaihtoehdoista regressiokertoimen tilastolliseen testaukseen liittyvä p-arvo

- vastausvaihtoehdot: i) $p < 0{,}001$ ii) $p < 0{,}01$ iii) $p < 0{,}05$ iv) $p > 0{,}05$

f) Baronin ja Kennyn askeleisiin 3 ja 4 liittyvän regressiokertoimen b_1 laskemiseen tarvittavan kaavan numero sivun 35 kaavakokoelmasta, kun työstressi on määritelty muuttujaksi x_1

g) Tehtävän f kaavalla lasketun regressiokertoimen arvo

Tehtävät jatkuvat seuraavalla sivulla!

h) regressiokertoimen testaukseen liittyvä testisuureen arvo

i) regressiokertoimen testaukseen liittyvä vapausasteluku

j) valitse vastauslomakkeen vaihtoehdoista regressiokertoimen tilastolliseen testaukseen liittyvä p-arvo

- vastausvaihtoehdot: i) $p < 0{,}001$ ii) $p < 0{,}01$ iii) $p < 0{,}05$ iv) $p > 0{,}05$

k) Baronin ja Kennyn askeleisiin 3 ja 4 liittyvän regressiokertoimen b_2 laskemiseen tarvittavan kaavan numero sivun 35 kaavakokoelmasta, kun työtyytyväisyys on määritelty muuttujaksi x_2

l) tehtävän k kaavalla lasketun regressiokertoimen arvo

m) regressiokertoimen testaukseen liittyvä testisuureen arvo

n) regressiokertoimen testaukseen liittyvä vapausasteluku

o) valitse vastauslomakkeen vaihtoehdoista regressiokertoimen tilastolliseen testaukseen liittyvä p-arvo

- vastausvaihtoehdot: i) $p < 0{,}001$ ii) $p < 0{,}01$ iii) $p < 0{,}05$ iv) $p > 0{,}05$

p) valitse onko tulosten perusteella muuttujien välillä:

i. täydellinen mediaatio
ii. osittainen mediaatio
iii. ei havaittavissa mediaatioefektiä.

Tehtävä 1.2

Tässä tehtävässä on vastattava kysymyksiin valitsemalla oikea vaihtoehto vastausvaihtoehtojen joukosta. Osassa tehtävistä saattaa olla useampia oikeita vastausvaihtoehtoja. Näissä tehtävissä täydet pisteet saadakseen täytyy löytää kaikki oikeat vaihtoehdot. Vastaa vastauslomakkeeseen 2. Vastauslomakkeessa osatehtävät ovat riveillä ja eri vastausvaihtoehdot sarakkeissa.

1. p-arvo on… (vastausvaihtoehdoissa sanan ”aineisto” voi käsittää laajasti, joko koko aineistona tai aineistosta laskettuna tunnuslukuna)

Vastausvaihtoehdot:

a) Todennäköisyys, että vastahypoteesi on tosi, sillä ehdolla, että on havaittu tietty aineisto tai tätä poikkeavampi tulos.

b) 1 - todennäköisyys, että nollahypoteesin on tosi, sillä ehdolla, että on havaittu tietty aineisto tai tätä poikkeavampi tulos

c) Todennäköisyys havaita tietty aineisto tai tätä poikkeavampi tulos, sillä ehdolla, että nollahypoteesi on tosi.

d) 1 - todennäköisyys havaita tietty aineisto tai tätä poikkeavampi tulos, sillä ehdolla, että nollahypoteesi on tosi

2. Spearmanin järjestyskorrelaatiokertoimen yhteydessä pätee aina…

Vastausvaihtoehdot:

a) Spearmanin järjestyskorrelaatiokerroin on aina itseisarvoltaan Pearsonin tulomomenttikerrointa suurempi, jos tutkittavien muuttujien välinen yhteys on monotoninen, eli aidosti laskeva tai nouseva, mutta epälineaarinen.

b) Spearmanin järjestyskorrelaatiokerroin on aina Pearsonin tulomomenttikerrointa suurempi, mikäli yhteys on epälineaarinen, mutta ei välttämättä monotoninen.

c) Mikäli muuttujat saavat arvoikseen järjestysnumeroita, on Spearmanin järjestyskorrelaatiokerroin aina yhtä suuri kuin vastaavien muuttujien Pearsonin tulomomenttikerroin.

d) Spearmanin järjestyskorrelaatiokertoimen avulla voidaan tutkia vain lineaarisia yhteyksiä.

3. Älykkyysosamäärän voidaan ajatella noudattavan normaalijakaumaa odotusarvolla (keskiarvolla) 100 ja keskihajonnalla 15. Mikä on todennäköisyys, että satunnaisesti valitun henkilön älykkyysosamäärä on pienempi tai yhtä suuri kuin 130, mutta suurempi tai yhtä suuri kuin 110?

Vastausvaihtoehdot:

a) likimäärin 0,23 eli 23 %

b) likimäärin 0,48 eli 48 %

c) likimäärin 0,75 eli 75 %

d) likimäärin 0,98 eli 98 %

4. Tiedetään, että standardoidun normaalijakauman ja khii-toiseen jakauman välillä vallitsee yhteys: Mikäli jokin satunnaisluku noudattaa standardoitua normaalijakaumaa, niin tämän luvun toinen potenssi noudattaa khii-toiseen jakaumaa vapausasteilla 1. Käänteisesti tiedetään, että mikäli luku noudattaa khii-toiseen jakaumaan vapausasteilla 1, niin tämän luvun neliöjuuri noudattaa standardoitua normaalijakaumaa. Eräs luku noudattaa khii-toiseen jakaumaa vapausasteilla 1 siten, että sitä pienempien tai yhtä suurten lukujen todennäköisyys on 90 %. Tästä luvusta otetaan neliöjuuri. Mikä on todennäköisyys havaita tämän muunnoksen jälkeen muunnettu luku tai sitä pienempi arvo?

Vastausvaihtoehdot:

a) 0,9000

b) 0,9500

c) 0,1000

d) 0,9965

5. Erästä psykologista testiä käytetään tietyn mielenterveyshäiriön diagnosoimiseen. Testin tulosten perusteella ne henkilöt, jotka ovat oikeasti sairaita, luokitellaan terveiksi 10 %:n todennäköisyydellä ja ne henkilöt, jotka ovat oikeasti terveitä, luokitellaan sairaiksi 6 %:n todennäköisyydellä. Kaikista tutkittavista 80 % on sairaita. Mikä on todennäköisyys, että sairaaksi luokiteltu henkilö on oikeasti sairas?

Vastausvaihtoehdot:

a) likimäärin 0,79 eli 79 %

b) likimäärin 0,80 eli 80 %

c) likimäärin 0,92 eli 92 %

d) likimäärin 0,98 eli 98 %

6. Eräässä tutkimuksessa tutkittiin miesten ja naisten eroja työtyytyväisyydessä. Tulosten analyysimenetelmänä käytetiin riippumattomien otosten t-testiä (vastahypoteesi: $\mu_{miehet} \neq \mu_{naiset}$). Havaituksi testisuureeksi saatiin 2,05 ja tutkimukseen osallistui 15 naista ja 12 miestä. Testituloksen perusteella p-arvolle pätee:

Vastausvaihtoehdot:

a) $p < 0,001$

b) $p < 0,01$

c) $p < 0,05$

d) $p > 0,05$

7. Todennäköisyyslaskennassa erillisille ja riippumattomille tapahtumille pätee, että…

Vastausvaihtoehdot:

a) tapahtumat ovat aina sekä erillisiä, että riippumattomia.

b) tapahtumat eivät voi olla sekä erillisiä, että riippumattomia.

c) ei-erilliset tapahtumat ovat aina myös toisistaan riippuvaisia.

d) ei-erilliset tapahtumat voivat olla joko toisistaan riippumattomia tai riippuvaisia.

8. Tutkija tutkii kolmea mielenterveyshäiriötä A, B ja C. Hän on kerännyt näitä sairauksia sairastavien joukosta 100 hengen aineiston ja hän haluaa testata, onko mielenterveyshäiriöiden ja sukupuolen välillä riippuvuutta. Testinä hän käyttää khii-toiseen riippumattomuustestiä. Mikä on tutkijan käyttämän testin vapausasteluku?

Vastausvaihtoehdot:
a) 2

b) 6

c) 98

d) 100

9. Muuttujan keskiarvo on 10 ja keskihajonta 5. Jos jokaisesta muuttujan arvosta vähennetään arvo 5, mitkä seuraavista väitteistä pitävät paikkaansa muunnoksen jälkeen?

Vastausvaihtoehdot:
a) keskiarvo on 5 ja keskihajonta 2,5

b) keskiarvo on 10 ja keskihajonta 5

c) keskiarvo on 5 ja keskihajonta on 5

d) muuttujan keskiarvon ja keskihajonnan arvosta ei voida tehdä päätelmiä ilman varsinaista aineistoa

10. Tarkastellaan kahta muuttujaa, joista ensimmäisessä keskiarvo on 10 ja keskihajonta 5 ja toisessa keskiarvo on 7 ja keskihajonta 2. Molempien muuttujien kohdalla muuttujien jokaisesta havaintoarvosta vähennetään arvo 3. Mitkä seuraavista väitteistä pitävät paikkaansa muunnoksen jälkeen?

Vastausvaihtoehdot:
a) Muuttujien välinen Pearsonin tulomomenttikorrelaatiokerroin muuttuu ja Spearmanin järjestyskorrelaatiokerroin muuttuu muunnoksen seurauksena.

b) Muuttujien välinen Pearsonin tulomomenttikorrelaatiokerroin ei muutu, mutta Spearmanin järjestyskorrelaatiokerroin muuttuu muunnoksen seurauksena.

c) Muuttujien välinen Pearsonin tulomomenttikorrelaatiokerroin ei muutu, eikä Spearmanin järjestyskorrelaatiokerroin muutu muunnoksen seurauksena.

d) Muuttujien välisessä regressioanalyysissä vakiotermin arvo muuttuu, mutta standardoimattoman regressiokertoimen arvo ei muutu.

Tehtävä 2

Vastaa ennakkomateriaalin ja kokeessa jaetun aineiston perusteella. Vastaa tehtäviin 2.1 ja 2.2 vastauslomakkeeseen 2 osatehtävien ohjeiden mukaisesti. Rasti valitsemiasi vastausvaihtoehtoja vastaavat soikiot vastauslomakkeeseen 2. Ellet ole vastannut osatehtävään mitään, tulkitaan vastaus vääräksi. Vastauslomakkeessa osatehtävät ovat riveillä ja eri vastausvaihtoehdot sarakkeissa.

Tehtävä 2.1

Vastaa ennakkomateriaalin perusteella seuraaviin väitteisiin / kysymyksiin. Vastausvaihtoehdot ovat a, b, c ja d. Väitteiden / kysymysten kohdalla voi olla useampi kuin yksi vastausvaihtoehto oikein, mutta jokaisessa kohdassa on ainakin yksi vastausvaihtoehto oikein. Jokaisen väitteen / kysymyksen kohdalla on löydettävä kaikki ja vain kaikki oikeat vastausvaihtoehdot, jotta saisi täydet pisteet.

1. Pyykösen ym:n Alzheimerin tautia ja ajokykyä käsittelevässä tutkimuksessa...

a) kaikkien osallistujien sisäänottokriteereitä olivat alle 75 vuoden ikä ja voimassa oleva ajokortti, kun taas poissulkukriteereitä kaikille osallistujille olivat aikaisempi toimintakykyä selvästi heikentävä neurologinen tai psykiatrinen sairaus ja väsymykseen vaikuttavan lääkkeen käyttö.
b) kaikkien osallistujien sisäänottokriteereitä olivat kyky ymmärtää tutkimuksen tarkoitus, voimassa oleva ajokortti, kyky ajaa autoa ilman ylimääräisiä hallintalaitteita ja äidinkielenä suomi.
c) kaikkien osallistujien sisäänottokriteereitä olivat kyky ajaa autoa ilman ylimääräisiä hallintalaitteita ja äidinkielenä suomi, kun taas Alzheimerin tautia sairastavilla yksi sisäänottokriteereistä oli, että oireiden alkamisesta oli kulunut alle vuosi.
d) poissulkukriteereitä kaikille osallistujille olivat kaikki näköongelmat ja aikaisempi toimintakykyä selvästi heikentävä neurologinen tai psykiatrinen sairaus, kun taas sisäänottokriteereitä kaikille osallistujille olivat alle 75 vuoden ikä ja suomi äidinkielenä.

2. Pyykösen ym:n Alzheimerin tautia ja ajokykyä käsittelevässä tutkimuksessa terveet verrokit keskimäärin (kun tulosta pidetään tilastollisesti merkitsevänä, jos $p < 0,05$)...

a) arvioivat oman ajosuoritteensa tilastollisesti merkitsevästi heikommaksi, tekivät tilastollisesti ei-merkitsevästi enemmän ajoneuvon sijaintivirheitä ja kokonaisuudessaan tilastollisesti merkitsevästi vähemmän virheitä simulaattoriajossa.
b) tekivät tilastollisesti ei-merkitsevästi vähemmän ajoneuvon sijaintivirheitä ja kokonaisuudessaan tilastollisesti ei-merkitsevästi vähemmän virheitä simulaattoriajossa kuin Alzheimerin tautia sairastavat potilaat.
c) olivat ennen simulaattoriajoa tilastollisesti merkitsevästi ja sen jälkeen tilastollisesti ei-merkitsevästi väsyneempiä ja arvioivat oman ajosuoritteensa tilastollisesti merkitsevästi heikommaksi kuin Alzheimerin tautia sairastavat potilaat.
d) ajoivat tilastollisesti ei-merkitsevästi nopeammin sekä olivat ennen simulaattoriajoa ja sen jälkeen tilastollisesti ei-merkitsevästi väsyneempiä kuin Alzheimerin tautia sairastavat potilaat.

3. MoCA-menetelmästä pätee Pyykösen ym:n Alzheimerin tautia ja ajokykyä käsittelevän artikkelin perusteella, että...

a) sen katkaisurajaa on väärien positiivisten välttämiseksi ehdotettu laskettavaksi, alle 20:n pistemäärä viittaa ajokiellon tarpeeseen ja että terveet verrokit saivat tässä tutkimuksessa keskimäärin enemmän kokonaispisteitä kuin Alzheimerin tautia sairastavat potilaat.
b) alle 20:n pistemäärä viittaa erikoissairaanhoidon tutkimusten harkinnan tarpeeseen, tämän tutkimuksen tulokset puoltavat ehdotusta katkaisurajan laskemiseksi 23:een ja että Alzheimerin tautia sairastavat potilaat saivat tässä tutkimuksessa tilastollisesti merkitsevästi (5 prosentin riskitasolla) keskimäärin vähemmän kokonaispisteitä kuin verrokit.
c) kukaan Alzheimerin tautia sairastavista potilaista ei saanut tässä tutkimuksessa yhtä paljon kokonaispisteitä kuin vähiten kokonaispisteitä saanut terve verrokki, katkaisurajaa on sairauksien seulonnassa väärien positiivisten välttämiseksi ehdotettu siirrettäväksi 26:sta 23:een ja että alle 20:n pistemäärä viittaa tiiviin seurannan tarpeeseen.
d) katkaisurajaa on väärien tulosten välttämiseksi ehdotettu muutettavaksi, alle 20:n pistemäärä viittaa ajokokeen harkinnan tarpeeseen ja että Alzheimerin tautia sairastavat potilaat saivat tässä tutkimuksessa keskimäärin tilastollisesti merkitsevästi (5 prosentin riskitasolla) vähemmän pisteitä käsitteenmuodostuksessa ja orientaatiossa kuin terveet verrokit.

4. Silvonen viittaa psykologian historiaa koskevassa artikkelissaan Jääskeläisen esittelemiin psykologian historian tutkimuksen viiteen näkökulmaan. Yksi näistä tarkastelee eri vaiheissa muodostuneita koulukuntia ja niiden esittämiä ajattelumalleja. Tästä näkökulmasta...

a) psykologian vakiintunut historiankirjoitus on ollut presentististä.
b) psykologia ei ole yhtenäinen kokonaisuus.
c) historiallinen tutkimus on osa tieteen itsereflektiota.
d) yhteiskunnallisten aatteellisten jännitteiden jäsentäminen voi selventää psykologian eri paradigmojen asemaa tieteen kentällä.

5. Idealistisen psykologian dualistinen ihmiskuva...

a) rajoitti varsinaisen sielunelämän kokeellisen tutkimuksen apperseeraamisen ilmiöön, jolla kuvattiin sielunelämän aktiivista puolta.
b) vastasi metodologisella tasolla erottelua ymmärtämiseen hengentieteiden metodina ja selittämiseen luonnontieteiden metodina.
c) myötävaikutti siihen, että varhaisvaiheen suomalainen psykologia löysi kosketuspinnan yhteiskunnallisiin käytäntöihin vain nousevan kansallisvaltion koululaitoksessa.
d) johti tutkimussuuntaan, jossa oleellisena ihmisessä pidettiin aistitoimintoja ja muistia.

6. Silvonen esittää psykologian kehityksen kolmiomallin yhteydessä, että...

a) psykologia voi olla perustellusti aidosti luonnontieteellisen mallin mukaan orientoituvaa tutkimusta, humanistisesta traditiosta ammentavaa tulkintataitoa ja ihmismielen ja yhteiskunnallisten rakenteiden yhteen kietoutumisen tutkimusta.
b) Foucault-vaikutteinen uuden historian teesi on, että sosiaalitieteiden ja sosiaalisen hallinnan näkökulman tulisi olla psykologian historiassa ensisijainen, sillä psykologia on joka tapauksessa osallisena näiden käytänteiden kentillä.
c) parhaimmillaan psykologian historia tarjoaisi reflektiivisen kuvan psykologian käytänteiden keskinäisistä yhteyksistä ja vaikutuksista toisiinsa.
d) mallin avulla psykologia voisi tavoitella yhtenäisen tieteen asemaa.

7. Keiski ja kumppanit tarkastelivat aikaisempia tutkimuksia perheväkivallasta. Mikä tai mitkä seuraavista väittämistä pitävät heidän katsauksensa perusteella paikkansa?

a) Perheissä väkivaltaa puolisoa tai lapsia kohtaan käyttävät enimmäkseen miehet.
b) Naiset käyttävät väkivaltaa esimerkiksi itsepuolustukseksi sekä lapsen ja puolison rankaisemiseksi.
c) Lapsuuden väkivaltakokemusten ja aikuisuuden väkivaltakäyttäytymisen suora yhteys on voimakas.
d) Perheväkivaltaa käyttävillä naisilla on useammin itseen kohdistuvaa väkivaltaa kuin perheväkivaltaa käyttävillä miehillä.

8. Mikä tai mitkä seuraavista väittämistä pitävät Keiskin ja kumppaneiden mukaan paikkansa naisille suunnatuista perheväkivaltakäyttäytymiseen liittyvistä interventioista?

a) Interventiot pohjautuvat feministiseen ajatteluun.
b) Interventioiden ansiosta rikoksen uusiutuminen on vähentynyt, mutta niillä ei ole ollut vaikutusta parisuhdetyytyväisyyteen.
c) Niiden keskeyttämisprosentti on suuri.
d) Interventioiden ansiosta osallistujien itsetunto on kohentunut ja emotionaalinen väkivaltakäyttäytyminen on vähentynyt.

9. Perheväkivaltaa käyttäneiden naisten ryhmäinterventiota koskeneista tuloksistaan tutkijat päättelivät, että...

a) osallistujien itsetuntemus koheni intervention aikana.
b) pienikin itsetuntemuksen koheneminen on hyväksi, sillä se vähentää väkivaltakäyttäytymistä.
c) yhteiskunnallista perheväkivaltakeskustelua on laajennettava siitä tausta-ajatuksesta, että tekijä on mies.
d) itseään vahingoittavan naisen kohdalla tulee huomioida myös toisiin kohdistuvan väkivaltakäyttäytymisen mahdollisuus.

10. Perheväkivaltaa käyttäneiden naisten ryhmäinterventiota tarkastelleessa tutkimuksessa tekijät arvioivat tulosten tulkintaan vaikuttaviksi heikkouksiksi...

a) itsearviointiin liittyvän kaunistelun, joka saattoi jättää piiloon perheväkivaltaan liittyviä sensitiivisiä aiheita.
b) kontrolliryhmän puuttumisen, minkä vuoksi vaikuttavuutta ei voinut arvioida.
c) vastaajien vähäisen kokonaismäärän, jonka takia aineistossa olevia todellisia yhteyksiä ei välttämättä saatu esiin.
d) matalan vastausprosentin, jonka vuoksi tuloksia ei voitu yleistää.

11. Kun interventiolla pyritään vaikuttamaan käyttäytymiseen, jossa motivaation laadulla (ulkoinen/sisäinen) tiedetään olevan merkitystä käyttäytymisen kannalta, on tarkoituksenmukaisinta soveltaa terveyskäyttäytymisen teoriana...

a) suojelumotivaatioteoriaa.
b) sosiaalis-kognitiivista teoriaa.
c) itsemääräämisteoriaa.
d) suunnitellun käyttäytymisen teoriaa.

12. Teoreettisten aihealueiden viitekehys (TDF) auttaa jäsentämään...

a) käyttäytymismuutostekniikoita.
b) interventioteorian valintaa.
c) keskitason teorioita.
d) kompleksisia adaptiivisia systeemeitä.

13. Käyttäytymismuutostekniikat määritellään siten, että ne...

a) sisältävät tekemisen substantiiveja (esimerkiksi tarjoaminen, neuvominen, järjestäminen, kehottaminen).
b) viittaavat tekniikan toteuttajan/toteuttajien toimiin.
c) voidaan toteuttaa intervention ”tekijän” toimesta tai itsenäisesti.
d) sisältävät termin ”käyttäytyminen”, joka viittaa yksittäiseen tekoon, toimintaan tai tapahtumaketjuun.

14. Koivusalon ym. artikkelin perusteella tiedetään, että…

a) pojat olivat yliedustettuina urheilijapainotteisen ja selkiintymättömän identiteetin ryhmissä.
b) tyttöjä on poikia vähemmän urheilun ja opiskelun yhdistävässä vahvan urheilija- ja opiskelijaidentiteetin ryhmässä.
c) urheilijan ja opiskelijan rooleihin yhtäaikainen sitoutuminen onnistuu tytöiltä poikia paremmin.
d) tyttöjen identiteetin on havaittu kehittyvän varhaisnuoruudessa nopeammin kuin poikien ja kehitys on myös monimuotoisempaa.

15. Koivusalon ym. artikkelin perusteella voidaan sanoa, että voimakas urheilijaidentiteetti ja epätasapaino opiskelun ja urheilun välillä on yhdistetty...

a) masentuneisuuteen ja heikompaan itsetuntoon.
b) sopeutumisvaikeuksiin urheilu-uran päättyessä.
c) sopeutumisvaikeuksiin työelämään siirryttäessä.
d) tunne-elämän ongelmiin urheilu-uran päättyessä.

16. Koivusalon ym. mukaan aikaisempien nuorten opiskelija- ja urheilijaidentiteettiä koskevien tutkimusten haasteita ovat:

a) aiemmissa tutkimuksissa ei ole esimerkiksi selvitetty sitä, kuinka suuri osa kaksoisuraa suorittavista nuorista sitoutuu molempiin kaksoisuran rooleihin.
b) opiskelija- ja urheilijaidentiteetin välistä yhteyttä on aiemmin tutkittu pääasiassa laadullisin menetelmin.
c) tutkimuksissa on usein keskitytty vain tietyn lajin urheilijoihin ja lisäksi tutkimukset on toteutettu pääosin Yhdysvalloissa.
d) suurin osa tutkimuksista on keskittynyt yliopisto-opintojen ja urheilun yhdistämiseen.

17. Koivusalon ym. artikkelin mukaan hyvinvointia tukevan tasapainon saavuttaminen opiskelun ja urheilun välillä vaatii...

a) yhteiskunnan vastuuta tarjota sellaiset puitteet, joissa tasapainoinen ja moniulotteinen identiteetin kehitys on mahdollista nuorille, jotka tähtäävät huipulle urheilussa.
b) urheiluun ja opintoihin liittyvän urakehityksen tasapainottamista mahdollisimman varhaisessa vaiheessa.
c) erityisesti huipputason urheilussa saavuttaville mahdollisuutta keskittyä ensisijaisesti urheilijaidentiteetin rakentamiseen.
d) että selkiintymättömän identiteetin ryhmään kuuluville tarjotaan opiskelumotivaatiota ja -kykyä tukevia interventioita.

18. Pohjolan ym. mukaan positiointiteorian kohdalla pätee, että...

a) se on osa tiedonsosiologista suuntausta.
b) se on kiinnostunut puhekäytännöistä ja siitä, miten niiden kautta tuotetaan ihmisten subjektiviteettia positioimalla eli asemoimalla ihmisiä.
c) sen mukaan sosiaalipsykologian ensisijainen tehtävä on ymmärtää merkitysten rakentumista ihmisten välisessä vuorovaikutuksessa.
d) sen mukaan keskusteluissa luodaan paikallisia moraalisia järjestyksiä, jotka määrittävät sen, mitä oikeuksia ja velvollisuuksia kullakin on.

19. Positiointiteorian käsitteiden avulla on Pohjolan ym. artikkelin perusteella aiemmin tutkittu muun muassa...

a) konfliktien muodostumista.
b) organisaatiokonsulttien käyttämiä syntaktisia rooleja.
c) organisaatioiden muutosviestintää.
d) organisaatiomuutokseen liittyvien vaikeuksien vuorovaikutuksellista käsittelyä.

20. Pohjolan ym. artikkelin mukaan ohjattavien osallisuutta työyhteisön kehittämiseen voidaan edistää houkuttelemalla heitä ”omistajuuspuheeseen”. Omistajuuspuheella tarkoitetaan...

a) puhetapaa, jossa esimiesasemassa olevat puhuvat heille itselleen tärkeistä asioista alaisilleen huomioiden alaisten tarpeet ja toiveet.
b) puhetapaa, jossa esimiesasemassa olevat puhuvat alaisia koskevista tärkeistä asioista huomioiden alaisten tarpeet ja toiveet.
c) puhetapaa, jossa työyhteisöä kehittävät konsultit huomioivat interventioissaan työyhteisön jäsenten erilaiset roolit ja odotukset.
d) puhetapaa, jossa työyhteisön jäsenet puhuvat heille tärkeistä asioista omista näkökulmistaan käsin.

21. Autismin kirjon henkilöillä on todettu esiintyvän...

a) joissain tapauksissa erityisen vahva musiikillinen muisti sekä kyky erotella ja myös nimetä sävelkorkeuksia riippumatta aiemmasta musiikillisesta koulutuksesta.
b) taipumusta keskittyä yksityiskohtiin ja puolestaan heikkoutta kyvyssä havaita kokonaisuuksia.
c) hitaampaa suoriutumista prosessointinopeutta vaativissa tehtävissä verrattuna kontrolleihin, mikä vaikuttaa heidän keskimääräiseen suoriutumiseensa yleistä älykkyyttä arvioivissa tehtävissä.
d) absoluuttinen sävelkorva, eli kyky nimetä sävelkorkeuksia ilman kontekstia, useammin kuin verrokeilla.

22. Savant-ilmiön esiintyminen autismikirjon häiriön yhteydessä...

a) on yleistä: noin 90 prosentilla savant-henkilöistä on autismikirjon häiriö.
b) on liitetty eräissä selitysmalleissa puutteisiin sosiaalisessa toiminnassa niin, että sosiaalisuudesta vastaavat aivokuoren alueet olisivat savanteilla siirtyneet muun kuin sosiaalisen tiedon käsittelyyn.
c) on liitetty eräissä selitysmalleissa siihen, että henkilöillä joilla on sekä autismia, että savant-kykyjä, on lisääntynyt kiinnostus kieleen ja sen prosessointiin, joka voi johtaa tarkkaan auditoriseen erotteluun esimerkiksi musiikin havaitsemisessa.
d) on liitetty eräissä selitysmalleissa siihen, että suppeisiin kiinnostuksen kohteisiin liittyvät yliharjoitellut kyvyt vievät tilaa korkeamman tason kognitiiviselta prosessoinnilta.

23. Uupumisasteinen väsymys…

a) kuvaa työntekijän emotionaalisten ja fyysisten voimavarojen tyhjentymistä niin, että tila ei poistu levolla tai loma-aikana ja tällöin työuupumus on jo edennyt pitkälle.
b) heikentää neljän viikon seurantaan perustuvan tutkimuksen mukaan työntekijän kykyä irrottautua työstä.
c) on eräiden pitkittäis- ja päiväkirjatutkimusten mukaan negatiivisesti yhteydessä psykologiseen irrottautumiseen työstä vapaa-ajalla.
d) jaetaan Maslachin työuupumuksen arviointimenetelmällä arvioituna sopimuksenvaraisesti kolmeen luokkaan sen perusteella, kuinka voimakkaita oireita tutkittava raportoi.

24. Työasioiden vatvomisesta työajan ulkopuolella tiedetään, että…

a) Kinnusen ym. tutkimuksessa havaittiin, että kahden vuoden seurannan aikana ongelmasuuntautunut, mutta ei tunnepitoinen vatvominen lisääntyi ei-uupuneiden ja lievän uupumuksen ryhmissä, kun taas molemmat vähenivät vähentyneen uupumuksen ryhmässä.
b) Kinnusen ym. tutkimuksessa havaittiin, että kahden vuoden seurannan aikana sekä ongelmasuuntautunut että tunnepitoinen vatvominen lisääntyivät ei-uupuneiden ja lievän uupumuksen ryhmissä, mutta ainoastaan tunnepitoisen vatvomisen lisääntymiseen liittyi uupumusasteisen väsymyksen lisääntyminen.
c) tunnepitoinen vatvominen on tutkimusten mukaan yhteydessä kortisolin eritykseen aamulla ja illalla, mikä puolestaan on negatiivisessa yhteydessä työntekijän palautumisen kokemuksiin.
d) tunnepitoinen vatvominen keskittyy työn negatiivisiin asioihin ja sen on todettu olevan yhteydessä krooniseen ja akuuttiin työhön liittyvään väsymykseen, kun taas ongelmasuuntautunut vatvominen on ratkaisukeskeistä ja sen on todettu olevan yhteydessä työntekijöiden vähäisempään väsymykseen.

25. Astrosyyteistä pitää paikkansa, että…

a) niistä vapautuvalla D-seriini-nimisellä välittäjäaineella on keskeinen rooli jo syntyneiden uusien solujen erilaistumisessa ja selviytymisessä.
b) ne ovat verisuoniin yhteydessä olevia hermotukisoluja, jotka sijaitessaan jyväissolukerroksen subgranulaarisella alueella voivat aikuisiässä edistää solujen jakautumista, erilaistumista ja kypsymistä.
c) edistävät neurogeneesiä vahvemmin aikuisuudessa kuin vastasyntyneellä.
d) ne ovat hermotukisoluja, samoin kuin solukuolemaan ohjattuja soluja poistavat mikrogliasolutkin.

26. Hippokampuksen mikroympäristössä neurogeneesiä säätelevät…

a) välittäjäaineet, kuten GABA ja glutamaatti, joiden syötteet syntyvät hippokampuksen kypsien solujen aktivaatiosta.
b) kasvutekijät, kuten aivoperäinen hermokasvutekijä BDNF (brain-derived neurotrophic factor), jonka määrään stressin on osoitettu vaikuttavan lisäävästi.
c) hormonit, kuten testosteroni ja estrogeeni, jotka edistävät neurogeneesiä lisäämällä hermosolujen tuoja- ja viejähaarakkeiden kasvua sekä synapsien muodostumista.
d) hormonit, kuten kortikosteroidi (ihmisellä kortisoli), jonka lyhytaikaisen, kertaluonteisen altistuksen on todettu aikuisilla uros- ja naarasrotilla vähentävän uusien solujen erilaistumista jyväissoluiksi.

27. Oppimisen hermostollisiin mekanismeihin on esitetty kuuluvan...

a) kestotehostuminen, joka tarkoittaa aiemmin muodostuneiden synaptisten yhteyksien vahvistumista ja jolle kypsät jyväissolut ovat otollisia, koska niillä on epäkypsiä soluja korkeampi lepojännite.
b) liian vahvojen olemassa olevien synaptisten yhteyksien vaimentaminen, jonka on esitetty liittyvän siihen, että uudet jyväissolut vaimentavat kypsiä jyväissoluja.
c) tarpeettomien olemassa olevien synaptisten yhteyksien vaimentaminen, jonka on esitetty liittyvän hippokampuksen ja aivokuoren välisen kommunikaation säätelyyn, erityisesti synaptisten yhteyksien hillitsemiseen.
d) dentate-piikki, joka on tyypillisesti unenaikainen, hiluksesta mitatussa kenttäpotentiaalissa tapahtuva hetkellinen positiivinen heilahdus, jonka seurauksena hippokampuksesta aivokuorelle kulkeva yhteys suljetaan.

28. Savant-ilmiötä on selitetty…

a) Andersonin älykkyysmallilla (theory of minimal cognitive architecture), jonka mukaan älykkyys koostuu sekä moduuleista että yleisestä älykkyydestä (g) ja on esitetty, että Savant-syndroomassa moduulien kyvyt ovat säilyneet.
b) Demetrioun mallilla, jossa älykkyys nähdään hierarkkisena systeeminä, jonka ensimmäisellä tasolla ovat erikoistuneet kognitiiviset järjestelmät ja joka sisältää kolmannella, toimintaa säätelevällä tasollaan myös metakognition.
c) aivojen modulaarisuudella, jonka mukaan savant-taidot ovat aivoissa alueellisesti erikoistuneita ja yleisestä älykkyydestä riippumattomia.
d) koherenssiteorian, eksekutiivisen teorian ja empatointi-systemointi -teorian avulla.

29. Uupumusryhmien taustatekijöistä pitää Kinnusen ym. tutkimuksen tulosten perusteella paikkaansa, että…

a) lisääntyneen uupumuksen ryhmässä oli naisia enemmän ja miehiä vähemmän verrattuna koko aineiston sukupuolijakaumaan, kun taas ei-uupuneiden ryhmässä naisia oli otoksen jakaumaan verrattuna vähemmän ja miehiä enemmän.
b) uupumusryhmät eivät eronneet toisistaan koulutustasossa tilastollisesti merkitsevästi.
c) lähtötilanteessa T1 uupumusryhmät erosivat toisistaan niin, että ei-uupuneiden ryhmässä tutkittavat tekivät vähiten työtunteja viikossa, ja tämä oli tilastollisesti merkitsevästi vähemmän kuin ryhmässä, jossa oireet olivat vakavia.
d) ei-uupuneiden ryhmässä työtunnit pysyivät samalla tasolla koko ajanjakson ajan, mutta vähentyvien oireiden ryhmässä työtunnit vähenivät tilastollisesti merkitsevästi.

30. Neurogeneesin muokkaamisesta pitää paikkaansa että…

a) kun hiirten neurogeneesiä on lisätty oppimistapahtuman jälkeisellä aerobisella harjoittelulla ja vähennetty geenitekniikan ja lääkeaineiden avulla, on pystytty säätelemään sitä, kuinka hyvin hiiret muistavat aiemmin opitun.
b) kun rotilla häirittiin vastamuodostuneiden solujen ohjelmoitua solukuolemaa, ne pystyivät edelleen parantamaan suoritustaan vesisokkelossa suunnistaessaan harjoituskerran aikana, mutta eivät enää muistaneet aiemmin oppimaansa reittiä vesisokkelossa.
c) kun hiirten neurogeneesiä vähennettiin juomaveteen sekoitetulla lääkeaineella, niiden suoriutuminen heikkeni tehtävässä, jossa tuli suunnistaa ulkoisten maamerkkien perusteella, mutta vain tilanteessa jossa lähtöpiste vaihteli.
d) kun aikuisten hiirten neurogeneesiä lisättiin yli kuukauden kestoisella aerobisella harjoittelulla oppimistapahtuman jälkeen, hiirten hippokampuksessa havaittiin verrokkeja enemmän uusia jyväissoluja ja ne muistivat aiemmin oppimaansa vahvemmin kuin verrokit.

Tehtävä 2.2

Tässä tehtävässä sinun tulee vastata 15 loogista päättelykykyä arvioivaan tehtävään. Jokaisessa tehtävässä on viisi vastausvaihtoehtoa (a-e), joista ainoastaan yksi on oikein. Useamman kuin yhden vastausvaihtoehdon valitseminen arvioidaan vääräksi vastaukseksi, samoin kuin vastauksen jättäminen tyhjäksi. Tehtävien vaikeustaso vaihtelee ja vaikeammista tehtävistä on mahdollisuus saada enemmän pisteitä kuin helpommista. Vastaa lomakkeeseen 2. Vastauslomakkeessa osatehtävät ovat riveillä ja eri vastausvaihtoehdot sarakkeissa.

1. Maija järjestää juhlat ja hän on kutsunut sinne kolme kaveriaan – Pekan, Markun ja Jarin. Nämä kertovat ennen juhlia seuraavat asiat:

Kaksi päivää ennen juhlia:
Pekka: Markku osallistuu juhliin
Markku: Jari osallistuu juhliin
Jari: Pekka osallistuu juhliin ainoastaan, jos minäkin osallistun

Päivää ennen juhlia:
Pekka: Jari osallistuu juhliin ainoastaan, jos minä en osallistu juhliin
Markku: Meistä kolmesta juhliin osallistuu parillinen määrä
Jari: Pekka osallistuu juhliin

Juhlapäivänä:
Pekka: Ei ole vielä vuosi 2018
Markku: Pekka osallistuu juhliin ainoastaan, jos minäkin osallistun
Jari: Ainakin yksi meistä kolmesta ei osallistu juhliin.

Maija tietää Pekasta, Markusta ja Jarista seuraavat asiat: Yksi heistä ei koskaan valehtele. Yksi heistä valehtelee parillisina päivinä, mutta muuten puhuu aina totta. Yksi heistä valehtelee niinä päivinä, jotka ovat jaollisia kolmella, mutta muuten puhuu aina totta. Kuka tai ketkä osallistuvat juhliin?

Vastausvaihtoehdot:

a) Pekka ja Markku
b) Pekka ja Jari
c) Markku ja Jari
d) ainoastaan Markku
e) Pekka, Markku ja Jari

2. Mirja harrastaa painonnostoa. Hänellä on neljä eri painoa. Hän punnitsee painot kaksi kerrallaan käyden läpi kaikki mahdolliset parit, jolloin näiden parien yhteispaino on 6, 8, 10, 12, 14, ja 16 kg. Kuinka paljon nämä yksittäiset neljä painoa painavat?

Vastausvaihtoehdot:

a) 1, 5, 7 ja 9 kg
b) 1, 4, 6 ja 10 kg
c) 2, 3, 7 ja 10 kg
d) 1, 2, 7 ja 9 kg
e) 2, 4, 6 ja 9 kg

3. Ville tapaa Aaron, Heikin, ja Niinan. Tutustumisleikkinä he alkavat arvuutella toistensa syntymäpäiviä. Ensimmäisenä arvataan Villen syntymäpäivää. Ville esittää muille 14 mahdollisuutta, joista yksi on hänen syntymäpäivänsä. Vaihtoehdot ovat:

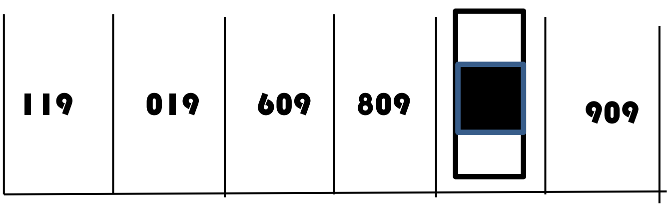
- 14.4.1999
- 19.2.2000
- 14.3.2000
- 15.3.2000
- 15.4.2000
- 16.4.2000
- 15.2.2001
- 15.3.2001
- 14.4.2001
- 16.4.2001
- 14.5.2001
- 16.5.2001
- 17.5.2001
- 17.2.2002

Hän paljastaa Aarolle kuukauden, Heikille päivän ja Niinalle vuoden siten, että muut eivät tiedä mitä muille on kerrottu. Kun hän on kertonut tämän, Aaro sanoo: ”En tiedä Villen syntymäpäivää, mutta ei tiedä Heikkikään”. Heikki sanoo: ”Se on totta, mutta Niinakaan ei tiedä Villen syntymäpäivää”. Niina sanoo: ”Kyllä ja Aaro ei ole keksinyt mikä Villen syntymäpäivä on”. Heikki vastaa: ” Nyt tiedän mikä on Villen syntymäpäivä”. Tässä vaiheessa Aaro sanoo: ”Kyllä, me tiedämme nyt kaikki Villen syntymäpäivän”. Mikä on Villen syntymäpäivä?

Vastausvaihtoehdot:

a) 15.3.2000
b) 15.4.2000
c) 16.4.2000
d) 15.3.2001
e) 16.5.2001

4. Parkkipaikalla on kuusi ruutua, jotka kaikki on numeroitu. Yhdessä ruudussa on auto, joka peittää numeron. Päättele oheisen kuvan perusteella mikä numero on siinä parkkiruudussa, jossa auto on.



Vastausvaihtoehdot:

a) 009
b) 109
c) 606
d) 709
e) 607

5. Korttipakasta, jossa kuuluisi olla 52 korttia, puuttuu kortteja. Kortit jaetaan siten, että kaikilla pelaajilla on täsmälleen sama määrä kortteja. Jos kortit jaetaan tasan neljän pelaajan kesken, niin jakamatta jää kolme korttia. Jos kortit jaetaan tasan kolmen pelaajan kesken jäljelle jää kaksi korttia ja jos kortit jaetaan tasan viiden pelaajan kesken, niin jäljelle jää kaksi korttia. Kuinka monta korttia on pakassa?

Vastausvaihtoehdot:

a) 43
b) 44
c) 46
d) 47
e) 48

6. Mikä luku tulee kysymysmerkin paikalle?
3+5+6=151830
5+5+6=253030
5+6+7=303542
5+5+3=251515
9+4+7=?

Vastausvaihtoehdot:

a) 283892
b) 362838
c) 362899
d) 366328
e) 635047

7. Monivalintatehtävässä annetaan yhdeksän (9) pistettä oikeasta vastauksesta ja jokaisesta väärästä vastauksesta vähennetään seitsemän (7) pistettä. Elina vastaa kaikkiin kysymyksiin ja hän saa pistemääräkseen testistä nolla (0) pistettä. Tiedetään, että testissä on alle 30 tehtävää. Kuinka monta tehtävää Elinan tekemässä testissä on?

Vastausvaihtoehdot:

a) ei voida ratkaista annetuilla tiedoilla
b) 16
c) 19
d) 21
e) 24

8. Mummo antoi perintönä puolet rahoistaan tyttärentyttärelleen ja puolet tyttärentyttären summasta pojanpojalleen. Yhden kuudesosan (1/6) koko perinnöstään hän antoi perintönä veljelleen ja jäljellejääneen 1000 € hän testamenttasi läheiselle koirahoitolalle. Kuinka monta euroa mummon perintö oli yhteensä?

Vastausvaihtoehdot:

a) ei voida ratkaista annetuilla tiedoilla
b) 10000 €
c) 11000 €
d) 12000 €
e) 15000 €

9. Pienessä lammessa kasvaa lumpeita. Joka ikinen päivä lumpeiden peittämä alue kasvaa kaksinkertaiseksi toisin sanoen tuplaantuu. Lumpeiden kasvuvauhti on niin huimaa, että 47 päivän kuluttua ne peittävät koko lammen. Kuinka monen päivän kohdalla lumpeet peittävät yhden neljäsosan lammesta?

Vastausvaihtoehdot:

a) 10
b) 20
c) 25
d) 35
e) 45

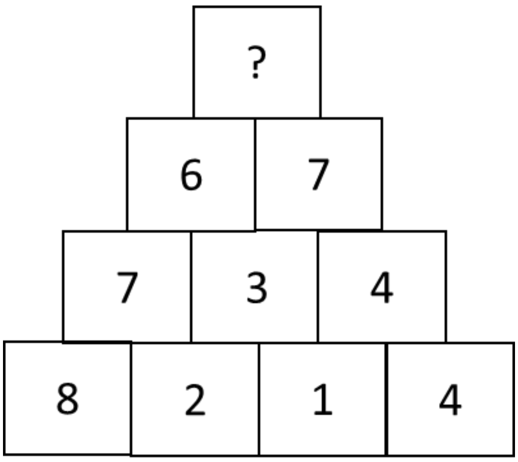
10. Bussit 1, 2 ja 3 kulkevat saman reitin joka päivä ja nämä bussit ovat sellaisia, joista yhtä matkustajat A, B, C, D, E, F ja G käyttävät päivittäin matkalla töihin. Tiedetään seuraavat asiat:
E eikä G eivät käytä bussia 1 niinä päivinä, kun B käyttää sitä.
G ei käytä bussia 2 niinä päivinä kuin D käyttää sitä.
Kun A ja F käyttävät samaa bussia, se on aina bussi numero 3
C käyttää aina bussia numero 3.

Mikäli B, C ja G haluaisivat matkustaa yhdessä töihin, niin mitä bussia/busseja he voisivat käyttää, kun edellä esitetyt ehdot ovat voimassa.

Vastausvaihtoehdot:

a) vain numeroa 1
b) vain numeroa 2
c) vain numeroa 3
d) vain numeroita 2 tai 3
e) mitä tahansa bussia 1, 2 tai 3

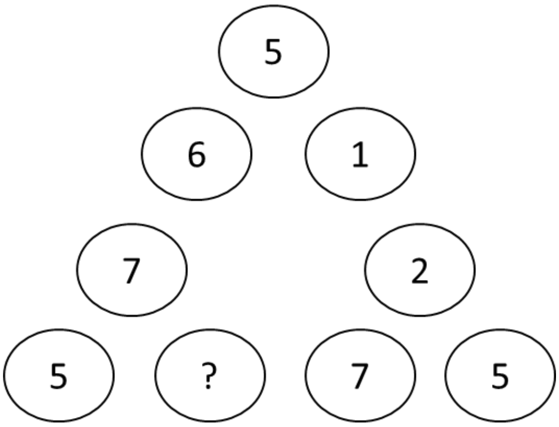
11. Mikä luku tulee kysymysmerkin paikalle?



Vastausvaihtoehdot:

a) 1
b) 4
c) 7
d) 12
e) 15

12. Mikä luku tulee kysymysmerkin paikalle?



Vastausvaihtoehdot:

a) 1
b) 2
c) 5
d) 6
e) 7

13. Mikä luku tulee kysymysmerkin paikalle?

2	6	8
3	7	3
6	4	?

Vastausvaihtoehdot:

a) 0
b) 1
c) 4
d) 6
e) 9

Seuraavissa kahdessa tehtävissä tehtävänannon alku on molemmissa sama. Kuitenkin viimeinen ehto vaihtuu tehtävän vaihtuessa. Aikaisemman tehtävän viimeinen ehto ei päde enää seuraavassa tehtävissä. Muuttuva osa on korostettu kursiivilla.

14. Janne osallistuu kesäleirillä kuuteen kilpailuun: pyöräilyyn, melontaan, ratsastukseen, rullaluisteluun, juoksuun ja uintiin. Hän sijoittuu jokaisessa kilpailussa viiden parhaan joukkoon (sijaluvuille 1.-5.). Eri kilpailuissa sijoitus on peräkkäinen ainoastaan, jos Jannen sijoituksen järjestysluvut ovat peräkkäiset. Janne sijoittuu peräkkäisille sijaluvuille melonnassa ja juoksussa. Myös hänen sijoituksensa rullaluistelussa ja uinnissa on peräkkäinen. Tiedetään myös, että hän sijoittuu korkeammalle pyöräilyssä kuin ratsastuksessa ja hän sijoittuu korkeammalle melonnassa kuin juoksussa.

Jos näiden tietojen lisäksi tiedetään vielä, että Janne sijoittuu korkeammalle juoksussa kuin pyöräilyssä ja korkeammalle pyöräilyssä kuin rullaluistelussa sekä uinnissa, niin mikä seuraavista väittämistä voidaan päätellä näillä tiedoilla?

Vastausvaihtoehdot:

a) Hän sijoittuu neljänneksi ratsastuksessa.
b) Hän sijoittuu neljänneksi rullaluistelussa.
c) Hänen sijoituksensa on sama ratsastuksessa ja rullaluistelussa.
d) Hänen sijoituksensa on sama ratsastuksessa ja uinnissa.
e) Hän sijoittuu korkeammalle ratsastuksessa kuin uinnissa.

15. Janne osallistuu kesäleirillä kuuteen kilpailuun: pyöräilyyn, melontaan, ratsastukseen, rullaluisteluun, juoksuun ja uintiin. Hän sijoittuu jokaisessa kilpailussa viiden parhaan joukkoon (sijaluvuille 1.-5.). Eri kilpailuissa sijoitus on peräkkäinen ainoastaan, jos Jannen sijoituksen järjestysluvut ovat peräkkäiset. Janne sijoittuu peräkkäisille sijaluvuille melonnassa ja juoksussa. Myös hänen sijoituksensa rullaluistelussa ja uinnissa on peräkkäinen. Tiedetään myös, että hän sijoittuu korkeammalle pyöräilyssä kuin ratsastuksessa ja hän sijoittuu korkeammalle melonnassa kuin juoksussa.

Jos näiden tietojen lisäksi tiedetään vielä, että Jannen sijoitus juoksussa on korkeampi kuin rullaluistelussa, jossa sijoitus on juoksuun nähden peräkkäinen ja hänen sijoituksensa uinnissa ja juoksussa ei ole sama, niin mikä seuraavista väitteistä on tosi joidenkin lajien kohdalla?

Vastausvaihtoehdot:

a) Hän sijoittuu sekä ensimmäiseksi ja toiseksi.
b) Hän sijoittuu sekä toiseksi että kolmanneksi.
c) Hän sijoittuu sekä toiseksi että neljänneksi.
d) Hän sijoittuu sekä toiseksi että viidenneksi.
e) Hän sijoittuu neljänneksi ja viidenneksi.

Tilastollisia taulukoita

Normitetun normaalijakauman kertymäfunktion arvoja

z	*0.00*	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*	*0.06*	*0.07*	*0.08*	*0.09*
0.00	0.5000000	0.5039894	0.5079783	0.5119665	0.5159534	0.5199388	0.5239222	0.5279032	0.5318814	0.5358564
0.10	0.5398278	0.5437953	0.5477584	0.5517168	0.5556700	0.5596177	0.5635595	0.5674949	0.5714237	0.5753454
0.20	0.5792597	0.5831662	0.5870644	0.5909541	0.5948349	0.5987063	0.6025681	0.6064199	0.6102612	0.6140919
0.30	0.6179114	0.6217195	0.6255158	0.6293000	0.6330717	0.6368307	0.6405764	0.6443088	0.6480273	0.6517317
0.40	0.6554217	0.6590970	0.6627573	0.6664022	0.6700314	0.6736448	0.6772419	0.6808225	0.6843863	0.6879331
0.50	0.6914625	0.6949743	0.6984682	0.7019440	0.7054015	0.7088403	0.7122603	0.7156612	0.7190427	0.7224047
0.60	0.7257469	0.7290691	0.7323711	0.7356527	0.7389137	0.7421539	0.7453731	0.7485711	0.7517478	0.7549029
0.70	0.7580363	0.7611479	0.7642375	0.7673049	0.7703500	0.7733726	0.7763727	0.7793501	0.7823046	0.7852361
0.80	0.7881446	0.7910299	0.7938919	0.7967306	0.7995458	0.8023375	0.8051055	0.8078498	0.8105703	0.8132671
0.90	0.8159399	0.8185887	0.8212136	0.8238145	0.8263912	0.8289439	0.8314724	0.8339768	0.8364569	0.8389129
1.00	0.8413447	0.8437524	0.8461358	0.8484950	0.8508300	0.8531409	0.8554277	0.8576903	0.8599289	0.8621434
1.10	0.8643339	0.8665005	0.8686431	0.8707619	0.8728568	0.8749281	0.8769756	0.8789995	0.8809999	0.8829768
1.20	0.8849303	0.8868606	0.8887676	0.8906514	0.8925123	0.8943502	0.8961653	0.8979577	0.8997274	0.9014747
1.30	0.9031995	0.9049021	0.9065825	0.9082409	0.9098773	0.9114920	0.9130850	0.9146565	0.9162067	0.9177356
1.40	0.9192433	0.9207302	0.9221962	0.9236415	0.9250663	0.9264707	0.9278550	0.9292191	0.9305634	0.9318879
1.50	0.9331928	0.9344783	0.9357445	0.9369916	0.9382198	0.9394292	0.9406201	0.9417924	0.9429466	0.9440826
1.60	0.9452007	0.9463011	0.9473839	0.9484493	0.9494974	0.9505285	0.9515428	0.9525403	0.9535213	0.9544860
1.70	0.9554345	0.9563671	0.9572838	0.9581849	0.9590705	0.9599408	0.9607961	0.9616364	0.9624620	0.9632730
1.80	0.9640697	0.9648521	0.9656205	0.9663750	0.9671159	0.9678432	0.9685572	0.9692581	0.9699460	0.9706210
1.90	0.9712834	0.9719334	0.9725711	0.9731966	0.9738102	0.9744119	0.9750021	0.9755808	0.9761482	0.9767045
2.00	0.9772499	0.9777844	0.9783083	0.9788217	0.9793248	0.9798178	0.9803007	0.9807738	0.9812372	0.9816911
2.10	0.9821356	0.9825708	0.9829970	0.9834142	0.9838226	0.9842224	0.9846137	0.9849966	0.9853713	0.9857379
2.20	0.9860966	0.9864474	0.9867906	0.9871263	0.9874545	0.9877755	0.9880894	0.9883962	0.9886962	0.9889893
2.30	0.9892759	0.9895559	0.9898296	0.9900969	0.9903581	0.9906133	0.9908625	0.9911060	0.9913437	0.9915758
2.40	0.9918025	0.9920237	0.9922397	0.9924506	0.9926564	0.9928572	0.9930531	0.9932443	0.9934309	0.9936128
2.50	0.9937903	0.9939634	0.9941323	0.9942969	0.9944574	0.9946139	0.9947664	0.9949151	0.9950600	0.9952012
2.60	0.9953388	0.9954729	0.9956035	0.9957308	0.9958547	0.9959754	0.9960930	0.9962074	0.9963189	0.9964274
2.70	0.9965330	0.9966358	0.9967359	0.9968333	0.9969280	0.9970202	0.9971099	0.9971972	0.9972821	0.9973646
2.80	0.9974449	0.9975229	0.9975988	0.9976726	0.9977443	0.9978140	0.9978818	0.9979476	0.9980116	0.9980738
2.90	0.9981342	0.9981929	0.9982498	0.9983052	0.9983589	0.9984111	0.9984618	0.9985110	0.9985588	0.9986051
3.00	0.9986501	0.9986938	0.9987361	0.9987772	0.9988171	0.9988558	0.9988933	0.9989297	0.9989650	0.9989992
3.10	0.9990324	0.9990646	0.9990957	0.9991260	0.9991553	0.9991836	0.9992112	0.9992378	0.9992636	0.9992886
3.20	0.9993129	0.9993363	0.9993590	0.9993810	0.9994024	0.9994230	0.9994429	0.9994623	0.9994810	0.9994991
3.30	0.9995166	0.9995335	0.9995499	0.9995658	0.9995811	0.9995959	0.9996103	0.9996242	0.9996376	0.9996505
3.40	0.9996631	0.9996752	0.9996869	0.9996982	0.9997091	0.9997197	0.9997299	0.9997398	0.9997493	0.9997585
3.50	0.9997674	0.9997759	0.9997842	0.9997922	0.9997999	0.9998074	0.9998146	0.9998215	0.9998282	0.9998347
3.60	0.9998409	0.9998469	0.9998527	0.9998583	0.9998637	0.9998689	0.9998739	0.9998787	0.9998834	0.9998879
3.70	0.9998922	0.9998964	0.9999004	0.9999043	0.9999080	0.9999116	0.9999150	0.9999184	0.9999216	0.9999247
3.80	0.9999277	0.9999305	0.9999333	0.9999359	0.9999385	0.9999409	0.9999433	0.9999456	0.9999478	0.9999499
3.90	0.9999519	0.9999539	0.9999557	0.9999575	0.9999593	0.9999609	0.9999625	0.9999641	0.9999655	0.9999670
4.00	0.9999683	0.9999696	0.9999709	0.9999721	0.9999733	0.9999744	0.9999755	0.9999765	0.9999775	0.9999784

t-jakaumaan liittyviä kriittisiä arvoja

	Todennäköisyys (jakauman oikean hännän pinta-ala)					
df	0.05	0.025	0.01	0.005	0.001	0.0005
1	6.313752	12.706205	31.820516	63.656741	318.308839	636.619249
2	2.919986	4.302653	6.964557	9.924843	22.327125	31.599055
3	2.353363	3.182446	4.540703	5.840909	10.214532	12.923979
4	2.131847	2.776445	3.746947	4.604095	7.173182	8.610302
5	2.015048	2.570582	3.364930	4.032143	5.893430	6.868827
6	1.943180	2.446912	3.142668	3.707428	5.207626	5.958816
7	1.894579	2.364624	2.997952	3.499483	4.785290	5.407883
8	1.859548	2.306004	2.896459	3.355387	4.500791	5.041305
9	1.833113	2.262157	2.821438	3.249836	4.296806	4.780913
10	1.812461	2.228139	2.763769	3.169273	4.143700	4.586894
11	1.795885	2.200985	2.718079	3.105807	4.024701	4.436979
12	1.782288	2.178813	2.680998	3.054540	3.929633	4.317791
13	1.770933	2.160369	2.650309	3.012276	3.851982	4.220832
14	1.761310	2.144787	2.624494	2.976843	3.787390	4.140454
15	1.753050	2.131450	2.602480	2.946713	3.732834	4.072765
16	1.745884	2.119905	2.583487	2.920782	3.686155	4.014996
17	1.739607	2.109816	2.566934	2.898231	3.645767	3.965126
18	1.734064	2.100922	2.552380	2.878440	3.610485	3.921646
19	1.729133	2.093024	2.539483	2.860935	3.579400	3.883406
20	1.724718	2.085963	2.527977	2.845340	3.551808	3.849516
21	1.720743	2.079614	2.517648	2.831360	3.527154	3.819277
22	1.717144	2.073873	2.508325	2.818756	3.504992	3.792131
23	1.713872	2.068658	2.499867	2.807336	3.484964	3.767627
24	1.710882	2.063899	2.492159	2.796940	3.466777	3.745399
25	1.708141	2.059539	2.485107	2.787436	3.450189	3.725144
26	1.705618	2.055529	2.478630	2.778715	3.434997	3.706612
27	1.703288	2.051831	2.472660	2.770683	3.421034	3.689592
28	1.701131	2.048407	2.467140	2.763262	3.408155	3.673906
29	1.699127	2.045230	2.462021	2.756386	3.396240	3.659405
30	1.697261	2.042272	2.457262	2.749996	3.385185	3.645959
40	1.683851	2.021075	2.423257	2.704459	3.306878	3.550966
50	1.675905	2.008559	2.403272	2.677793	3.261409	3.496013
60	1.670649	2.000298	2.390119	2.660283	3.231709	3.460200
70	1.666914	1.994437	2.380807	2.647905	3.210789	3.435015
80	1.664125	1.990063	2.373868	2.638691	3.195258	3.416337
90	1.661961	1.986675	2.368497	2.631565	3.183271	3.401935
100	1.660234	1.983972	2.364217	2.625891	3.173739	3.390491
200	1.652508	1.971896	2.345137	2.600634	3.131480	3.339835
300	1.649949	1.967903	2.338842	2.592316	3.117620	3.323252
400	1.648672	1.965912	2.335706	2.588176	3.110731	3.315015
500	1.647907	1.964720	2.333829	2.585698	3.106612	3.310091

χ^2-jakauman kriittisiä arvoja (todennäköisyys = jakauman oikean hännän pinta-ala)

df	*0.95*	*0.9*	*0.1*	*0.05*	*0.025*	*0.01*	*0.005*	*0.001*
1	0.00393	0.01579	2.70554	3.84146	5.02389	6.63490	7.87944	10.82757
2	0.10259	0.21072	4.60517	5.99146	7.37776	9.21034	10.59663	13.81551
3	0.35185	0.58437	6.25139	7.81473	9.34840	11.34487	12.83816	16.26624
4	0.71072	1.06362	7.77944	9.48773	11.14329	13.27670	14.86026	18.46683
5	1.14548	1.61031	9.23636	11.07050	12.83250	15.08627	16.74960	20.51501
6	1.63538	2.20413	10.64464	12.59159	14.44938	16.81189	18.54758	22.45774
7	2.16735	2.83311	12.01704	14.06714	16.01276	18.47531	20.27774	24.32189
8	2.73264	3.48954	13.36157	15.50731	17.53455	20.09024	21.95495	26.12448
9	3.32511	4.16816	14.68366	16.91898	19.02277	21.66599	23.58935	27.87716
10	3.94030	4.86518	15.98718	18.30704	20.48318	23.20925	25.18818	29.58830
11	4.57481	5.57778	17.27501	19.67514	21.92005	24.72497	26.75685	31.26413
12	5.22603	6.30380	18.54935	21.02607	23.33666	26.21697	28.29952	32.90949
13	5.89186	7.04150	19.81193	22.36203	24.73560	27.68825	29.81947	34.52818
14	6.57063	7.78953	21.06414	23.68479	26.11895	29.14124	31.31935	36.12327
15	7.26094	8.54676	22.30713	24.99579	27.48839	30.57791	32.80132	37.69730
16	7.96165	9.31224	23.54183	26.29623	28.84535	31.99993	34.26719	39.25235
17	8.67176	10.08519	24.76904	27.58711	30.19101	33.40866	35.71847	40.79022
18	9.39046	10.86494	25.98942	28.86930	31.52638	34.80531	37.15645	42.31240
19	10.11701	11.65091	27.20357	30.14353	32.85233	36.19087	38.58226	43.82020
20	10.85081	12.44261	28.41198	31.41043	34.16961	37.56623	39.99685	45.31475
21	11.59131	13.23960	29.61509	32.67057	35.47888	38.93217	41.40106	46.79704
22	12.33801	14.04149	30.81328	33.92444	36.78071	40.28936	42.79565	48.26794
23	13.09051	14.84796	32.00690	35.17246	38.07563	41.63840	44.18128	49.72823
24	13.84843	15.65868	33.19624	36.41503	39.36408	42.97982	45.55851	51.17860
25	14.61141	16.47341	34.38159	37.65248	40.64647	44.31410	46.92789	52.61966
26	15.37916	17.29188	35.56317	38.88514	41.92317	45.64168	48.28988	54.05196
27	16.15140	18.11390	36.74122	40.11327	43.19451	46.96294	49.64492	55.47602
28	16.92788	18.93924	37.91592	41.33714	44.46079	48.27824	50.99338	56.89229
29	17.70837	19.76774	39.08747	42.55697	45.72229	49.58788	52.33562	58.30117
30	18.49266	20.59923	40.25602	43.77297	46.97924	50.89218	53.67196	59.70306
40	26.50930	29.05052	51.80506	55.75848	59.34171	63.69074	66.76596	73.40196
50	34.76425	37.68865	63.16712	67.50481	71.42020	76.15389	79.48998	86.66082
60	43.18796	46.45889	74.39701	79.08194	83.29767	88.37942	91.95170	99.60723
70	51.73928	55.32894	85.52704	90.53123	95.02318	100.42518	104.21490	112.31693
80	60.39148	64.27784	96.57820	101.87947	106.62857	112.32879	116.32106	124.83922
90	69.12603	73.29109	107.56501	113.14527	118.13589	124.11632	128.29894	137.20835
100	77.92947	82.35814	118.49800	124.34211	129.56120	135.80672	140.16949	149.44925

Lukujen neliöjuuria. Taulukon sarakkeilla on lukujen sadasosat ja riveillä kymmenet. Esimerkiksi sarakkeella 4 ja rivillä 10 oleva luku 20.248 on luvun 410 neliöjuuri kolmen desimaalin tarkkuudella.

	SARAKE									
RIVI	0	1	2	3	4	5	6	7	8	9
0	0.000	10.000	14.142	17.321	20.000	22.361	24.495	26.458	28.284	30.000
1	1.000	10.050	14.177	17.349	20.025	22.383	24.515	26.476	28.302	30.017
2	1.414	10.100	14.213	17.378	20.050	22.405	24.536	26.495	28.320	30.033
3	1.732	10.149	14.248	17.407	20.075	22.428	24.556	26.514	28.337	30.050
4	2.000	10.198	14.283	17.436	20.100	22.450	24.576	26.533	28.355	30.067
5	2.236	10.247	14.318	17.464	20.125	22.472	24.597	26.552	28.373	30.083
6	2.449	10.296	14.353	17.493	20.149	22.494	24.617	26.571	28.390	30.100
7	2.646	10.344	14.387	17.521	20.174	22.517	24.637	26.589	28.408	30.116
8	2.828	10.392	14.422	17.550	20.199	22.539	24.658	26.608	28.425	30.133
9	3.000	10.440	14.457	17.578	20.224	22.561	24.678	26.627	28.443	30.150
10	3.162	10.488	14.491	17.607	20.248	22.583	24.698	26.646	28.460	30.166
11	3.317	10.536	14.526	17.635	20.273	22.605	24.718	26.665	28.478	30.183
12	3.464	10.583	14.560	17.664	20.298	22.627	24.739	26.683	28.496	30.199
13	3.606	10.630	14.595	17.692	20.322	22.650	24.759	26.702	28.513	30.216
14	3.742	10.677	14.629	17.720	20.347	22.672	24.779	26.721	28.531	30.232
15	3.873	10.724	14.663	17.748	20.372	22.694	24.799	26.739	28.548	30.249
16	4.000	10.770	14.697	17.776	20.396	22.716	24.819	26.758	28.566	30.265
17	4.123	10.817	14.731	17.804	20.421	22.738	24.839	26.777	28.583	30.282
18	4.243	10.863	14.765	17.833	20.445	22.760	24.860	26.796	28.601	30.299
19	4.359	10.909	14.799	17.861	20.469	22.782	24.880	26.814	28.618	30.315
20	4.472	10.954	14.832	17.889	20.494	22.804	24.900	26.833	28.636	30.332
21	4.583	11.000	14.866	17.916	20.518	22.825	24.920	26.851	28.653	30.348
22	4.690	11.045	14.900	17.944	20.543	22.847	24.940	26.870	28.671	30.364
23	4.796	11.091	14.933	17.972	20.567	22.869	24.960	26.889	28.688	30.381
24	4.899	11.136	14.967	18.000	20.591	22.891	24.980	26.907	28.705	30.397
25	5.000	11.180	15.000	18.028	20.616	22.913	25.000	26.926	28.723	30.414
26	5.099	11.225	15.033	18.055	20.640	22.935	25.020	26.944	28.740	30.430
27	5.196	11.269	15.067	18.083	20.664	22.956	25.040	26.963	28.758	30.447
28	5.292	11.314	15.100	18.111	20.688	22.978	25.060	26.981	28.775	30.463
29	5.385	11.358	15.133	18.138	20.712	23.000	25.080	27.000	28.792	30.480
30	5.477	11.402	15.166	18.166	20.736	23.022	25.100	27.019	28.810	30.496
31	5.568	11.446	15.199	18.193	20.761	23.043	25.120	27.037	28.827	30.512
32	5.657	11.489	15.232	18.221	20.785	23.065	25.140	27.055	28.844	30.529
33	5.745	11.533	15.264	18.248	20.809	23.087	25.159	27.074	28.862	30.545
34	5.831	11.576	15.297	18.276	20.833	23.108	25.179	27.092	28.879	30.561
35	5.916	11.619	15.330	18.303	20.857	23.130	25.199	27.111	28.896	30.578
36	6.000	11.662	15.362	18.330	20.881	23.152	25.219	27.129	28.914	30.594
37	6.083	11.705	15.395	18.358	20.905	23.173	25.239	27.148	28.931	30.610
38	6.164	11.747	15.427	18.385	20.928	23.195	25.259	27.166	28.948	30.627
39	6.245	11.790	15.460	18.412	20.952	23.216	25.278	27.185	28.965	30.643
40	6.325	11.832	15.492	18.439	20.976	23.238	25.298	27.203	28.983	30.659
41	6.403	11.874	15.524	18.466	21.000	23.259	25.318	27.221	29.000	30.676

	Sarake									
RIVI	0	1	2	3	4	5	6	7	8	9
42	6.481	11.916	15.556	18.493	21.024	23.281	25.338	27.240	29.017	30.692
43	6.557	11.958	15.588	18.520	21.048	23.302	25.357	27.258	29.034	30.708
44	6.633	12.000	15.620	18.547	21.071	23.324	25.377	27.276	29.052	30.725
45	6.708	12.042	15.652	18.574	21.095	23.345	25.397	27.295	29.069	30.741
46	6.782	12.083	15.684	18.601	21.119	23.367	25.417	27.313	29.086	30.757
47	6.856	12.124	15.716	18.628	21.142	23.388	25.436	27.331	29.103	30.773
48	6.928	12.166	15.748	18.655	21.166	23.409	25.456	27.350	29.120	30.790
49	7.000	12.207	15.780	18.682	21.190	23.431	25.475	27.368	29.138	30.806
50	7.071	12.247	15.811	18.708	21.213	23.452	25.495	27.386	29.155	30.822
51	7.141	12.288	15.843	18.735	21.237	23.473	25.515	27.404	29.172	30.838
52	7.211	12.329	15.875	18.762	21.260	23.495	25.534	27.423	29.189	30.854
53	7.280	12.369	15.906	18.788	21.284	23.516	25.554	27.441	29.206	30.871
54	7.348	12.410	15.937	18.815	21.307	23.537	25.573	27.459	29.223	30.887
55	7.416	12.450	15.969	18.841	21.331	23.558	25.593	27.477	29.240	30.903
56	7.483	12.490	16.000	18.868	21.354	23.580	25.612	27.495	29.257	30.919
57	7.550	12.530	16.031	18.894	21.378	23.601	25.632	27.514	29.275	30.935
58	7.616	12.570	16.062	18.921	21.401	23.622	25.652	27.532	29.292	30.952
59	7.681	12.610	16.093	18.947	21.424	23.643	25.671	27.550	29.309	30.968
60	7.746	12.649	16.125	18.974	21.448	23.664	25.690	27.568	29.326	30.984
61	7.810	12.689	16.155	19.000	21.471	23.685	25.710	27.586	29.343	31.000
62	7.874	12.728	16.186	19.026	21.494	23.707	25.729	27.604	29.360	31.016
63	7.937	12.767	16.217	19.053	21.517	23.728	25.749	27.622	29.377	31.032
64	8.000	12.806	16.248	19.079	21.541	23.749	25.768	27.641	29.394	31.048
65	8.062	12.845	16.279	19.105	21.564	23.770	25.788	27.659	29.411	31.064
66	8.124	12.884	16.310	19.131	21.587	23.791	25.807	27.677	29.428	31.081
67	8.185	12.923	16.340	19.157	21.610	23.812	25.826	27.695	29.445	31.097
68	8.246	12.961	16.371	19.183	21.633	23.833	25.846	27.713	29.462	31.113
69	8.307	13.000	16.401	19.209	21.656	23.854	25.865	27.731	29.479	31.129
70	8.367	13.038	16.432	19.235	21.679	23.875	25.884	27.749	29.496	31.145
71	8.426	13.077	16.462	19.261	21.703	23.896	25.904	27.767	29.513	31.161
72	8.485	13.115	16.492	19.287	21.726	23.917	25.923	27.785	29.530	31.177
73	8.544	13.153	16.523	19.313	21.749	23.937	25.942	27.803	29.547	31.193
74	8.602	13.191	16.553	19.339	21.772	23.958	25.962	27.821	29.563	31.209
75	8.660	13.229	16.583	19.365	21.794	23.979	25.981	27.839	29.580	31.225
76	8.718	13.266	16.613	19.391	21.817	24.000	26.000	27.857	29.597	31.241
77	8.775	13.304	16.643	19.416	21.840	24.021	26.019	27.875	29.614	31.257
78	8.832	13.342	16.673	19.442	21.863	24.042	26.038	27.893	29.631	31.273
79	8.888	13.379	16.703	19.468	21.886	24.062	26.058	27.911	29.648	31.289
80	8.944	13.416	16.733	19.494	21.909	24.083	26.077	27.928	29.665	31.305
81	9.000	13.454	16.763	19.519	21.932	24.104	26.096	27.946	29.682	31.321
82	9.055	13.491	16.793	19.545	21.954	24.125	26.115	27.964	29.698	31.337
83	9.110	13.528	16.823	19.570	21.977	24.145	26.134	27.982	29.715	31.353
84	9.165	13.565	16.852	19.596	22.000	24.166	26.153	28.000	29.732	31.369
85	9.220	13.601	16.882	19.621	22.023	24.187	26.173	28.018	29.749	31.385
86	9.274	13.638	16.912	19.647	22.045	24.207	26.192	28.036	29.766	31.401
87	9.327	13.675	16.941	19.672	22.068	24.228	26.211	28.054	29.783	31.417

	Sarake									
RIVI	0	1	2	3	4	5	6	7	8	9
88	9.381	13.711	16.971	19.698	22.091	24.249	26.230	28.071	29.799	31.432
89	9.434	13.748	17.000	19.723	22.113	24.269	26.249	28.089	29.816	31.448
90	9.487	13.784	17.029	19.748	22.136	24.290	26.268	28.107	29.833	31.464
91	9.539	13.820	17.059	19.774	22.159	24.310	26.287	28.125	29.850	31.480
92	9.592	13.856	17.088	19.799	22.181	24.331	26.306	28.142	29.866	31.496
93	9.644	13.892	17.117	19.824	22.204	24.352	26.325	28.160	29.883	31.512
94	9.695	13.928	17.146	19.849	22.226	24.372	26.344	28.178	29.900	31.528
95	9.747	13.964	17.176	19.875	22.249	24.393	26.363	28.196	29.917	31.544
96	9.798	14.000	17.205	19.900	22.271	24.413	26.382	28.213	29.933	31.559
97	9.849	14.036	17.234	19.925	22.293	24.434	26.401	28.231	29.950	31.575
98	9.899	14.071	17.263	19.950	22.316	24.454	26.420	28.249	29.967	31.591
99	9.950	14.107	17.292	19.975	22.338	24.474	26.439	28.267	29.983	31.607

Neliöjuuren laskusääntöjä: $\sqrt{\boldsymbol{A} \cdot \boldsymbol{B}} = \sqrt{\boldsymbol{A}} \cdot \sqrt{\boldsymbol{B}}$ **ja** $\sqrt{\frac{\boldsymbol{A}}{\boldsymbol{B}}} = \frac{\sqrt{\boldsymbol{A}}}{\sqrt{\boldsymbol{B}}}$

Kaavoja

Ohessa on kaavoja, joista joistakin saattaa olla apua tehtävien ratkaisemisessa. Kaavat on numeroitu ja kunkin kaavan numero on merkitty kaavan vasemmalle puolelle. Kaavojen numeroita käytetään tehtävissä, joissa vaaditaan oikean kaavan merkitsemistä vastauslomakkeeseen. HUOM! Osa kaavoista on virheellisiä.

1. $b_0 = \frac{\sum x_i - b_1(\sum y_1)}{n} = \bar{x} - b_1\bar{y}$

2. $b_0 = \frac{\sum y_i - b_1(\sum x_1)}{n} = \bar{y} - b_1\bar{x}$

3. $b_1 = \frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\bar{x})(y_i-\bar{y})}$

4. $b_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$

5. $b_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$

6. $b_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum y_i^2)}$

7. $b_1 = \frac{n(\sum x_i{}^2) - (\sum y_i{}^2)}{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}$

8. $b_1 = \frac{n(\sum x_i{}^2) - (\sum x_i)^2}{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}$

9. $b_1 = \frac{(\sum y_i)^2 - \frac{(\sum x_i)(\sum y_i)}{n}}{(\sum x_i{}^2) - \frac{(\sum y_i)^2}{n}}$

10. $b_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)^2}{n}}{(\sum x_i{}^2) - \frac{(\sum x_i)(\sum y_i)}{n}}$

11. $b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$

12. $b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$

13. $b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_2 y)^2}$

14. $b_2 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_2 y)^2}$

15. $b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 y)^2}$

16. $b_2 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 y)^2}$

17. $t = \frac{b}{SE_b}$

18. $t = \frac{SE_b}{b}$

19. $r = \frac{\sum_{i=1}^{n} xy - n\bar{x}\bar{y}}{\sqrt{[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2][\sum_{i=1}^{n} y_i^2 - n\bar{y}^2]}}$

20. $r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{[\sum_{i=1}^{n}(x_i-\bar{x})^2][\sum_{i=1}^{n}(y_i-\bar{y})^2]}}$

21. $r = \frac{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}}{[\sum_{i=1}^{n}(x_i-\bar{x})^2][\sum_{i=1}^{n}(y_i-\bar{y})^2]}$

22. $r = \frac{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}}{[\sum_{i=1}^{n}(x_i-\bar{x})^2][\sum_{i=1}^{n}(y_i-\bar{y})^2]}$

23. $r = \frac{s_{xy}}{s_x \cdot s_y}$

24. $r = \frac{s_x \cdot s_y}{s_{xy}}$

25. $r = 1 - \frac{6\cdot\sum_{i=1}^{n} d_i^2}{n^3 - n}$

26. $r = 1 - \frac{6\cdot\sum_{i=1}^{n} d_i^2}{n^2 - 1}$

27. $r = 1 - \frac{6\cdot\sum_{i=1}^{n} d_i^2}{n - 2}$

28. $P(A|B) = \frac{P(A)\cdot P(B|A)}{P(B)}$

Psykologian valintakoeyhteistyössä mukana olevien koulutusohjelmien valintakoemateriaalit 2019

Hakukohteet

Korkeakoulujen yhteishaku, kevät 2019

- Helsingin yliopisto: Päähaku, psykologian kandiohjelma, psykologian kandidaatti ja psykologian maisteri tai filosofian maisteri (3 v + 2,5/2 v)
- Helsingin yliopisto: Maisterihaku, psykologian maisteriohjelma, psykologian maisteri (2,5 v)
- Itä-Suomen yliopisto: Psykologia, psykologian kandidaatti ja maisteri (3 v + 2,5 v)
- Itä-Suomen yliopisto: Maisterihaku, Psykologia, Joensuu, psykologian maisteri (2,5 v)
- Jyväskylän yliopisto: Psykologian kandidaatti- ja maisteriohjelma, psykologian kandidaatti ja maisteri (3 v + 2,5 v)
- Jyväskylän yliopisto: Avoimen väylä, psykologian kandidaatti- ja maisteriohjelma, psykologian kandidaatti ja filosofian maisteri (3 v + 2 v)
- Tampereen yliopisto: Psykologian tutkinto-ohjelma, psykologian kandidaatti ja maisteri (3 v + 2,5 v)
- Turun yliopisto: Psykologia, psykologian kandidaatti ja psykologian maisteri (3 v + 2,5 v)

Valintakoemateriaalit

Tiistaina 9.4. klo 12.00 julkaistu materiaali (tilastomateriaalin korjaussivun viimeisin päivitys 12.4. klo 17.00). Materiaali on suomenkielinen.

	Artikkelit & tilastomatematiikan materiaali
1	Kantola, L. & Pirttimäki, T. & Nokia, M. S. (2017). Aikuisiän neurogeneesi hippokampuksessa mahdollistaa joustavan toiminnan. Psykologia, 52 (06), 436-456.
2	Keiski, P., Helminen, M., Lindroos, M., Kommeri, H., & Paavilainen, E. (2018) Nainen perheväkivallan tekijänä – ryhmäinterventio väkivaltakäyttäytymisen loppumiseksi. Sosiaalilääketieteellinen aikakauslehti 2018: 55: 143–155.
3	Kinnunen, Ulla & Feldt, Taru & Korpela, Kalevi & Mauno, Saija & Sianoja, Marjaana. (2017). Uupumusasteisen väsymyksen pysyvyys kahden vuoden aikana: yhteydet työstä palautumiseen vapaa-ajalla. Psykologia, 52 (04), 293-306.
4	Koivusalo, Laura & Aunola, Kaisa & Bertram, Raymond & Ryba, Tatiana V. (2018). Urheilija vai opiskelija? Urheilulukiolaisten identiteettiprofiilit. Liikunta ja tiede, 55 (2-3), 80-87.
5	Kontu, Elina & Ullgren, Iira-Maria & Törmänen, Minna & Nislin, Mari & Pirttimaa, Raija. (2013). Savant-lahjakkuus ja yleisen älykkyyden käsite. NMI-bulletin, 2013, Vol. 23, No. 1.
6	Linnansaari, A & Hankonen, N. (2019) Miten terveyskäyttäytymiseen voidaan vaikuttaa? Kirjasta: Terveyden psykologia (toim. Sanna Sinikallio), 5, s. 89-134.
7	Pohjola, M., Kykyri, V.-L., & Laitila, A. (2018). Neuvottelu vastuusta ja toimijuudesta yhteistyön kehittämiseen liittyvässä konsultaatiossa. Psykologia, 53 (05-06), 386-401, 449.
8	Psykologian valintakoe 2019 - Tilastomatematiikan materiaali ja korjaussivu.
9	Pyykönen, H., ym. (2019) Ajokyvyn arviointi MoCA-menetelmällä Alzheimerin taudin varhaisvaiheessa. Lääkärilehti 11/2019 vsk 74, 686-694.
10	Silvonen, J. (2017) Sielutieteestä psykologiaan – suomalaisen psykologiahistorian haasteita. Psykologia, 52 (02-03), 84-94.

Psykologian valintakoe 2019

Tilastomatematiikan materiaali

Aluksi

Tässä materiaalissa esitetään tiiviisti tilastotieteen peruskäsitteet ja analyysit, siinä laajuudessa kuin niitä edellytetään psykologian alan vuoden 2019 valinnoissa. Lisäksi materiaalissa on myös esitetty esitettyihin teemoihin liittyen valintakoekysymyksiä Psykologian valinnoista (Helsinki, Tampere ja Turku) aikaisemmilta vuosilta. Vanhat koetehtävät on pyritty valitsemaan sopivan kattavasti niin, että niiden ratkaisemisesta voisi olla aidosti hyötyä valintakokeeseen valmistautumisessa. Tehtäviin esitetään myös ratkaisumallit, jotka eivät kuitenkaan kaikissa tapauksissa ole ainoat oikeat, sillä kuten usein tilastomatematiikan tehtävissä, oikeaan ratkaisuun voi päätyä monella tavalla.

Materiaalin sisältöä kirjoitettaessa on pyritty sopivaan tasapainoon ytimekkään ja perinpohjaisen ilmaisun välillä. Materiaali ei pyri olemaan kattava kuvaus kaikista teemoista, joihin törmää esimerkiksi tilastotieteen johdantokursseilla, vaan tässä keskitytään ainoastaan tänä vuonna valintakokeessa keskeisimpiin aihealueisiin.

Koska tähän materiaaliin on otettu tehtäviä psykologian vanhojen valintakokeiden originaaliversioista, ei taitto – ikävä kyllä – ole aina paras mahdollinen. Ulkoasun puutteista huolimatta asiat toivottavasti selviävät lukijalle. Vanhat valintakoetehtävät alkavat monisteessa sivulta 43 alkaen.

Tilastolliset menetelmät ovat keskeisessä roolissa psykologiatieteessä, joka on hyvin menetelmäintensiivinen. Lisäksi suurin osa psykologisista arviointimenetelmistä on ns. psykometrisia, eli niiden kehittämisessä ja luotettavuuden arvioinnissa tilastotieteellä on keskeinen rooli. Materiaali toivon mukaan auttaa myös tulkitsemaan ja avaamaan muiden valintakoemateriaalina olevien artikkelien tuloksia ja helpottaa päättelyn perustalla olevan ajatusmallin ymmärtämistä.

Merkinnöistä

Lisämateriaalissa käytetään kahta tapaa summamerkinnän indekseissä. Ne kuvataan alla ja tarkoittavat siis täysin samaa.

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

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

Lähtökohtaisesti monisteessa käytetään latinalaisia aakkosia viitattaessa otoksen tunnuslukuihin esimerkiksi havaintoarvo (x), keskiarvo ($\bar{x}$) jne. vastaavasti kreikkalaisia aakkosia käytetään viitattaessa tietyn perusjoukon tunnuslukuihin, kuten esimerkiksi keskiarvo (μ), keskihajonta (σ) jne.

Tieteellinen tutkimus ja tilastollinen ajattelu

Tieteellisen tutkimuksen ja tilastollisen ajattelun / päättelyn filosofiasta voisi kirjoittaa useita kirjoja. Tässä monisteessa nämä asiat pääosin sivuutetaan. Tilastollisen päättelyn yhteydessä keskeistä on kuitenkin muuttujien mitta-asteikot tai tarkemmin mittauksen tasot, jotka kuvaavat tietyn muuttujan sisältämää informaatiota. Mittauksen tasojen määrittäminen liittyy keskeisesti tilastollisten analyysien oletuksiin, ja tyypillisesti eri analyyseissä joudutaan tekemään jokin oletus siitä, millä mittauksen tasolla tutkittava ilmiö on mitattu. Alla olevassa taulukossa on esitelty mittauksen tasot ja niihin liittyvät tärkeimmät ominaisuudet.

Asteikko / taso	kuvaus	Esimerkki
Nominaali- eli luokitteluasteikko	Tällä asteikolla voidaan tilastoida havaintoja, jotka voidaan luokitella johonkin ryhmään	Sukupuoli, ammatti
Ordinaali- eli järjestysasteikko	Tämän asteikon ryhmät voidaan järjestää yksikäsitteiseen järjestykseen jonkin kriteerin avulla	Koulutustaso, ikäluokka, juoksukilpailun tulokset
Intervalli- eli välimatka-asteikko	Tällä asteikolla voidaan havainnosta laskea erotus ja siten tarkastella havaintojen välistä etäisyyttä	Lämpötila Celsius-asteikolla
Suhdeasteikko	Asteikolla on absoluuttinen nollakohta	Lämpötila Kelvin-asteikolla, pituus, paino jne.

Varsinkaan psykologiassa ei mittauksen tason määrittäminen ole aina yksiselitteistä. Esimerkiksi älykkyysosamäärän voi ajatella olevan välimatka-asteikollinen, mutta toisaalta ei ole aina kaikissa tilanteissa järkevää ajatella, että esimerkiksi etäisyys pistemäärien 80–100 välillä olisi sama kuin etäisyys pistemäärien 100–120 välillä. Kuitenkin suurimman osan psykologisista mittaustuloksista (esim. kyselylomakkeiden pistemäärät) voi ajatella käyttäytyvän siten, että ne noudattavat tyydyttävää välimatka-asteikkoa.

Perustunnusluvut ja aineiston kuvailu

Tilastollisten aineistojen sisältämää informaatiota tiivistetään tyypillisesti erilaisiksi tunnusluvuiksi. Tässä kappaleessa on esitetty lyhyesti keskeisimmät tunnusluvut, joiden avulla voidaan kuvailla aineiston, erityisesti sen muuttujien, sisältämää informaatiota.

Sijaintiluvut

Sijaintiluvut kuvaavat nimensä mukaisesti muuttujien jakauman sijaintia. Keskeisimmät sijaintiluvut ja niiden kaavat on esitetty seuraavassa taulukossa.

Sijaintiluku	Kaava	Esimerkki
Moodi	Numeerisesti yleisin arvo	Aineistossa on 50 miestä ja 60 naista. Muuttujan sukupuoli moodiarvo on nainen
Mediaani	Aineiston järjestyksessä keskimmäinen havainto	Muuttujan arvot ovat 1, 2, 3, 4, 5. Muuttujan mediaani on arvo 3, koska se on järjestetyn aineiston keskimmäinen havainto. Jos havaintoja on parillinen määrä, niin myös järjestysasteikollisen muuttujan kohdalla ilmoitetaan mediaanina kahden keskimmäisen havainnon aritmeettinen keskiarvo.
Aritmeettinen keskiarvo	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	Muuttujan arvot ovat 2, 3, 5 ja 9. Tällöin (aritmeettinen) keskiarvo on: $\bar{x} = \frac{2+3+5+9}{4} = 4{,}75$
Fraktiilit		
Alakvartiili	Luku, jonka alapuolella on 25 % suuruusjärjestykseen järjestetyn muuttujan havainnoista	
Yläkvartiili	Luku, jonka alapuolella on 75 % suuruusjärjestykseen järjestetyn muuttujan havainnoista	

Hajontaluvut

Hajontalukujen avulla pyritään kuvaamaan aineiston vaihtelua, eli aineiston homogeenisuutta tai heterogeenisyyttä. Keskeisimmät hajontaluvut on esitetty alla olevassa taulukossa

Hajontaluku	Kaava	Esimerkki / tulkinta
Vaihteluväli	Aineiston pienin eli minimiarvo ja suurin eli maksimiarvo	Muuttujan arvot ovat 2, 10, 1, 4, 12 ja 8. Aineiston vaihteluväli on (1, 12)
Kvartiiliväli	Aineiston alakvartiilin ja yläkvartiilin arvo (kts. sijaintiluvut)	
Vaihteluvälin pituus	Maksimiarvo – minimiarvo	Muuttujan arvot ovat 2, 10, 1, 4, 12 ja 8. Aineiston vaihteluvälin pituus on 12 – 1 = 11
Kvartiilivälin pituus	Yläkvartiili – alakvartiili	
(otos) Keskihajonta	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$	Havaintoarvojen keskimääräinen vaihtelu keskiarvon ympärillä. Voidaan laskea myös kaavalla: $s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}}$

Esimerkkejä

Tutkitaan pientä aineistoa, jossa havaintoarvot ovat 2, 3, 3, 4, 8, 8, 12, 23 ja 25. Keskiarvo voidaan aikaisemmin todetun mukaisesti laskea kaavalla:

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

$$\bar{x} = \frac{2+3+3+4+8+8+12+23+25}{9} = \frac{88}{9} = 9\tfrac{7}{9} \approx 9{,}78$$

tai mikäli huomioidaan, että osa muuttujan havaintojenarvoista ovat samoja, seuraavasti

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{2 + 2\cdot 3 + 4 + 2\cdot 8 + 12 + 23 + 25}{1+2+1+2+1+1+1} = \frac{88}{9} = 9\tfrac{7}{9} \approx 9{,}78$$

Keskihajonta kuvaa havaintoarvojen keskimääräistä poikkeamista keskiarvosta, ja se on sopiva tunnusluku kertomaan arvojen vaihtelusta välimatka- ja suhdeasteikollisilla muuttujilla. Käsin laskien 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, joka on luokan todellisen alarajan ja todellisen ylärajan keskiarvo

Seuraavassa esitetään keskihajonnan laskeminen kummallakin tavalla pienelle aineistolle. Käsin laskemisessa käytetään apuna taulukkoon laskettuja välivaiheita.

	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
Σ	15795		1168,44	15593795

$$\bar{x} = \frac{15795}{16} = 987{,}1875 \approx 987{,}2$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1\,168.44}{15}} = \sqrt{77.896} \approx 8{,}83$$

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

$$= \sqrt{\frac{15\,593\,795 - 15\,592\,626.5625}{15}} = \sqrt{\frac{1\,168{,}4375}{15}} \approx 8{,}83$$

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

Laskemista voi vielä helpottaa vähentämällä ensin kaikista luvuista 900, sillä keskihajonta riippuu ainoastaan aineiston vaihtelusta keskiarvon ympärillä ja vakion vähentäminen jokaisesta havaintoarvosta muuttaa ainoastaan aineiston sijaintilukuja ei vaihtelua keskiarvon ympärillä.

	nettopaino − 900	x_i^2
	83	6889
	76	5776
	90	8100
	97	9409
	83	6889
	69	4761
	86	7396
	95	9025
	78	6084
	96	9216
	98	9604
	86	7396
	99	9801
	88	7744
	92	8464
	79	6241
Σ	1395	122795

$$\bar{x} = \frac{1395}{16} = 87{,}1875 \approx 87{,}2$$

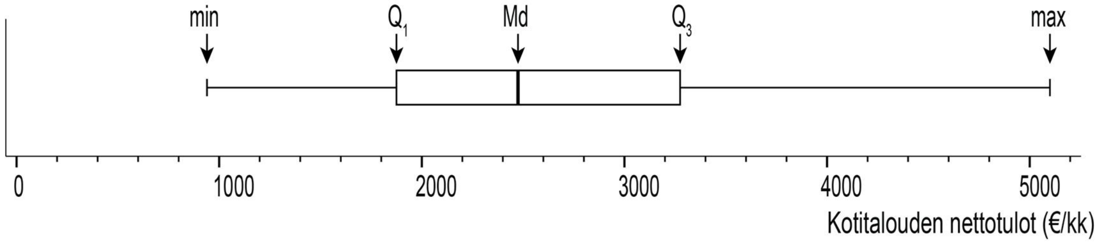
$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}} = \sqrt{\frac{122\,795 - \frac{1\,395^2}{16}}{15}} = \sqrt{\frac{122\,795 - 121\,626{,}5625}{15}} = \sqrt{\frac{1\,168{,}4375}{15}} \approx 8{,}83$$

Kun ilmoitetaan alkuperäisen muuttujan keskiarvo, tulee lukuun 87,2 muistaa lisätä 900. Tässä huomataan samalla miten vakion lisääminen (tai vähentäminen) kaikkien havaintoarvojen kohdalla vaikuttaa: keskiarvo muuttuu lisätyn (tai vähennetyn) vakion suuruudella, mutta keskihajonta ei muutu.

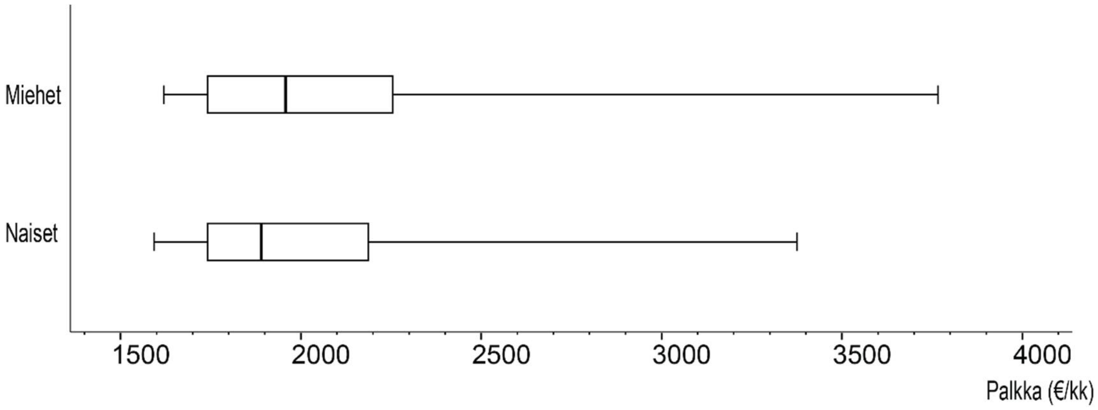
Kun hajonta lasketaan suoraan perusjoukon arvoista, kaavoissa on jakajana (n − 1) sijasta n.

Viisilukuinen yhteenveto ja laatikkokuvio

Muuttujan viisilukuinen yhteenveto on muuttujan pienimmän arvon, alakvartiilin, mediaanin, yläkvartiilin ja suurimman arvon järjestetty joukko (min, Q_1, Md, Q_3, max). Nämä viisi lukua jakavat siis aineiston neljään yhtä suureen osaan: havainnoista 25 % on välillä min–Q_1, 25 % välillä Q_1–Md, 25 % välillä Md–Q_3 ja 25 % välillä Q_3–max. Laatikkokuvio on viisilukuisen yhteenvedon kuvallinen esitys. Siitä on alla esimerkki erään aineiston muuttujasta kotitalouksien nettotulot (€/kk).



Laatikkokuvio on käyttökelpoinen myös ryhmäerojen kuvailussa. Alla esitetty miesten ja naisten palkkojen kuvailu, josta on helppo havaita erojen olevan erityisesti suurten palkkojen kohdalla.



Todennäköisyyslaskenta

Todennäköisyys on tietyn tapahtuman esiintyvyyden mitta. Todennäköisyys on reaaliluku, joka saa aina arvot suljetulta väliltä 0–1. Todennäköisyys saa arvon 0, kun tietty tapahtuma on mahdoton ja arvon 1, kun se tapahtuu varmasti. Mitä lähempänä todennäköisyys on arvoa 1, sitä yleisemmin se tapahtuu. Seuraavaksi esitellään ns. klassisen todennäköisyyden keskeisimmät laskusäännöt. Klassisessa todennäköisyyslaskennassa todennäköisyys määritellään suotuisien tapahtumien lukumäärän ja kaikkien mahdollisten tapahtumien lukumäärän osamääränä

$$p(A) = \frac{N(Suotuisat\ tapahtumat)}{N(kaikki\ tapahtumat)},$$

Esimerkiksi tarkastellaan nopanheittoa. Käytettäessä perinteistä noppaa on mahdollista saada kuusi eri silmälukua 1, 2, 3, 4, 5 ja 6 (6 kpl). Halutaan laskea todennäköisyys, että lopputuloksena on silmäluku 3 tai sitä pienempi silmäluku, jolloin suotuisia tapahtumia ovat arvot 1, 2 ja 3 (3 kpl). Pyydetty todennäköisyys on siis 3/6 eli sievennettynä 1/2.

Todennäköisyyslaskennassa tapahtuman A todennäköisyys on jokin reaaliluku 0:n ja 1:n väliltä eli

$$0 \leq p(A) \leq 1$$

Tapahtuman A vastatapahtuman eli komplementtitapahtuman ei-A todennäköisyys on

$$p(A^c) = 1 - p(A)$$

koska tapahtumalle ja sen vastatapahtumalle täytyy päteä kaava $p(A) + p(A^c) = 1$. Esimerkiksi todennäköisyys saada harhattoman nopan heitossa silmäluku 1, 2, 3, 4 tai 5 on

$$p(silmäluku\ 1, 2, 3, 4\ \ tai\ 5) = 1 - p(silmäluku\ 6) = 1 - \frac{1}{6} = \frac{5}{6}$$

Jos kaksi tapahtumaa A ja B ovat erillisiä eli toisensa poissulkevia tapahtumia, niin todennäköisyys sille, että joko tapahtuma A tai tapahtuma B tapahtuu (myös molemmat voivat tapahtua samanaikaisesti) on

$$p(A\ tai\ B) = p(A) + p(B) - p(A\ ja\ B)$$

Esimerkiksi korttipakassa on 52 korttia. Mikä on todennäköisyys, että nostettu kortti on hertta (13 kpl) tai ässä (neljä kappaletta)? Vastaus:

$$p(hertta\ tai\ ässä) = p(hertta) + p(ässä) - p(herttaässa) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}$$

Mikäli kaksi tapahtumaa ovat toisistaan riippumattomia, niin todennäköisyys, että molemmat tapahtuvat, on

$$p(A\ ja\ B) = p(A) \cdot p(B)$$

Kaavaa voidaan laajentaa myös useampaan erilliseen tapahtumaan. Esimerkki: Noppaa heitetään kaksi kertaa. Mikä on todennäköisyys saada molemmilla heitoilla silmäluku 6?

$$p(Noppa\ 1 = 6\ ja\ Noppa\ 2 = 6) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Mikäli tapahtumat A ja B eivät ole toisistaan riippumattomia, niin todennäköisyys sille, että molemmat tapahtuvat, on:

$p(A\ ja\ B) = p(A) \cdot p(B \mid A)$, jossa $p(B \mid A)$ tarkoittaa todennäköisyyttä, että B tapahtuu ehdolla, että A on tapahtunut. Tätä sääntöä kutsutaan myös yleiseksi kertolaskusäännöksi.

Seuraavaksi tarkastellaan hieman lisää tapahtumien erillisyyttä, riippumattomuutta ja ehdollista todennäköisyyttä sekä esitellään lisäksi kokonaistodennäköisyyden laskeminen ja Bayesin teoreema. Yhteislaskusäännön kohdalla vaikuttaa kuitenkin se, ovatko tapahtumat erillisiä (eli toisensa poissulkevia) vai eivät. Yleistä yhteenlaskusääntöä voi käyttää aina. Erillisten tapausten kohdalla sääntö yksinkertaistuu, kun termin $P(A\ ja\ B)$, joka on nolla, voi jättää pois. Jos tapahtumia on enemmän kuin kaksi ja ne kaikki ovat erillisiä, erillisten tapahtumien yhteenlaskusääntöä voi käyttää. Useamman kuin kahden tapahtuman yleinen yhteenlaskusääntö on hieman monimutkaisempi eikä sitä esitellä tässä materiaalissa. Riittää todeta, että ratkaisuissa voi käyttää apuna liitäntälakia.

$P(A\ tai\ B\ tai\ C) = P((A\ tai\ B)\ tai\ C) = P(A\ tai\ (B\ tai\ C))$.

Tähdennetään vielä, että tapahtumien erillisyys ja riippumattomuus ovat siis eri asioita.

Tapahtumat A ja B ovat erillisiä, jos $A\ ja\ B = \emptyset$ eli tyhjä joukko, jolloin $P(A\ ja\ B) = 0$.

Tapahtumat A ja B ovat riippumattomia, jos $P(A\ ja\ B) = P(A) \cdot P(B)$.

Tapahtumat eivät voi siis olla sekä erillisiä että riippumattomia. Ei-erilliset tapahtumat voivat olla joko riippumattomia tai riippuvaisia.

Yleinen kertolaskusääntö:

$P(A\ ja\ B) = P(A) \cdot P(B|A)$, jonka voi myös esittää $P(A\ ja\ B) = P(B) \cdot P(A|B)$.

Siis $P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$, josta voidaan johtaa Bayesin kaava

$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$ ja $P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}$.

Riippumattomuus voidaan ilmaista siis myös seuraavasti. Tapahtumat A ja B ovat riippumattomia, jos $P(A) \cdot P(B|A) = P(A) \cdot P(B)$. Kun molemmat puolet supistetaan *P(A)*:lla, saadaan $P(B|A) = P(B)$. Vastaavasti tapahtumat A ja B ovat riippumattomia, jos $P(A|B) = P(A)$. Eli tapahtumat A ja B ovat riippumattomia, jos A:n tapahtuminen ei vaikuta B:n tapahtumisen todennäköisyyteen eikä vastaavasti B:n tapahtuminen vaikuta A:n tapahtumisen todennäköisyyteen.

Kokonaistodennäköisyyttä tarvitaan tilanteissa, joissa jokin tapahtuma voi toteutua usealla eri tavalla (ehdolla). Tarkastellaan esimerkkiä. Kutsuille on varattu kolme korillista limonadia. Yhdessä korissa on 6 kolajuomaa, toisessa korissa on 8 kolajuomaa ja kolmannessa korissa on 12 kolajuomaa. Liisan isä hakee kellarista sattumanvaraisesti yhden korin, ja näkemättä koria ja pulloja Liisa valitsee korista yhden pullon. Mikä on todennäköisyys, että Liisa saa kolajuoman?

Merkitään $P(A_1)$ = {isä valitsee korin 1} = $\frac{1}{3}$, $P(A_2)$ = {isä valitsee korin 2} = $\frac{1}{3}$, $P(A_3)$ = {isä valitsee korin 3} = $\frac{1}{3}$ ja P(B) = {Liisa saa kolajuoman}.

$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3)$

$$= \frac{1}{3} \cdot \frac{6}{24} + \frac{1}{3} \cdot \frac{8}{24} + \frac{1}{3} \cdot \frac{12}{24} = \frac{3}{36} + \frac{4}{36} + \frac{6}{36} = \frac{13}{36}$$

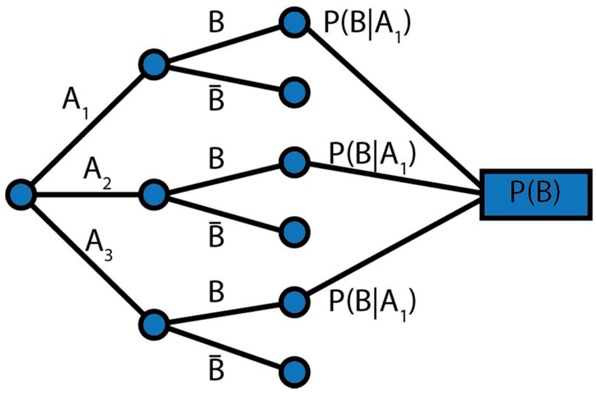
Tässä tapauksessa tämä on luonnollisesti sama kuin kolajuomien kokonaismäärä jaettuna kaikkien limonadipullojen määrällä $\frac{6+8+12}{3\cdot 24} = \frac{26}{72} = \frac{13}{36}$. Toisissa tilanteissa, joissa $P(A_1)$, $P(A_2)$, $P(A_3)$,…, $P(A_n)$ eivät kaikki ole yhtä suuria, vastaavuus ei ole näin suoraviivainen.

Bayesin teoreeman avulla voidaan esimerkin tilanteessa vastata seuraavaan kysymykseen. Kun tiedetään, että Liisa sai kolajuoman, niin millä todennäköisyydellä isä toi kellarista korin 1?

$$P(A_1|B) = \frac{P(A_1) \cdot P(B|A_1)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3)} =$$

$$\frac{\frac{1}{3} \cdot \frac{6}{24}}{\frac{1}{3} \cdot \frac{6}{24} + \frac{1}{3} \cdot \frac{8}{24} + \frac{1}{3} \cdot \frac{12}{24}} = \frac{\frac{3}{36}}{\frac{13}{36}} = \frac{3}{13}$$

Esimerkin voi havainnollistaa seuraavalla puurakenteella.



Kokonaistodennäköisyys on kaikkien ”alusta loppuun” kulkevien polkujen yhdistelmä.

Kun tiedetään, että jokin tapahtuma on tapahtunut, niin Bayesin teoreeman avulla voidaan selvittää, mikä on tapahtumaan johtaneiden eri vaihtoehtojen todennäköisyys. Bayesin teoreemaa voidaan käyttää vastaavissa tilanteissa monissa yhteyksissä. Kuvataan asiaa vielä kahden esimerkin avulla:

Tiedetään, että virustautia sairastaa 5 % perusjoukosta. Viruksen toteamiseksi on olemassa laboratoriotesti, joka antaa joko positiivisen tai negatiivisen tuloksen (positiivinen tulos tarkoittaa, että testin mukaan henkilöllä on tautia aiheuttava virus, ja negatiivinen testitulos tarkoittaa, että henkilöllä ei ole virusta). Niiden henkilöiden kohdalla, joilla on virus, testi antaa oikean (positiivisen) tuloksen todennäköisyydellä 0.95. Terveiden henkilöiden kohdalla testi antaa oikean (negatiivisen) tuloksen todennäköisyydellä 0.90. Satunnaisesti valittu perusjoukon henkilö menee testiin ja testin tulos on positiivinen. Millä todennäköisyydellä tällä henkilöllä on virus?

Kootaan annetut tiedot yhteen. Merkitään V = henkilöllä on virus, T = henkilöllä ei ole virusta eli henkilö on terve, P = testin tulos on positiivinen ja N = testin tulos on negatiivinen. Huomioidaan, että $T = \bar{V}$ ja $N = \bar{P}$.

On ehkä kuitenkin helpompi käyttää ensin mainittuja merkintöjä. $P(V) = 0.05, P(T) = P(\bar{V}) = 1 - 0,05 = 0,95$, $P(P|V) = 0,95$ ja $P(N|T) = 0,90$.

Ratkaistava todennäköisyys on puolestaan $P(V|P)$. Lasketaan ensin kokonaistodennäköisyys sille, että testi antaa positiivisen tuloksen.

$$P(P) = P(V) \cdot P(P|V) + P(\bar{V}) \cdot P(P|\bar{V}) = 0,05 \cdot 0,95 + (1 - 0,05) \cdot (1 - 0,90)$$
$$= 0,05 \cdot 0,95 + 0,95 \cdot 0,10 = 0,1425$$

Bayesin teoreeman perusteella

$P(V|P) = \frac{0,05 \cdot 0,95}{0,05 \cdot 0,95 + 0,95 \cdot 0,10} = \frac{0,0475}{0,1425} = \frac{1}{3}$.

Eli testatuista henkilöistä, joiden testitulos on positiivinen, vain joka kolmannella on virus.

Tarkastellaan vielä yhtä esimerkkiä. Kokeessa on yksi monivalintatehtävä, jossa on viisi vastausvaihtoehtoa. Vaihtoehdoista yksi on oikea ja neljä ovat vääriä. Tehtävään vastanneet henkilöt ovat joko hyvin valmistautuneita, melko hyvin valmistautuneita tai täysin valmistautumattomia. Hyvin valmistautuneita on 40 %, melko hyvin valmistautuneita samoin 40 % ja valmistautumattomia on 20 %. Hyvin valmistautuneet valitsevat oikean vaihtoehdon todennäköisyydellä 0,90, melko hyvin valmistautuneet todennäköisyydellä 0,50 ja valmistautumattomat arvaavat satunnaisesti.

a) Kuinka iso osa monivalintatehtävän vastauksista on oikein?

b) Jos hakija on vastannut oikein, millä todennäköisyydellä hän on hyvin valmistautunut?

Kohdassa a) on selvitettävä kokonaistodennäköisyys.

$P = 0,4 \cdot 0,9 + 0,4 \cdot 0,5 + 0,2 \cdot 0,2 = 0,6$.

Kohta b) ratkeaa Bayesin teoreeman avulla. $P = \frac{0,4 \cdot 0,9}{0,4 \cdot 0,9 + 0,4 \cdot 0,5 + 0.2 \cdot 0,2} = 0,6$.

Tehtäviä (vastaukset ovat seuraavalla sivulla)

1) Eräässä käräjäoikeudessa niistä syytetyistä, jotka ovat syyllistyneet rikokseen, vapautetaan 8 % ja niistä syytetyistä, jotka eivät ole syyllistyneet rikokseen, tuomitaan 6 %. Kaikista syytetyistä 80 % on syyllisiä.

    a) Mikä on todennäköisyys, että tuomittu henkilö on syyllistynyt rikokseen?

    b) Mikä on todennäköisyys, että tuomittu henkilö ei ole syyllistynyt rikokseen?

    c) Mikä on todennäköisyys, että vapautettu henkilö on syyllistynyt rikokseen?

2) Syventävälle kurssille osallistumisen vaatimuksena on, että opiskelijalla on peruskurssilla saatavat tiedot. Nämä voi saavuttaa joko avoimen yliopiston kurssilla tai kirjatentissä tai yliopiston kurssilla tai kirjatentissä. Oletetaan, että edellä mainitut neljä tapaa ovat toisensa poissulkevia. Syventävälle kurssille osallistuvista 30 % on suorittanut avoimen yliopiston peruskurssin ja 10 % kirjatentin, 40 % on suorittanut yliopiston peruskurssin ja 20 % kirjatentin. Syventävän kurssin hyväksytyn suorituksen on saanut 80 % avoimessa yliopistossa peruskurssin suorittaneista ja 90 % yliopistossa peruskurssin suorittaneista.

    a) Mikä on todennäköisyys, että syventävästä kurssista hylätyn suorituksen saanut on käynyt avoimen yliopiston peruskurssin?

    b) Mikä on todennäköisyys, että syventävästä kurssista hylätyn suorituksen saanut on suorittanut yliopiston peruskurssin kirjatentillä?

3) Kaupassa myynnissä olevista tomaateista 1/3 tulee puutarhalta A, 1/4 puutarhalta B ja loput puutarhalta C. Puutarhan A tomaateista 94 % on myyntikelpoisia, puutarhan B tomaateista 90 % on myyntikelpoisia ja puutarhan C tomaateista vain 88 % on myyntikelpoisia.

a) Millä todennäköisyydellä satunnaisesti valittu tomaatti on myyntiin kelpaamaton?

b) Millä todennäköisyydellä myyntiin kelpaamaton tomaatti on peräisin puutarhalta A?

Tehtävien vastaukset

1. a) $P = \frac{0.8\cdot(1-0.08)}{0.8\cdot(1-0.08)+(1-0.8)\cdot 0.06} = \frac{184}{187} \approx 0{,}984$

b) $P = \frac{(1-0.8)\cdot 0.06}{0.8\cdot(1-0.08)+(1-0.8)\cdot 0.06} = \frac{3}{187} = 1 - \frac{184}{187} \approx 0{,}016$

c) $P = \frac{0.8\cdot 0.08}{0.8\cdot 0.08+(1-0.8)\cdot(1-0.06)} = \frac{16}{63} \approx 0{,}254$

2. a) $P = \frac{0.3\cdot(1-0{,}8)}{(0{,}3+0{,}1)\cdot(1-0.8)+(0.4+0.2)\cdot(1-0.9)} = \frac{3}{7} \approx 0{,}429$

b) $P = \frac{0.2\cdot(1-0.9)}{(0{,}3+0{,}1)\cdot(1-0{,}8)+(0{,}4+0.2)\cdot(1-0.9)} = \frac{1}{7} \approx 0{,}143$

3. a) $P = \frac{1}{3}\cdot(1-0{,}94) + \frac{1}{4}\cdot(1-0{,}90) + \frac{5}{12}\cdot(1-0{,}88) = \frac{19}{200} = 0{,}095$

b) $= \frac{\frac{1}{3}\cdot(1-0{,}94))}{\frac{1}{3}\cdot(1-0{,}94)+\frac{1}{4}\cdot(1-0{,}90)+\frac{5}{12}\cdot(1-0{,}88)} = \frac{4}{19} \approx 0{,}211$

Kombinatoriikka

Todennäköisuuslaskennassa törmätään usein kysymykseen, mikä on kaikkien mahdollisten tapahtumien ja suotuisien tapahtumien lukumäärä. Tähän kysymykseen pystytään tyypillisesti vastaamaan kombinatoriikan laskusäännöillä.

Esimerkiksi kuinka monella eri tavalla n määrä ihmisiä voidaan asettaa jonoon? Tähän antaa vastauksen permutaatio, jossa ensimmäiselle paikalle voidaan asettaa 10 henkilöä, toiselle 9, kolmannelle 8 jne. Eri tapoja järjestää jono on siis n! (n kertoma) lukumäärä, eli $n! = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot ... \cdot 1$, eli esimerkin tapauksessa mahdollisia jonoja on $10! = 10 \cdot 9 \cdot 8 \cdot ... \cdot 1 = 3628800$ kappaletta.

Vastaavasti tilanteessa, jossa kymmenen ihmisen joukosta arvotaan 5 henkilöä, jotka asetetaan jonoon, voidaan kaikkien mahdollisten jonojen lukumäärä laskea hyödyntäen variaation käsitettä. Esimerkin tilanteessa mahdollisia jonoja on $10 \cdot 9 \cdot 8 \cdot ... \cdot 6 = 30240$ kappaletta. Yleisesti, jos tapahtumassa on n kappaletta alkioita ja näistä halutaan tarkastella sellaisia variaatioita, joissa esiintyy k kappaletta alkioita, niin mahdollisten tapahtumien lukumäärä on $n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot ... \cdot (n-k)$, joka voidaan laskea myös kaavalla $\frac{n!}{(n-k)!}$. Esimerkiksi esimerkissä mainitussa tilanteessa lukumäärä on siis $\frac{n!}{(n-k)!} = \frac{10!}{(10-5)!} = \frac{3628800}{120} = 30240$.

Variaation kohdalla lähdetään ajatuksesta, että alkioiden järjestyksellä on merkitystä. Sen sijaan kombinaatiossa valitaan n alkiota sisältävästä joukosta k kappaletta alkioita, mutta huomioimatta erillistä järjestystä. Esimerkiksi arpalipukkeet on numeroitu numeroilla 1–10. Nostetaan kaksi arpalipuketta ja kysymys kuuluu, kuinka monella tavalla voidaan valita lipukkeet, joissa on numerot 1 ja 2. Tässä tilanteessa ei ole merkitystä sillä, saadaanko ensin numero 1 ja sitten numero 2 vaiko toisin päin. Kombinaatioiden lukumäärä voidaan laskea kaavalla

$$\frac{n!}{k! \cdot (n-k)!},$$

josta käytetään tyypillisesti myös merkintää $\binom{n}{k}$.

Esimerkki: Jos lotossa arvotaan 7 numeroa 39 numeron joukosta. Kuinka monta erilaista lottoriviä on olemassa?

Vastaus:

$$\binom{39}{7} = \frac{39!}{7! \cdot (39-7)!} = 15380937$$

Todennäköisyysjakaumat

Satunnaismuuttujan käyttäytymistä voidaan kuvata todennäköisyysjakaumalla. Todennäköisyysjakauma on tilastollinen malli, joka kuvaa eri havaintoarvojen esiintymistodennäköisyyksiä. Keskeisimmät todennäköisyysjakaumat ovat normaalijakauma, t-jakauma ja χ^2-jakauma, jotka ovat keskeisessä roolissa mm. tilastollisessa päätöksenteossa ja hypoteesintestauksessa.

Normaalijakauma

Normaalijakauma on symmetrinen todennäköisyysjakauma, jonka sijainnin määrittää jakauman odotusarvo (keskiarvo) ja muodon keskihajonta. Seuraavalla sivulla olevassa taulukossa on z välillä 0,00–3,69. Z-arvon kokonaisosa ja kymmenesosa luetaan 1. sarakkeesta ja sadasosa 1. riviltä (z). Taulukosta sarakkeen ja rivin leikkauskohdasta löytyy kertymäfunktion arvo.

Siis esimerkiksi jos tarvitsee selvittää $P(Z < 1{,}75) = \Phi(1{,}75)$, jossa Φ on kertymäfunktion arvo, taulukon kolmannelta sivulta etsitään sarakkeesta z luku 1,7 ja riviltä z luku 0,05. 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 komplementtitapahtuman säännöllä.

Se, että käytetään aina standardoitua normaalijakaumaa, jossa keskiarvo on nolla (0) ja keskihajonta yksi (1), voi vaikuttaa ensisilmäyksellä rajoittavalta. Kuitenkin mikä tahansa satunnaismuuttuja voidaan standardoida kaavalla.

$$z_i = \frac{x_i - \bar{x}}{s}$$

Eli jokaisesta havaintoarvosta vähennetään muuttujan keskiarvo ja jaetaan tämä erotus muuttujan hajonnalla. Esimerkiksi jos tiedetään (tai tunnusluvut on laskettu aineistosta), että älykkyysosamäärän keskiarvo on 100 ja keskihajonta on 15 ja älykkyysosamäärä noudattaa normaalijakaumaa, niin todennäköisyys havaita älykkyysosamäärä, joka on alle 130, voidaan laskea seuraavasti:

$X \sim N(100, 15)$, niin $Z = ((X - 100) / 15) \sim N(0, 1)$. Tällöin $p(X < 130)$

$= p(Z((130 - 100) / 15)) = \Phi(2) \approx 0{,}977$

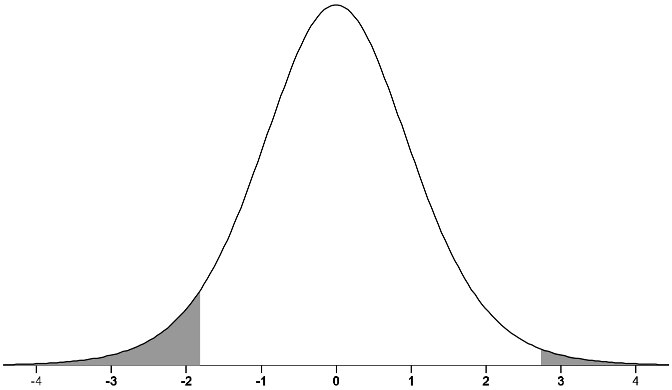
Taulukko: Normaalijakauman kertymäfunktion arvoja

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000000	0,5039894	0,5079783	0,5119665	0,5159534	0,5199388	0,5239222	0,5279032	0,5318814	0,5358564
0,10	0,5398278	0,5437953	0,5477584	0,5517168	0,5556700	0,5596177	0,5635595	0,5674949	0,5714237	0,5753454
0,20	0,5792597	0,5831662	0,5870644	0,5909541	0,5948349	0,5987063	0,6025681	0,6064199	0,6102612	0,6140919
0,30	0,6179114	0,6217195	0,6255158	0,6293000	0,6330717	0,6368307	0,6405764	0,6443088	0,6480273	0,6517317
0,40	0,6554217	0,6590970	0,6627573	0,6664022	0,6700314	0,6736448	0,6772419	0,6808225	0,6843863	0,6879331
0,50	0,6914625	0,6949743	0,6984682	0,7019440	0,7054015	0,7088403	0,7122603	0,7156612	0,7190427	0,7224047
0,60	0,7257469	0,7290691	0,7323711	0,7356527	0,7389137	0,7421539	0,7453731	0,7485711	0,7517478	0,7549029
0,70	0,7580363	0,7611479	0,7642375	0,7673049	0,7703500	0,7733726	0,7763727	0,7793501	0,7823046	0,7852361
0,80	0,7881446	0,7910299	0,7938919	0,7967306	0,7995458	0,8023375	0,8051055	0,8078498	0,8105703	0,8132671
0,90	0,8159399	0,8185887	0,8212136	0,8238145	0,8263912	0,8289439	0,8314724	0,8339768	0,8364569	0,8389129
1,00	0,8413447	0,8437524	0,8461358	0,8484950	0,8508300	0,8531409	0,8554277	0,8576903	0,8599289	0,8621434
1,10	0,8643339	0,8665005	0,8686431	0,8707619	0,8728568	0,8749281	0,8769756	0,8789995	0,8809999	0,8829768
1,20	0,8849303	0,8868606	0,8887676	0,8906514	0,8925123	0,8943502	0,8961653	0,8979577	0,8997274	0,9014747
1,30	0,9031995	0,9049021	0,9065825	0,9082409	0,9098773	0,9114920	0,9130850	0,9146565	0,9162067	0,9177356
1,40	0,9192433	0,9207302	0,9221962	0,9236415	0,9250663	0,9264707	0,9278550	0,9292191	0,9305634	0,9318879
1,50	0,9331928	0,9344783	0,9357445	0,9369916	0,9382198	0,9394292	0,9406201	0,9417924	0,9429466	0,9440826
1,60	0,9452007	0,9463011	0,9473839	0,9484493	0,9494974	0,9505285	0,9515428	0,9525403	0,9535213	0,9544860
1,70	0,9554345	0,9563671	0,9572838	0,9581849	0,9590705	0,9599408	0,9607961	0,9616364	0,9624620	0,9632730
1,80	0,9640697	0,9648521	0,9656205	0,9663750	0,9671159	0,9678432	0,9685572	0,9692581	0,9699460	0,9706210
1,90	0,9712834	0,9719334	0,9725711	0,9731966	0,9738102	0,9744119	0,9750021	0,9755808	0,9761482	0,9767045
2,00	0,9772499	0,9777844	0,9783083	0,9788217	0,9793248	0,9798178	0,9803007	0,9807738	0,9812372	0,9816911
2,10	0,9821356	0,9825708	0,9829970	0,9834142	0,9838226	0,9842224	0,9846137	0,9849966	0,9853713	0,9857379
2,20	0,9860966	0,9864474	0,9867906	0,9871263	0,9874545	0,9877755	0,9880894	0,9883962	0,9886962	0,9889893
2,30	0,9892759	0,9895559	0,9898296	0,9900969	0,9903581	0,9906133	0,9908625	0,9911060	0,9913437	0,9915758
2,40	0,9918025	0,9920237	0,9922397	0,9924506	0,9926564	0,9928572	0,9930531	0,9932443	0,9934309	0,9936128
2,50	0,9937903	0,9939634	0,9941323	0,9942969	0,9944574	0,9946139	0,9947664	0,9949151	0,9950600	0,9952012
2,60	0,9953388	0,9954729	0,9956035	0,9957308	0,9958547	0,9959754	0,9960930	0,9962074	0,9963189	0,9964274
2,70	0,9965330	0,9966358	0,9967359	0,9968333	0,9969280	0,9970202	0,9971099	0,9971972	0,9972821	0,9973646
2,80	0,9974449	0,9975229	0,9975988	0,9976726	0,9977443	0,9978140	0,9978818	0,9979476	0,9980116	0,9980738
2,90	0,9981342	0,9981929	0,9982498	0,9983052	0,9983589	0,9984111	0,9984618	0,9985110	0,9985588	0,9986051
3,00	0,9986501	0,9986938	0,9987361	0,9987772	0,9988171	0,9988558	0,9988933	0,9989297	0,9989650	0,9989992
3,10	0,9990324	0,9990646	0,9990957	0,9991260	0,9991553	0,9991836	0,9992112	0,9992378	0,9992636	0,9992886
3,20	0,9993129	0,9993363	0,9993590	0,9993810	0,9994024	0,9994230	0,9994429	0,9994623	0,9994810	0,9994991
3,30	0,9995166	0,9995335	0,9995499	0,9995658	0,9995811	0,9995959	0,9996103	0,9996242	0,9996376	0,9996505
3,40	0,9996631	0,9996752	0,9996869	0,9996982	0,9997091	0,9997197	0,9997299	0,9997398	0,9997493	0,9997585
3,50	0,9997674	0,9997759	0,9997842	0,9997922	0,9997999	0,9998074	0,9998146	0,9998215	0,9998282	0,9998347
3,60	0,9998409	0,9998469	0,9998527	0,9998583	0,9998637	0,9998689	0,9998739	0,9998787	0,9998834	0,9998879

Studentin t-jakauma

Studentin t-jakauma (lyhyemmin t-jakauma) on symmetrinen, mutta sen muoto määräytyy vapausasteiden (df) mukaan. Vapausasteiden tarkempi määrittely sivuutetaan ja ne voi ajatella tiettyjen jakaumien muotoon vaikuttavana parametrina. Studentin t-jakauman taulukossa (seuraava sivu) esitetään tiettyihin merkitsevyystasoihin liittyvät testisuureen kriittiset arvot. Taulukon vasemmasta reunasta valitaan vapausasteluku. Vapausasteiden riviltä nähdään testisuureen kriittiset arvot.

Jos esimerkiksi t-jakauman vapausasteluku on 10, niin todennäköisyys, että $t > 1.812$ on 0.05 (likiarvo kahden desimaalin tarkkuudella) ja todennäköisyys että $t > 2.764$ on 0.01 (likiarvo tämäkin). Vastaavasti symmetrian takia todennäköisyys, että $t < -1.812$ on myös 0.05 ja $t < -2.764$ on 0.01.

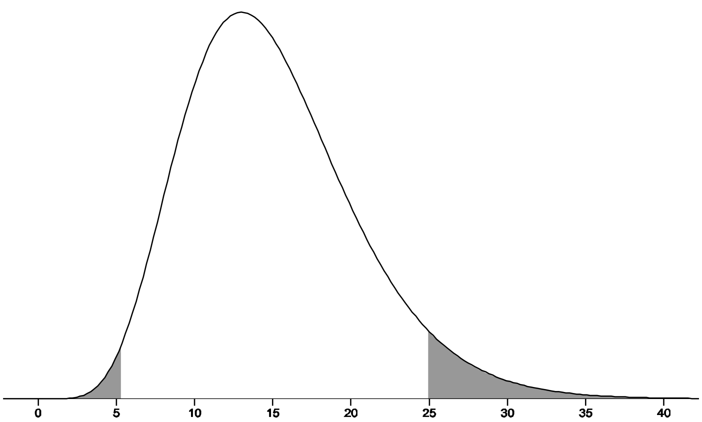


Kuva: Esimerkki t-jakauman tiheysfunktiosta

χ^2-jakauma

χ^2-jakauma ei ole symmetrinen ja sen muoto määräytyy vapausasteiden mukaan. χ^2-jakauman kriittiset arvot on taulukoitu vapausasteiden ja merkitsevyystason mukaan. Taulukko on esitetty Studentin t-jakauman taulukon jälkeen.

Jos esimerkiksi χ^2-jakauman vapausasteluku on 15, niin todennäköisyys että $\chi^2 < 7,26$ on 0,05 (siis 1−0,95 likiarvo kahden desimaalin tarkkuudella) ja todennäköisyys että $\chi^2 > 24,996$ on 0,05 (likiarvo).



Kuva: Esimerkki χ^2-jakauman tiheysfunktiosta

Taulukko: t-jakauman kriittisiä arvoja eri todennäköisyyksillä

	todennäköisyys					
df	0,05	0,025	0,01	0,005	0,001	0,0005
1	6,313752	12,706205	31,820516	63,656741	318,308839	636,619249
2	2,919986	4,302653	6,964557	9,924843	22,327125	31,599055
3	2,353363	3,182446	4,540703	5,840909	10,214532	12,923979
4	2,131847	2,776445	3,746947	4,604095	7,173182	8,610302
5	2,015048	2,570582	3,364930	4,032143	5,893430	6,868827
6	1,943180	2,446912	3,142668	3,707428	5,207626	5,958816
7	1,894579	2,364624	2,997952	3,499483	4,785290	5,407883
8	1,859548	2,306004	2,896459	3,355387	4,500791	5,041305
9	1,833113	2,262157	2,821438	3,249836	4,296806	4,780913
10	1,812461	2,228139	2,763769	3,169273	4,143700	4,586894
11	1,795885	2,200985	2,718079	3,105807	4,024701	4,436979
12	1,782288	2,178813	2,680998	3,054540	3,929633	4,317791
13	1,770933	2,160369	2,650309	3,012276	3,851982	4,220832
14	1,761310	2,144787	2,624494	2,976843	3,787390	4,140454
15	1,753050	2,131450	2,602480	2,946713	3,732834	4,072765
16	1,745884	2,119905	2,583487	2,920782	3,686155	4,014996
17	1,739607	2,109816	2,566934	2,898231	3,645767	3,965126
18	1,734064	2,100922	2,552380	2,878440	3,610485	3,921646
19	1,729133	2,093024	2,539483	2,860935	3,579400	3,883406
20	1,724718	2,085963	2,527977	2,845340	3,551808	3,849516
21	1,720743	2,079614	2,517648	2,831360	3,527154	3,819277
22	1,717144	2,073873	2,508325	2,818756	3,504992	3,792131
23	1,713872	2,068658	2,499867	2,807336	3,484964	3,767627
24	1,710882	2,063899	2,492159	2,796940	3,466777	3,745399
25	1,708141	2,059539	2,485107	2,787436	3,450189	3,725144
26	1,705618	2,055529	2,478630	2,778715	3,434997	3,706612
27	1,703288	2,051831	2,472660	2,770683	3,421034	3,689592
28	1,701131	2,048407	2,467140	2,763262	3,408155	3,673906
29	1,699127	2,045230	2,462021	2,756386	3,396240	3,659405
30	1,697261	2,042272	2,457262	2,749996	3,385185	3,645959
40	1,683851	2,021075	2,423257	2,704459	3,306878	3,550966
50	1,675905	2,008559	2,403272	2,677793	3,261409	3,496013
60	1,670649	2,000298	2,390119	2,660283	3,231709	3,460200
70	1,666914	1,994437	2,380807	2,647905	3,210789	3,435015
80	1,664125	1,990063	2,373868	2,638691	3,195258	3,416337
90	1,661961	1,986675	2,368497	2,631565	3,183271	3,401935
100	1,660234	1,983972	2,364217	2,625891	3,173739	3,390491

Taulukko: χ^2-jakauman kriittisiä arvoja eri todennäköisyyksillä

	Todennäköisyys							
df	0,95	0,9	0,1	0,05	0,025	0,01	0,005	0,001
1	0,00393	0,01579	2,70554	3,84146	5,02389	6,63490	7,87944	10,82757
2	0,10259	0,21072	4,60517	5,99146	7,37776	9,21034	10,59663	13,81551
3	0,35185	0,58437	6,25139	7,81473	9,34840	11,34487	12,83816	16,26624
4	0,71072	1,06362	7,77944	9,48773	11,14329	13,27670	14,86026	18,46683
5	1,14548	1,61031	9,23636	11,07050	12,83250	15,08627	16,74960	20,51501
6	1,63538	2,20413	10,64464	12,59159	14,44938	16,81189	18,54758	22,45774
7	2,16735	2,83311	12,01704	14,06714	16,01276	18,47531	20,27774	24,32189
8	2,73264	3,48954	13,36157	15,50731	17,53455	20,09024	21,95495	26,12448
9	3,32511	4,16816	14,68366	16,91898	19,02277	21,66599	23,58935	27,87716
10	3,94030	4,86518	15,98718	18,30704	20,48318	23,20925	25,18818	29,58830
11	4,57481	5,57778	17,27501	19,67514	21,92005	24,72497	26,75685	31,26413
12	5,22603	6,30380	18,54935	21,02607	23,33666	26,21697	28,29952	32,90949
13	5,89186	7,04150	19,81193	22,36203	24,73560	27,68825	29,81947	34,52818
14	6,57063	7,78953	21,06414	23,68479	26,11895	29,14124	31,31935	36,12327
15	7,26094	8,54676	22,30713	24,99579	27,48839	30,57791	32,80132	37,69730
16	7,96165	9,31224	23,54183	26,29623	28,84535	31,99993	34,26719	39,25235
17	8,67176	10,08519	24,76904	27,58711	30,19101	33,40866	35,71847	40,79022
18	9,39046	10,86494	25,98942	28,86930	31,52638	34,80531	37,15645	42,31240
19	10,11701	11,65091	27,20357	30,14353	32,85233	36,19087	38,58226	43,82020
20	10,85081	12,44261	28,41198	31,41043	34,16961	37,56623	39,99685	45,31475
21	11,59131	13,23960	29,61509	32,67057	35,47888	38,93217	41,40106	46,79704
22	12,33801	14,04149	30,81328	33,92444	36,78071	40,28936	42,79565	48,26794
23	13,09051	14,84796	32,00690	35,17246	38,07563	41,63840	44,18128	49,72823
24	13,84843	15,65868	33,19624	36,41503	39,36408	42,97982	45,55851	51,17860
25	14,61141	16,47341	34,38159	37,65248	40,64647	44,31410	46,92789	52,61966
26	15,37916	17,29188	35,56317	38,88514	41,92317	45,64168	48,28988	54,05196
27	16,15140	18,11390	36,74122	40,11327	43,19451	46,96294	49,64492	55,47602
28	16,92788	18,93924	37,91592	41,33714	44,46079	48,27824	50,99338	56,89229
29	17,70837	19,76774	39,08747	42,55697	45,72229	49,58788	52,33562	58,30117
30	18,49266	20,59923	40,25602	43,77297	46,97924	50,89218	53,67196	59,70306
40	26,50930	29,05052	51,80506	55,75848	59,34171	63,69074	66,76596	73,40196
50	34,76425	37,68865	63,16712	67,50481	71,42020	76,15389	79,48998	86,66082

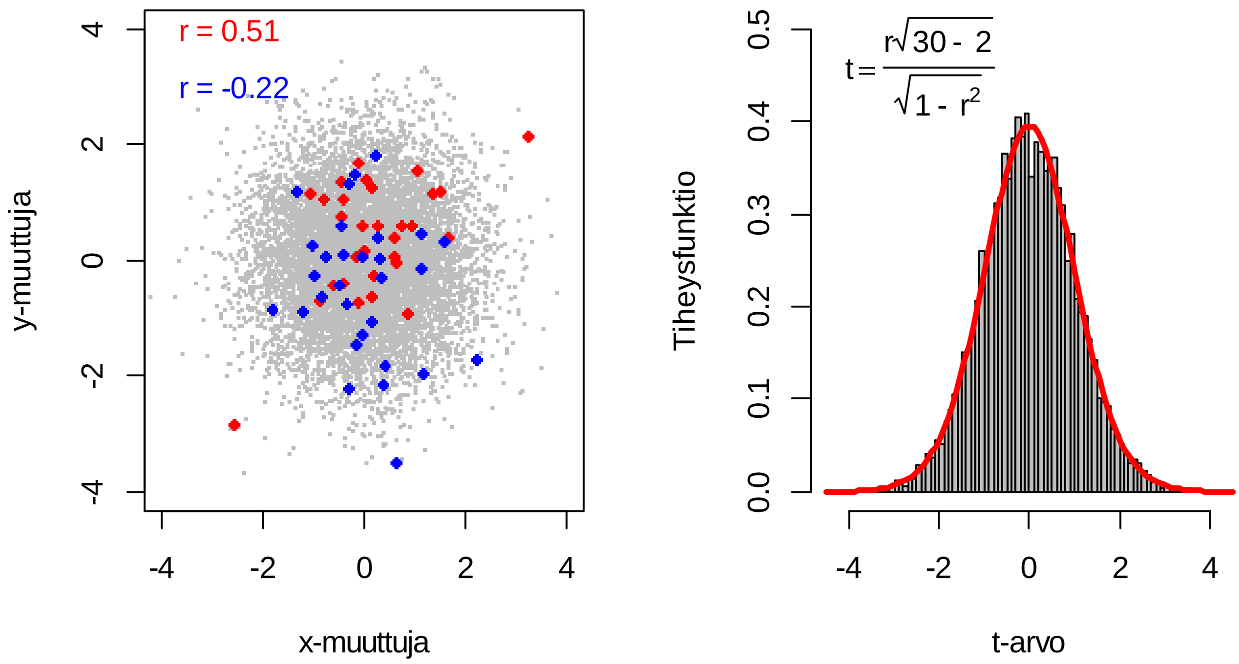
Tilastollinen päätöksenteko ja hypoteesien testaaminen

Tilastollisessa päätöksenteossa ja erityisesti hypoteesin testauksessa pyritään aineiston perusteella tekemään päätelmiä tilanteesta perusjoukossa otoksen perusteella. Tavallisemmin ns. klassisessa frekventistisessä tulkinnassa tiivistetään otantavirheen tuottama vaihtelu tunnuslukuun, jota kutsutaan p-arvoksi. P-arvo on tietty todennäköisyys, joka lasketaan tietystä tilastollisesta testistä lasketun testisuureen perusteella. Vielä tarkemmin voidaan sanoa, että **p-arvo on todennäköisyys sille, että saadaan otoksesta lasketun arvon verran poikkeava tai sitä vielä poikkeavampi tulos nollahypoteesin mukaisesta arvosta olettaen, että nollahypoteesi on tosi**. P-arvon tulkinnassa saattaa tulla vastaan useita väärinymmärryksiä, joita tässä pyritään ennalta ehkäisemään. **P-arvo ei ole** todennäköisyys sille, että nollahypoteesi on tosi, **p-arvo ei ole** todennäköisyys hylätä väärä nollahypoteesi **eikä p-arvo ole** todennäköisyys sille, että havaittu tulos olisi saatu sattumalta. Karkeasti voisi sanoa, että mitä pienempi p-arvo on, sitä vahvempi ”todiste” otoksesta laskettu testisuure on nollahypoteesia vastaan. Tässä materiaalissa ei todisteta kaavojen toimivuutta tai johdeta tilastollisia kaavoja esitettyä muotoa pidemmälle. Lisäksi keskitytään ainoastaan keskeisimpiin tilastollisiin testeihin.

Alla on esitetty tilastollisen päätöksenteon keskeisimmät virhemahdollisuudet taulukkomuodossa.

	Todellisuus	
Päätös	**H_0 epätosi**	**H_0 tosi**
H_0 hylätään	1. Oikea päätös	2. Hylkäämisvirhe (tyypin 1 virhe eli väärä positiivinen päätös)
H_0 jää voimaan	3. Hyväksymisvirhe (tyypin 2 virhe eli väärä negatiivinen päätös)	4. Oikea päätös

Klassinen tilastollinen päättely perustuu siis pitkälti ajatukseen siitä, että perusjoukosta otetaan otoksia äärettömänä toistokokeena. Esimerkiksi seuraavassa kuvassa on vasemmalla esitetty kahden muuttujan hajontakuvaaja. Harmaalla merkityt pisteet muodostavat perusjoukon, jonka koko on suuri. Havaitaan, että perusjoukossa ei muuttujien välillä vaikuta olevan minkäänlaista yhteyttä, eli korrelaatiokerroin, jonka käsitteeseen palataan myöhemmin tässä monisteessa, on perusjoukossa nolla. Myöhemmin havaitaan, että tämä tilanne on korrelaatiokertoimen tilastollisen merkitsevyyden testauksessa nollahypoteesin mukainen. Tästä perusjoukosta otetaan otoksia, joiden jokaisen otoskoko tässä esimerkissä on 30 havaintoa. Kuvaajaan on väritetty pisteet kahdesta otoksesta. Punaisella väritettyjen pisteiden välinen otoskorrelaatio on 0,51 ja sinisellä väritettyjen pisteiden muodostamassa otoksessa otoskorrelaatio on −0,22. Vastaavia otoksia otetaan tästä perusjoukosta 10000 kappaletta ja jokaiselle lasketulle korrelaatiolle on tehty muuttujamuunnos histogrammissa esitetyn kaavan mukaisesti. Tähänkin kaavaan palataan myöhemmin korrelaatiota käsittelevässä kappaleessa. Histogrammista havaitaan, että 10000 otoksen otoskorrelaatiot käyttäytyvät varsin systemaattisesti. Itse asiassa voidaan osoittaa, että ne noudattavat t-jakaumaa vapausasteilla $n - 2$ ($n = 30$ tässä esimerkissä). Korrelaatiokertoimen tapauksessa otantavirheestä aiheutuva vaihtelu voidaan siis mallintaa t-jakaumaa hyödyntäen, jolloin voidaan laskea yksittäiselle korrelaatiokertoimelle todennäköisyys, että sellainen tai sitä suurempi korrelaatiokerroin esiintyy sattumalta sillä ehdolla, että nollahypoteesi on tosi. Mikäli tämä todennäköisyys, jota kutsutaan myös p-arvoksi, on hyvin pieni, ryhdytään klassisessa frekventistisessä päättelyssä epäilemään taustalla olevaa oletusta nollahypoteesista. Vastaava päättelyketju tehdään jokaisen klassisen tilastollisen analyysin kohdalla. On syytä huomata, että päättely on hieman takaperoista, sillä siinä lähdetään epäilemään taustalla olevaa oletusta nollahypoteesin oikeellisuudesta, ei sitä, että sattumalta olisi otannassa osunut otantavirheen takia kohdalle harvinainen aineisto.



Tyypillisesti tilastollisessa päättelyssä valitaan niin kutsutuksi riskitasoksi jokin pieni todennäköisyys, kuten 5 %, 1 % tai 0,1 %, eli halutaan, että havaittu p-arvo on alle tämän tietyn riskitason, jotta voidaan sanoa, että tulos on tilastollisesti merkitsevä ja nollahypoteesi hylätään. Tarkasti ottaen p-arvo on havaitun testisuureen itseisarvon todennäköisyysjakauman hännille jäävän jakauman pinta-ala t-testeissä ja z-testeissä, jotka tyypillisesti ovat kaksisuuntaisia ja oikealle jäävän hännän pinta-ala yksisuuntaisissa testeissä (esim. khii toiseen -testi). P-arvo voidaan siis laskea selvittämällä todennäköisyys $p(|t| \geq havaittu\ testisuure)$.

Khii toiseen -riippumattomuustesti

Khii toiseen -riippumattomuustestin avulla pyritään selvittämään, voidaanko aineiston perusteella päätellä, että kahden nominaali-asteikollisen (kategorisen) muuttujan välillä on yhteyttä perusjoukossa.

Testissä asetetaan seuraavat hypoteesit:

H0: Muuttujien välillä ei ole riippuvuutta

H1: Muuttujien välillä on riippuvuutta

Testisuure voidaan laskea ristiintaulukosta, jossa on r kappaletta rivejä ja s kappaletta sarakkeita kaavalla:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(o_{rs}-e_{rs})^2}{e_{rs}}$$

Kaavassa o_{rs} on ristiintaulukon rivin r ja sarakkeen s solun havaittu frekvenssi ja e_{rs} on vastaavan solun odotettu frekvenssi. Odotettu frekvenssi voidaan laskea tietyn rivin ja tietyn sarakkeen frekvenssien rivi ja sarakesummien avulla seuraavilla kaavoilla.

$$e_{ij} = \frac{(\sum_{i=1}^{r} o_i)(\sum_{i=1}^{s} o_s)}{N}$$

Kaavassa lasketaan tietyn solun odotettu frekvenssi laskemalla ensin kyseessä olevan rivin r havaittujen frekvenssien summa ja kerrotaan tämä vastaavan sarakkeen havaittujen frekvenssien summalla. Tämä tulo jaetaan vielä aineiston kokonais-N:llä, eli otoskoolla. Mikäli jokaisessa solussa odotetun frekvenssin arvo on suurempi tai yhtä suuri kuin 5, niin noudattaa näiden laskukaavojen avulla laskettu testisuure khii toiseen -jakaumaa vapausasteilla $(r-1)(s-1)$, jossa r ja s ovat järjestyksessä rivien ja sarakkeiden lukumäärä.

Esimerkki. Tarkastellaan seuraavaa frekvenssitaulukkoa, jossa muodostettu ristiintaulukko eräästä otoksesta (N=500), jossa on tutkittu USA:ssa demokraattien ja republikaanien kantoja erääseen lakialoitteeseen. Taulukkoon on laskettu valmiiksi myös rivien ja sarakkeiden summat.

	Puolesta	**Ei osaa sanoa**	**Vastaan**	**Yhteensä**
Demokraatti	138	83	64	285
Republikaani	64	67	84	215
Yhteensä	202	150	148	500

Alla olevaan taulukkoon on laskettu odotetut frekvenssit jokaiselle ristiintaulukon solulle.

	Puolesta	**Ei osaa sanoa**	**Vastaan**	**Yhteensä**
Demokraatti	285·202/500=115.14	285·150/500=85.5	285·148/500=84.36	285
Republikaani	215·202/500=86.86	215·150/500=64.5	215·148/500=63.64	215
Yhteensä	202	150	148	500

Testisuureen laskemin on tämän jälkeen suoraviivaista:

$$X^2 = \frac{(138-115.14)^2}{115.14} + \frac{(83-85.5)^2}{85.5} + \cdots + \frac{(84-63.64)^2}{63.64} = 22.152$$

Testisuure noudattaa khii toiseen -jakaumaa vapausasteilla $(2-1)(3-1)=2$. Khii toiseen -jakauman kertymäfunktion taulukosta (kappale todennäköisyysjakaumat) havaitaan, että 5 % riskitason kriittinen arvo (eli arvo, jonka oikealla puolella on 5 % jakauman pinta-alasta) on likimäärin 5.99. Koska testisuure 22.152 > 5.99, on testiin liittyvä p-arvo pienempi kuin 5 %, eli tyypillisesti merkitään $p < 0{,}05$.

Yhden otoksen t-testi ja luottamusväli

Yhden otoksen t-testillä pyritään selvittämään voiko otos olla peräisin perusjoukosta, joka on normaalisti jakautunut tietyllä keskiarvolla. Toisin sanoen tutkimuksen kohteena on tarkastella, poikkeaako tietyn muuttujan keskiarvo perusjoukossa jostain tietystä teoreettisesta arvosta.

Yhden otoksen t-testi lasketaan kaavalla:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{N}}$$

Jossa $\bar{x}$ on otoksesta laskettu keskiarvo, μ_0 perusjoukon keskiarvo, johon havaittua otoskeskiarvoa halutaan verrata, N on otoksen koko ja s otoskeskihajonta, joka lasketaan kaavalla

$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$ tai kaavalla $s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}}$.

Tyypillisesti yhden otoksen t-testissä asetetaan seuraavat hypoteesit:

H0: $\mu = \mu_0$, eli perusjoukossa keskiarvo on sama teoreettinen arvo, johon halutaan verrata

H1: $\mu \neq \mu_0$, eli perusjoukossa keskiarvo on erisuuri kuin teoreettinen arvo, johon halutaan verrata

Perusoppikirjoissa esitetään usein myös ns. yksisuuntaisen testauksen hypoteesit, joissa otetaan lisäksi kantaa keskiarvojen eron suuntaan, mutta nämä sivuutetaan, koska ne ovat psykologiassa harvinaisempia.

Mikäli voidaan olettaa, että otos on peräisin normaalijakaumaa noudattavasta perusjoukosta, havainnot ovat toisistaan riippumattomia ja kiinnostuksen kohteena oleva muuttuja on mitattu vähintään välimatka-asteikolla, noudattaa testisuure t-jakaumaa vapausasteilla $N - 1$.

Esimerkki: tutkija on kerännyt pienen aineiston tietyllä älykkyysmittarilla. Aineisto on esitetty alla olevassa taulukossa, ja taulukkoon on laskettu valmiiksi havaintoarvojen summa ja havaintoarvojen toiset potenssit ja niiden summa. Tutkijaa kiinnostaa, voidaanko tämän otoksen perusteella väittää, että siinä populaatiossa, jonka perusjoukosta tämä otos on kerätty, keskiarvo poikkeaa arvosta 100.

	1	2	3	4	5	6	7	8	9	10	summa
x	120	110	100	95	130	105	90	95	135	125	1105
x^2	14400	12100	10000	9025	16900	11025	8100	9025	18225	15625	124425

Keskihajonta voidaan laskea annettujen arvojen perusteella seuraavasti:

$s = \sqrt{\frac{124425 - \frac{1105^2}{10}}{10-1}} = 16{,}06$ ja $\bar{x} = \frac{1105}{10} = 110{,}5$

Joten testisuure on

$$t = \frac{110{,}5 - 100}{16{,}06 / \sqrt{10}} \approx 2{,}07$$

Koska testisuure noudattaa t-jakaumaa (oletetaan tässä, että analyysin oletukset ovat voimassa) vapausasteilla 10 − 1 = 9. T-jakauman kriittinen arvo on 5 %:n riskitasolla näillä vapausasteilla likimäärin 2.26, joten koska havaittu p-arvo on tätä kriittistä arvoa pienempi, jää nollahypoteesi voimaan 5 %:n riskitasolla.

Luottamusväli voidaan johtaa suoraan yhden otoksen t-testin kaavasta. Luottamusväli keskiarvolle on väli, jolla keskiarvon ajatellaan sijaitsevan perusjoukossa tietyllä luottamustasolla (yleensä 95 %, 99 % tai 99,9 %). Luottamusvälin ylä- ja alaraja voidaan laskea kaavalla.

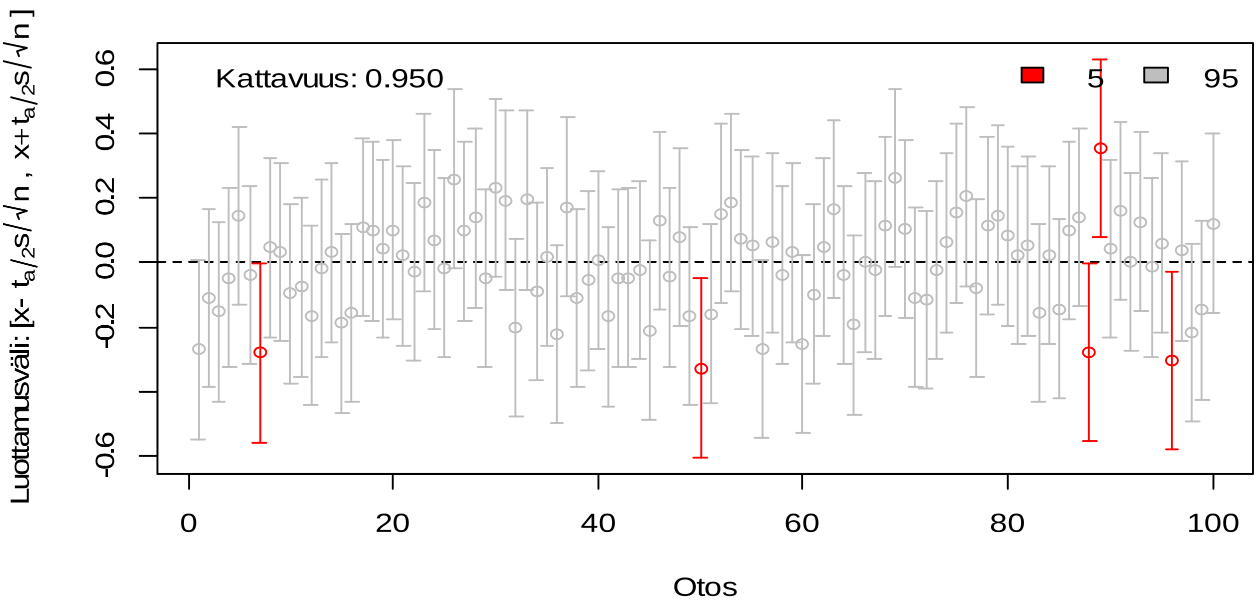
$$Alaraja = \bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \; Yläraja = \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Kaavoissa $t_{\alpha/2}$on tietyn luottamustason kriittinen arvo, jonka voi katsoa t-jakauman taulukosta (vapausasteet n − 1).

Esimerkki: erään tutkimuksen perusteella aineistosta havaittu keskiarvo on 10 ja keskihajonta 2. Otoskoko tutkimuksessa oli 100. Keskiarvon 95 %:n luottamusväli on

$10 \pm 2{,}26 \cdot \frac{2}{\sqrt{100}}$, eli alaraja on 9,54 ja yläraja on 10,45.

Luottamusvälin tulkinta on vastaava kuin perinteisissä tilastollisissa testeissä. Jos voidaan ajatella, että otetaan 100 kpl otoksia samasta perusjoukosta ja samalla otoskoolla, niin odotusarvoisesti perusjoukon keskiarvo sisältyy luottamusväliin todennäköisyydellä, joka on valittu luottamustaso. Alla olevassa kuvassa on esitetty esimerkki tilanteesta, jossa perusjoukosta, jossa keskiarvo on nolla (0), on otettu 100 kappaletta satunnaisotoksia. Jokaiselle satunnaisotokselle on laskettu erikseen 95 %:n luottamusväli. Kuvasta havaitaan, että 95 %:ssa otoksista luottamusväli ylittää perusjoukon todellisen keskiarvon. Kuvaan on punaisella merkitty luottamusvälit, jotka eivät ylitä todellista keskiarvoa.



Kahden riippumattoman otoksen t-testi

Kahden riippumattoman otoksen t-testissä tutkitaan, voidaanko aineiston perusteella olettaa, että kahden ryhmän keskiarvot poikkeavat toisistaan perusjoukossa. Testissä tehdään seuraavat oletukset:

1) molemmat otokset ovat peräisin normaalisti jakautuneesta perusjoukosta

2) havainnot ovat toisistaan riippumattomia

3) vertailtava muuttuja on mitattu vähintään välimatka-asteikolla

4) molempien ryhmien (otosten) keskihajonnat perusjoukossa ovat yhtä suuret

Testissä asetetaan tyypillisesti seuraavat hypoteesit:

$H0\colon \mu_1 = \mu_2$

$H1\colon \mu_1 \neq \mu_2$

Eli nollahypoteesina on: ryhmien (otosten) keskiarvot eivät poikkea toisistaan perusjoukossa ja vastahypoteesina: keskiarvojen välillä on eroa perusjoukossa.

t-testisuureen kaava on:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ jossa } s^2 = \frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{(n_1-1)+(n_2-1)} \text{ ja siis } s = \sqrt{\frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{(n_1-1)+(n_2-1)}}$$

Testisuure noudattaa oletusten vallitessa t-jakaumaa vapausasteilla $n_1 + n_2 - 2$.

Alla olevassa esimerkissä on esitetty pieni aineisto, jossa on kaksi ryhmää. Taulukkoon on laskettu valmiiksi ryhmien keskiarvot ja keskihajonnat. Molemmissa ryhmissä on 10 havaintoa, eli kokonais-N on 20.

	Havaintoarvot										**Keskiarvo**	**Keskihajonta**
Ryhmä 1	120	110	100	95	130	105	90	95	135	125	110,50	16,06
Ryhmä 2	105	95	95	100	110	100	92	95	110	120	102,20	8,92

$$s^2 = \sqrt{\frac{(10-1)\cdot 16.06 + (10-1)\cdot 8.92}{(10-1)+(10-1)}} \approx 3{,}53 \text{ Jolloin } t = \frac{110.50 - 102.20}{3{,}53 \cdot \sqrt{\left(\frac{1}{10} + \frac{1}{10}\right)}} \approx 5{,}258$$

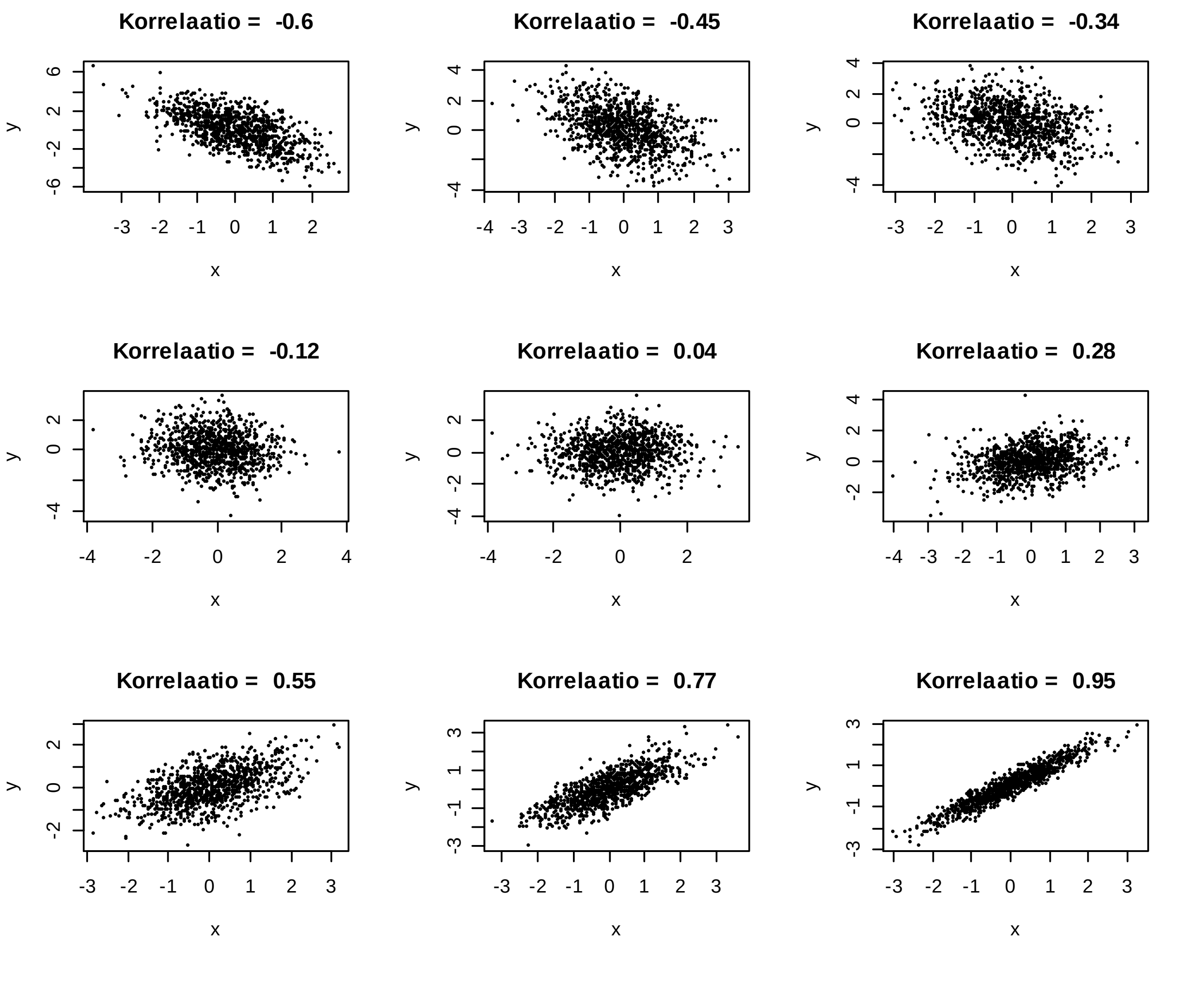
Oletusten ollessa voimassa noudattaa tämä testisuure t-jakaumaa vapausasteilla 10 + 10 − 2 = 18, jonka kriittinen arvo t-jakaumataulukosta katsoen on likimäärin 2,10. Tulos on siis tilastollisesti merkitsevä 5 %:n riskitasolla. Itseasiassa tulos on tilastollisesti merkitsevä myös 0,1 %:n riskitasolla, koska tämän riskitason kriittinen arvo on likimäärin 3.92. Tämän perusteella nollahypoteesi: ryhmien keskiarvot perusjoukossa ovat yhtä suuret, hylätään.

Korrelaatio

Korrelaatio on todennäköisyyslaskennassa ja tilastotieteessä käytetty käsite, joka kuvaa kahden muuttujan välistä lineaarista riippuvuutta. Korrelaatiokerroin tarkoittaa aineistosta laskettua havaintojen välistä korrelaatiota. Tarkkaan ottaen se on numeerinen mitta satunnaismuuttujien väliselle lineaariselle riippuvuudelle. Riippumattomien muuttujien välillä ei ole korrelaatiota.

Korrelaatiokerroin ei riipu käytetyistä yksiköistä. Mitä enemmän korrelaatiokerroin poikkeaa nollasta, sitä voimakkaampaa muuttujien välinen riippuvuus on. Arvo 1 tarkoittaa, että muuttujien välillä on täydellinen lineaarinen riippuvuus (−1 tarkoittaa täydellistä negatiivista lineaarista riippuvuutta), ts. toisen muuttujan voi laskea tarkasti lineaarisesti toisen arvosta.

Korrelaatio voidaan laskea usealla eri tavalla muuttujien mitta-asteikosta ja käyttötarkoituksesta riippuen. Tavallisesti korrelaatiolla tarkoitetaan Pearsonin tulomomenttikorrelaatiokerrointa. Nimestä huolimatta sen esitti ensimmäisenä Francis Galton. Jos tarkasteltavat muuttujat on mitattu vain järjestysasteikolla, niin silloin on parempi käyttää Spearmanin järjestyskorrelaatiokerrointa. Sanalla korrelaatiokerroin (joskus vain korrelaatio) tarkoitetaan yleensä Pearsonin korrelaatiokerrointa. Alla olevassa kuvassa on havainnollistettu kahden muuttujan välistä yhteyttä hajontakuvaajilla. Kuviin on merkitty myös muuttujien välille laskettu Pearsonin tulomomenttikorrelaatiokertoimen arvo.



Pearsonin tulomomenttikorrelaatiokerroin

Kun aineistossa on n kpl havaintoja, voidaan Pearsonin tulomomenttikorrelaatiokerroin laskea kaavalla:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

jossa $\bar{x}$ ja $\bar{y}$ ovat satunnaismuuttujien x ja y aritmeettisia keskiarvoja. Käsin laskettaessa on usein helpompi käyttää seuraavaa kaavaa, joka johtaa samaan lopputulokseen:

$$r = \frac{\sum_{i=1}^{n} xy - n\bar{x}\bar{y}}{\sqrt{\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right]}}$$

Esimerkki:

Tutkitaan kahden muuttujan x ja y välistä yhteyttä. Muuttujien arvot on esitetty alla olevassa taulukossa, johon on laskettu valmiiksi myös käsin laskua varten tarvittavat apumuuttujat.

	x	y	xy	x^2	y^2
1	10	270	2700	100	72900
2	14	320	4480	196	102400
3	21	430	9030	441	184900
4	30	530	15900	900	280900
5	28	530	14840	784	280900
summa	103	2080	46950	2421	922000

Lisäksi lasketaan muuttujien aritmeettiset keskiarvot, jotka ovat järjestyksessä muuttujille x ja y 20,6 ja 416. Sijoittamalla nämä keskiarvot ja taulukossa esitetyt luvut käsinlaskukaavaan saadaan.

$$r = \frac{46950 - 5 \cdot 20{,}6 \cdot 416}{\sqrt{(2421 - 5 \cdot 20{,}6^2) \cdot (922000 - 5 \cdot 416^2)}} = \frac{4102}{\sqrt{16970624}} \approx 0{,}996$$

Pearsonin tulomomenttikorrelaatiokerroin voidaan laskea edellä esitettyjen kaavojen lisäksi hyvin usealla eri tavalla. Alla on esitetty eri vaihtoehtoja, ja se minkä kaavan käyttö on milloinkin helpointa, riippuu siitä mitä tunnuslukuja on helposti saatavilla esimerkiksi tehtävänannon yhteydessä.

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

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

Hajontojen ja kovarianssin avulla korrelaatiokertoimen voi laskea kaavalla

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

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

Kovarianssikerroin, lyhyemmin kovarianssi, 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.

Joskus korrelaatiokertoimen voi laskea helpoiten ’käsin’ seuraavien kaavojen avulla.

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

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

Spearmanin järjestyskorrelaatiokerroin

Spearmanin järjestyskorrelaatiokertoimessa muuttujien arvot korvataan järjestysluvuilla (1, 2, 3, 4, … jne.) ja lasketaan näiden järjestyslukujen välille normaali Pearsonin tulomomenttikorrelaatiokerroin. Kyseessä on siis pikemminkin muuttujiin tehtävä järjestyslukumuunnos kuin varsinainen oma itsenäinen korrelaatiokerroin. Mikäli kaksi havaintoa saavat saman järjestysluvun, niin järjestysluku korvataan kahden järjestysluvun keskiarvolla. Tyypillisesti, mikäli muuttujien arvot ovat suoraan järjestyslukuja tai ne on muunnettu järjestysluvuiksi, lasketaan Spearmanin järjestyskorrelaatiokerroin kaavalla:

$$r = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n^3 - n}$$

Jossa d_i on toisiaan vastaavien muuttujien x ja y järjestyslukujen erotus. Seuraavaksi todistetaan, että Pearsonin ja Spearmanin korrelaatiokertoimien kaavat ovat identtiset, mikäli havaintoarvot on muunnettu järjestysluvuiksi. Todistaminen vaatii perustietoa aritmeettisten lukujonojen summauksesta, ja näitä tietoja ei vaadita valintakokeessa, vaikka ne on tässä esitetty.

Tarkastellaan korrelaatiokertoimen kaavaa, joka on muotoa (kaavoissa summaus tehdään yli kaikkien havaintojen, vaikka summauksen indeksejä ei ole erikseen kirjoitettu):

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}$$

Kaavassa kohta $\sum(x_i - \bar{x})^2$ voidaan kirjoittaa muotoon

$$\sum(x_i - \bar{x})^2 = \sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

sekä x:n että y:n suhteen. Jos perinteiset muuttujan arvot on korvattu järjestysluvuilla (i = 1,….., n), voidaan yllä oleva kaava kirjoittaa muodossa (tässä hyödynnetään lukujonojen summan ehtoja, jotka on esitetty lukion pitkässä matematiikassa ja voidaan osoittaa esim. induktiolla. Tarvittaessa näihin lukujonoja koskeviin peruskaavoihin voi tutustua esim. Wikipediasta (http://fi.wikipedia.org/wiki/Summa):

$$\sum i^2 - \frac{(\sum i)^2}{n} = \frac{n(n+1)(2n+1)}{n} - \frac{\left(\frac{n(n+1)}{2}\right)^2}{n} = \frac{n^3 - n}{12}$$

Nyt siis korrelaatiokertoimen kaavan nimittäjässä olevassa tulossa olevat molemmat termit voidaan korvata yllä esitetyllä kaavalla.

Seuraavaksi kirjoitetaan korrelaatiokertoimen kaavan osoittaja hieman toiseen muotoon. Muodostetaan ensin apumuuttuja e, joka on muotoa $e_i = (x_i - \bar{x}) - (y_i - \bar{y})$

Muodostetaan seuraavaksi e:n toinen potenssi ja summataan se yli kaikkien havaintojen, jolloin saadaan:

$$\sum e_i^2 = \sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2 - 2\sum(x_i - \bar{x})(y_i - \bar{y})$$

Ratkaistaan edellä olevasta kaavasta tulosumma $\sum(x_i - \bar{x})(y_i - \bar{y})$, joka on korrelaatiokertoimen kaavassa osoittajana

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{2}\left(\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2 - \sum e_i^2\right)$$

Kun edellä olevassa lausekkeessa korvataan alkuperäiset arvot järjestysluvuilla, on $e_i = i - j = d_i$ eli e on muuttujien alkuperäisten (keskistettyjen) arvojen järjestyslukujen erotus ja vastaavasti

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{2}\cdot\left(\frac{n^3 - n}{12} + \frac{n^3 - n}{12} - \sum d_i^2\right)$$

Yhdistämällä edellinen tulos aikaisemmin saatuihin tuloksiin korrelaatiokertoimen kaavassa saadaan:

$$\frac{\frac{1}{2}\cdot\left(\frac{n^3-n}{12} + \frac{n^3-n}{12} - \sum d_i^2\right)}{\sqrt{\left(\frac{n^3-n}{12}\right)\left(\frac{n^3-n}{12}\right)}} = \frac{2\cdot\frac{n^3-n}{12} - \sum d_i^2}{2\cdot\frac{n^3-n}{12}} = 1 - \frac{6\cdot\sum d_i^2}{n^3-n},$$

joka on Spearmanin järjestyskorrelaatiokertoimen kaava, mikä oli todistettava.

Esimerkki: Tarkastellaan aineistoa, jossa on kymmenen havaintoa. Luvut on järjestyslukumuunnoksen havainnollistamiseksi järjestetty muuttujan x suhteen kasvavaan järjestykseen. Mikäli kahdella luvulla on alkuperäisissä muuttujissa sama arvo, on järjestysluvuksi annettu vastaavien järjestyslukujen keskiarvo (esim. muuttujassa x on kaksi kappaletta havaintoarvoja 120, jotka ovat järjestyksessä kahdeksas ja yhdeksäs luku, jolloin näiden havaintoarvojen järjestyslukujen keskiarvo on 8,5).

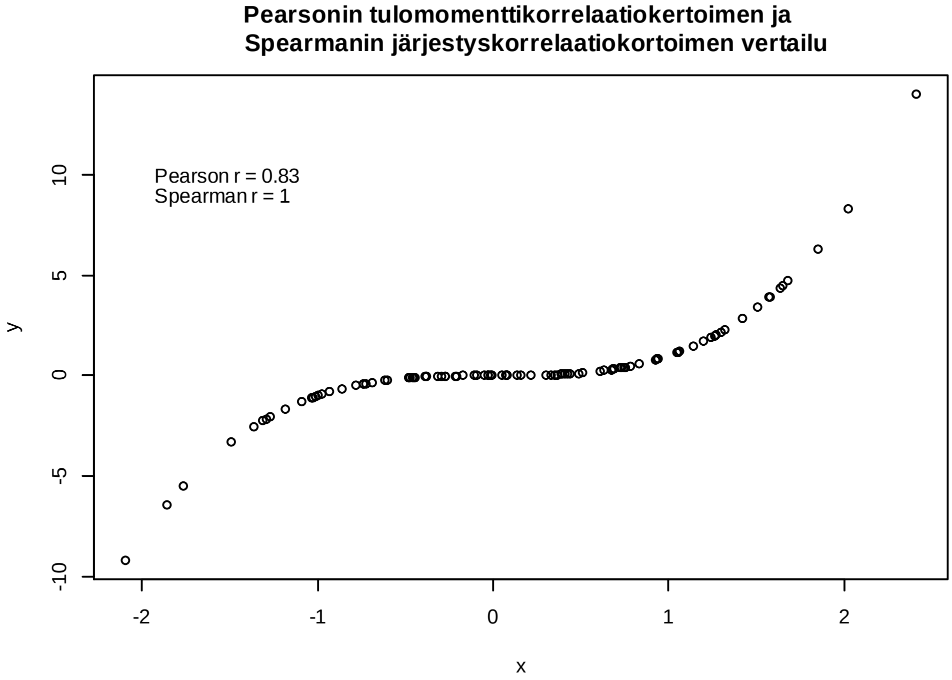
	x	y	Järjestys x	Järjestys y	Järjestyslukujen erotus (d)	d^2
1	70	17	1	2	-1	1
2	80	16	2	1	1	1
3	90	20	3.5	4.5	-1	1
4	90	19	3.5	3	0.5	0.25
5	100	20	5.5	4.5	1	1
6	100	21	5.5	6	-0.5	0.25
7	110	22	7	7.5	-0.5	0.25
8	120	22	8.5	7.5	1	1
9	120	24	8.5	10	-1.5	2.25
10	140	23	10	9	1	1
summa	1020	204	55	55	0	9

Sijoittamalla tarvittavat luvut Spearmanin järjestyskorrelaation kaavaan, saadaan järjestyskorrelaatioksi:

$$r = 1 - \frac{6 \cdot \sum d_i^2}{n^3 - n} = 1 - \frac{6 \cdot 9}{10^3 - 10} = \approx 0{,}945$$

Pearsonin tulomomenttikorrelaatiokertoimen ja Spearmanin järjestyskorrelaatiokertoimen vertailu

Koska Spearmanin järjestyskorrelaatiokerroin perustuu järjestyslukumuunnokseen, on se Pearsonin tulomomenttikorrelaatiokertoimeen verrattuna vähemmän herkkä poikkeamille muuttujien välisen yhteyden lineaarisuudesta tilanteissa, joissa muuttujien välinen yhteys on monotoninen, eli aidosti kasvava tai aidosti laskeva. Seuraavassa kuvassa on esitetty esimerkki, jossa muuttujien välinen yhteys on aidosti kasvava, eli järjestyksessä seuraava muuttujan arvo on aina suurempi kuin edellinen. Muuttujien välinen yhteys ei ole kuitenkaan lineaarinen. Tässä tilanteessa havaitaan, että Spearmanin järjestyskorrelaatiokerroin on 1, mutta Pearsonin tulomomenttikorrelaatiokerroin on selkeästi tätä pienempi.



Korrelaatiokertoimen tilastollisen merkitsevyyden testaus

Koska psykologisessa tutkimuksessa käytännössä katsoen aina käsitellään otosta, on tärkeää tilastollisen testauksen avulla pyrkiä tekemään päätelmiä siitä, poikkeaako otoksessa havaittu yhteys jossakin perusjoukossa arvosta 0.

Tilastollisessa hypoteesin testauksessa asetetaan tyypillisesti seuraavat hypoteesit.

$H_{0:}\, \rho = 0$

$H_{1:}\, \rho \neq 0$

Eli nollahypoteesina on: perusjoukossa perusjoukon korrelaatiokerroin ρ on yhtä suuri kuin nolla ja vastahypoteesina, että perusjoukossa korrelaatiokerroin poikkeaa arvosta nolla. Testisuure on t, joka voidaan laskea kaavalla:

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}},$$

jossa r on otoksesta laskettu korrelaatiokerroin, n otoskoko ja joka noudattaa t-jakaumaa vapausasteilla $n - 2$.

Esimerkki:

Aineistosta laskettu otoskorrelaatiokerroin on 0.5 ja otoksen koko on 102. Tällöin korrelaatiokertoimeen liittyvän testisuureen arvo on

$$t = \frac{0{,}5 \cdot \sqrt{100-2}}{\sqrt{1-r0{,}5^2}} = \frac{0{,}5 \cdot 10}{\sqrt{1-0.25}} = \frac{5}{\sqrt{0{,}75}} \approx 5{,}77$$

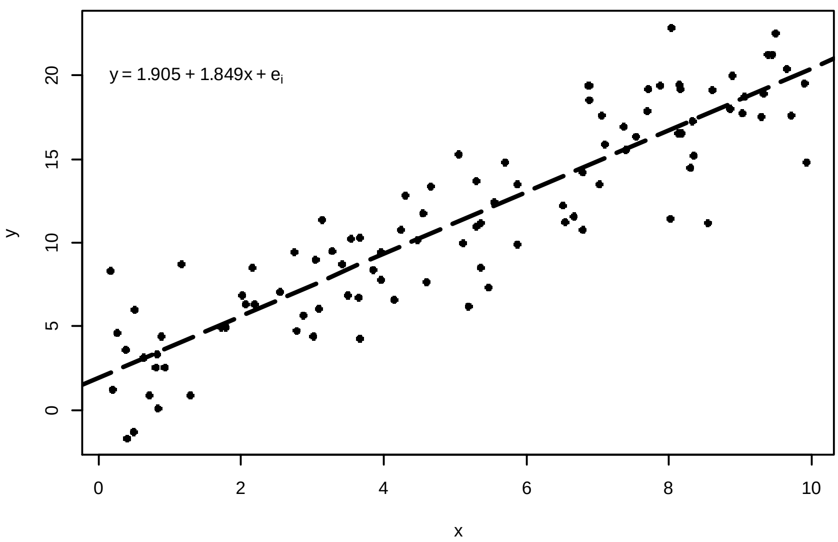
Testisuure noudattaa t-jakaumaa vapausasteilla $102 - 2 = 100$. T-jakauman kertymäfunktion taulukosta havaitaan, että näillä vapausasteilla 5 % merkitsevyystason kriittinen arvo on likimäärin 1.98 (huom: testi on kaksisuuntainen, eli jakauman molempiin häntiin jää 2.5 % jakauman pinta-alasta). Koska havaittu testisuureen arvo on suurempi kuin kriittinen arvo, on korrelaatiokerroin tilastollisesti merkitsevä 5 %:n riskitasolla.

Regressioanalyysi

Regressioanalyysillä tutkitaan yhden tai useamman ennustavan, eli riippumattoman muuttujan yhteyttä ennustettavaan eli riippuvaan muuttujaan. Regressioanalyysi pyrkii siis selittämään jonkin ennustettavan muuttujan havaittujen arvojen vaihtelun joidenkin ennustavien muuttujien havaittujen arvojen avulla. Regressioanalyysissä pyritään estimoimaan ns. regressiosuoran yhtälö, joka on yhden selittäjän regressioanalyysin tapauksessa muotoa:

$$y = b_0 + b_1 x + e_i$$

Kaavassa selittävä muuttuja on merkitty x:llä, selitettävä y:llä ja e_i on ennusteeseen liittyvä ennustevirhe, joka oletetaan normaalijakautuneeksi odotusarvolla 0 ja tietyllä vakiovarianssilla, joka estimoidaan aineistosta. Alla olevassa kuvassa on esitetty hajontakuvaaja x:n ja y:n välillä eräässä otoksessa ja aineiston perusteella laskettu regressiosuora (katkoviiva) ja sen yhtälö



Tavallisesti regressiokertoimet ratkaistaan havaitusta aineistosta pienimmän neliösumman menetelmällä, jossa etsitään ne arvot b_0 ja b_1, jotka minimoivat yhtälön $\sum_{i=1}^{n}(y - b_0 + b_1 x)^2$. Voidaan osoittaa, että edellä mainittu yhtälö saavuttaa pienimmän arvonsa, kun parametrit, eli regressiokertoimet b_0 ja b_1 saavat arvot

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Nämä regressiokertoimet voidaan tulkita siten, että b_0 on piste, jossa regressiosuora leikkaa y-akselin, eli ennustettu arvo, kun ennustajan x arvo on 0, ja b_1 on suoran kulmakerroin, eli kun x ”muuttuu” yhden yksikön verran, niin ennustettu y:n arvo kasvaa kulmakertoimen verran. Vaihtoehtoisesti regressiosuoran kulmakerroin $b_1$voidaan laskea myös kaavalla $b_1 = \frac{S_{xy}}{S_x}$, jossa $S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$, eli muuttujien x ja y välinen kovarianssi ja $S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$, eli muuttujan x keskihajonta. Mikäli tunnetaan muuttujien välinen korrelaatiokerroin r, voidaan regressiosuoran kulmakerroin laskea myös kaavalla $b_1 = r\frac{S_y}{S_x}$, jossa S_y on muuttujan y keskihajonta. Toisaalta mikäli molemmat muuttujat on standardoitu siten, että niiden keskiarvo on 0 ja hajonta on 1, niin regressiokerroin on sama kuin muuttujien välinen korrelaatio ja regressiosuoran vakiotermi saa aina arvon 0. Tällöin regressiokerrointa kutsutaan myös standardoiduksi regressiokertoimeksi.

Esimerkki: Tutkitaan pientä aineistoa, jossa otoskoko on 10 havaintoa. Alla olevassa taulukossa on esitetty havaintoarvot, tarvittavat apumuuttujat ja lisäksi on laskettu muuttujien summat ja tarvittavat keskiarvot.

	x	y	x2	xy
1	4	6	16	24
2	5	7	25	35
3	3	3	9	9
4	1	0	1	0
5	5	4	25	20
6	6	5	36	30
7	7	9	49	63
8	5	10	25	50
9	3	7	9	21
10	1	4	1	4
summa	40	55	196	256
Keskiarvo	4	5,5		

Aikaisemmin esitettyjen yhtälöjen perusteella regressiosuoran parametrit, eli kulmakerroin ja vakio ovat siis.

$b_1 = \frac{10 \cdot 256-40\cdot 55}{10\cdot 196-40^2} = 1$ ja $b_0 = 5{,}5 - 1 \cdot 4 = 1{,}5$

Regressiokertoimen tilastollisen merkitsevyyden testaus

Koska tyypillisesti psykologiassa käsitellään otosta, on kiinnostuksen kohteena luonnollisesti hypoteesintestaus, jonka avulla pyritään tekemään päätelmiä muuttujien välisestä yhteydestä varsinaisessa tutkittavassa perusjoukossa. Tilastollisen merkitsevyyden testausta varten asetetaan seuraavat hypoteesit:

$H_0\colon \beta = 0$ ja $H_1\colon \beta \neq 0$. Eli nollahypoteesina on, että perusjoukossa tietty regressioparametri on yhtä suuri kuin nolla ja vastahypoteesina, että tietty regressioparametri poikkeaa perusjoukossa nollasta.

Tätä hypoteesintestausta varten tarvitaan tieto regressiokertoimeen liittyvästä keskivirheestä. Regressiokertoimen vakiotermiin liittyvät keskivirheet voidaan lausua kaavoilla.

$$SE_{b0} = \sqrt{S_e^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]}$$

Ja

$$SE_{b1} = \sqrt{\frac{S_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

Kaavoissa S_e^2 on ennustevirheiden, eli havaitun arvon ja ennustetun arvon välisen erotuksen varianssi, joka lasketaan kaavalla $S_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2}$. Kaavassa e_i on tiettyyn havaintoon liittyvä ennustevirhe eli havaitun arvon ja ennustetun arvon erotus. On huomattava, että myös jakajassa on perinteiseen otosvarianssiin verrattuna jakajana $n - 2$, koska regressioennusteeseen liittyy kaksi parametria.

Tarkastellaan aikaisemmin käytettyä esimerkkiä. Pienimmän neliösumman estimoinnin perusteella tiedämme, että regressiokertoimet ovat $b_0 = 1{,}5$ ja $b_1 = 1$. Alla olevassa taulukossa on laskettu ennustettu arvo jokaiselle havainnolle kaavalla $\hat{y} = 1{,}5 + 1 \cdot x$. Lisäksi taulukkoon on laskettu ennustevirheet, eli residuaalit sekä niiden toinen potenssi.

	x	y	$(x_i - \bar{x})^2$	ennuste	residuaali, e	e^2
1	4	6	0	5.5	0.5	0.25
2	5	7	1	6.5	0.5	0.25
3	3	3	1	4.5	-1.5	2.25
4	1	0	9	2.5	-2.5	6.25
5	5	4	1	6.5	-2.5	6.25
6	6	5	4	7.5	-2.5	6.25
7	7	9	9	8.5	0.5	0.25
8	5	10	1	6.5	3.5	12.25
9	3	7	1	4.5	2.5	6.25
10	1	4	9	2.5	1.5	2.25
summa	40	55	36	55	0	42.5
Keskiarvo	4	5.5	3.6	5.5	0	4.25

Residuaalien otosvarianssi on $S_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{42{,}5}{10-2} \approx 5{,}31$, joten vakiotermin b_0 keskivirhe on:

$SE_{b0} = \sqrt{5{,}31\left[\frac{1}{10} + \frac{4^2}{36}\right]} \approx 1{,}70$ ja vastaavasti regressiokertoimen $b_1$keskivirhe on $SE_{b1} = \sqrt{\frac{5{,}31}{36}} \approx 0{,}38$

Tilastollisen merkitsevyys testauksen testisuure on aina muotoa $t = \frac{b}{SE}$ joka noudattaa t-jakaumaa vapausasteilla $n - 2$, eli regressiokertoimien testisuureet ovat $t_{b0} = \frac{1{,}5}{1{,}7} \approx 0{,}88$ ja $t_{b1} = \frac{1}{0{,}38} \approx 2{,}63$. t-jakauman 5 % merkitsevyystason kriittinen arvo vapausasteilla 8 on 2,306, eli regressiokerroin b_1 on tilastollisesti merkitsevä 5 % riskitasolla, mutta vakiotermi ei ole tilastollisesti merkitsevä. Tyypillisesti vakiotermin tilastollinen merkitsevyys ei kuitenkaan ole erityisen kiinnostavaa, vaan tutkimuksen kannalta on oleellista tehdä päätelmiä muuttujien välisestä yhteydestä, eli varsinaisesta regressiokertoimesta.

Regressiomallin sopivuuden arvioiminen ja malliin liittyvien oletusten tarkasteleminen on ehdottoman tärkeää jokaisen analyysin yhteydessä, mutta nämä tarkastelut ohitetaan tässä monisteessa ja varsinaisessa valintakokeessa voi olettaa oletusten olevan voimassa. Oletuksista tässä monisteessa on mainittu ainoastaan ennustevirheiden normaalisuus, ja muita oletuksia ovat virhevarianssin homoskedastisuus, eli voidaan olettaa, että jokaisen henkilön kohdalla ennustevirheen jakauma noudattaa normaalijakaumaa vakiovirhevarianssilla. Lisäksi oletetaan, että havainnot ovat riippumattomia toisistaan. Nämä oletukset ovat keskeisiä regressioanalyysiin liittyvän hypoteesintestauksen yhteydessä. Varsinaiseen regressiosuoran sovitukseen ei liity oletuksia, eli regressiokertoimet voidaan laskea pienimmän neliösumman menetelmällä mistä tahansa aineistosta.

Usean selittäjän regressio

Usean selittäjän regressioanalyysissä ennustavia tekijöitä on kaksi tai useampia. Regressiosuoran yhtälö on tilanteessa, jossa selittäviä tekijöitä on k kappaletta

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k + e_i$$

Usean selittäjän regressiomallien käsin laskeminen vaati jo matriisilaskennan perusteiden tuntemusta, mutta kahdella ennustajalla analyysi on vielä laskettavissa suhteellisen helposti. Kahden selittäjän regressiomallin tapauksessa regressiosuoran yhtälö on

$$y = b_0 + b_1x_1 + b_2x_2 + e_i$$

Eli estimoitavia regressioparametrejä on kolme kappaletta: vakio b_0, muuttujan x_1 regressiokerroin b_1 ja muuttujan x_2 regressiokerroin b_2.

Pienimmän neliösumman menetelmällä voidaan ratkaista parametrien arvot. Kaavoissa käytetään hieman erilaista merkintätapaa, sillä on erittäin tärkeää huomata, että kaavoissa kaikki havaintoarvot on ensin keskistetty, eli jokaisesta havaintoarvosta on vähennetty muuttujan keskiarvo. Kaavoissa keskistettyjä muuttujia merkitään pienellä kirjaimella ja keskistämättömiä isoilla kirjaimilla.

$$b_1 = \frac{(\sum x_2^2)\,(\sum x_1y) - (\sum x_1x_2)(\sum x_2y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$

$$b_2 = \frac{(\sum x_1^2)\,(\sum x_2y) - (\sum x_1x_2)(\sum x_1y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$

Vastaavasti suorana yleistyksenä yhden selittäjän tapauksesta voidaan johtaa vakiotermille b_0 kaava, joka on

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_1\bar{X}_2$$

Kaavoista huomataan, että niissä huomioidaan molempien selittävien muuttujien yhteisvaihtelu, eli regressiokertoimen tulkinta on tietyn muuttujan itsenäinen yhteys, kun toisen selittäjän selittämä vaihtelu on vakioitu. Seuraavassa esimerkissä tarkastelemme jälleen pientä aineistoa, josta on tarkoitus laskea kahden selittäjän regressioanalyysin tulos. Alla olevassa taulukossa on varsinaiset muuttujat x1, x2 ja y sekä regressiokertoimien laskemiseen tarvittavat apumuuttujat.

	Alkuperäiset muuttujat			Keskistetyt muuttujat			Apumuuttujat (muodostettu keskistetyistä muuttujista)				
	X_1	X_2	Y	X_1-4	X_2-5,5	Y-7	x_1^2	x_2^2	x_1*y	x_2*y	x_1*x2
1	4	6	8	0	0,5	1	0	0,25	0	0,5	0
2	5	7	7	1	1,5	0	1	2,25	0	0	1,5
3	3	3	4	-1	-2,5	-3	1	6,25	3	7,5	2,5
4	1	0	7	-3	-5,5	0	9	30,25	0	0	16,5
5	5	4	6	1	-1,5	-1	1	2,25	-1	1,5	-1,5
6	6	5	9	2	-0,5	2	4	0,25	4	-1	-1
7	7	9	10	3	3,5	3	9	12,25	9	10,5	10,5
8	5	10	9	1	4,5	2	1	20,25	2	9	4,5
9	3	7	6	-1	1,5	-1	1	2,25	1	-1,5	-1,5
10	1	4	4	-3	-1,5	-3	9	2,25	9	4,5	4,5
Summa	40	55	70	0	0	0	36	78,5	27	31	36
Keskiarvo	4	5,5	7	0	0	0	3,6	7,85	2,7	3,1	3,6

Sijoittamalla kaavaan keskistettyjen muuttujiin liittyvät summat saadaan:

$b_1 = \frac{78{,}5 \cdot 27 - 36 \cdot 31}{36 \cdot 78{,}5 - 36^2} \approx 0{,}6559$ ja $b_2 = \frac{36 \cdot 31 - 36 \cdot 27}{36 \cdot 78{,}5 - 36^2} \approx 0{,}094$

ja

$b_0 = 7 - 0{,}6559 \cdot 4 - 0{,}094 \cdot 5{,}5 \approx 3{,}858$

Joissakin tilanteissa on helpompi laskea tarvittavien apumuuttujien summat suoraan ilman keskistystä hyödyntäen seuraavia kaavoja. Kaavoissa pienillä kirjaimilla kirjoitetut muuttujat on keskistetty ja isoilla kirjaimilla kirjoitetut ovat alkuperäisiä muuttujan arvoja.

$$\sum(x_1 y) = \sum(X_1 - \bar{X}_1)(Y - \bar{Y}) = \sum X_1 Y - \frac{\sum X_1 \sum Y}{N}$$

$$\sum(x_2 y) = \sum(X_2 - \bar{X}_2)(Y - \bar{Y}) = \sum X_2 Y - \frac{\sum X_2 \sum Y}{N}$$

$$\sum(x_1 x_2) = \sum(X_1 - \bar{X}_1)(X_2 - \bar{X}_2) = \sum X_2 Y - \frac{\sum X_2 \sum Y}{N}$$

Esimerkiksi aikaisemmin alla olevassa taulukossa on laskettu aikaisemman esimerkin muuttujien tulot suoraan alkuperäisistä havaintoarvoista.

	X1	X2	Y	X1*Y	X2*Y	X1*x1
1	4	6	8	32	48	24
2	5	7	7	35	49	35
3	3	3	4	12	12	9
4	1	0	7	7	0	0
5	5	4	6	30	24	20
6	6	5	9	54	45	30
7	7	9	10	70	90	63
8	5	10	9	45	90	50
9	3	7	6	18	42	21
10	1	4	4	4	16	4
Summa	40	55	70	307	416	256
Keskiarvo	4	5.5	7	30.7	41.6	25.6

Tällöin

$$\sum(x_1 y) = \sum(X_1 - \bar{X}_1)(Y - \bar{Y}) = \sum X_1 Y - \frac{\sum X_1 \sum Y}{N} = 307 - \frac{40 \cdot 70}{10} = 27,$$

joka on sama tulos kuin alkuperäisen taulukon summan x_1*y arvo.

Regressiokertoimen tilastollisen merkitsevyyden testaus kahden selittäjän regressioanalyysissä

Kahden selittäjän regressioanalyysin regressiokertoimien tilastollisen merkitsevyyden testaus poikkeaa jonkin verran yhden selittäjän regressioanalyysistä. Regressiokertoimien keskivirheet voidaan laskea seuraavilla kaavoilla. Vakiotermin keskivirheen laskeminen sivuutetaan, koska vakion merkitsevyyden testaaminen ei ole tyypillisesti kovinkaan mielenkiintoista.

$$SE_{b1} = \sqrt{\frac{S_{y.12}^2}{\sum[(X_1 - \bar{X}_1)^2(1 - r_{12}^2)]}}$$

$$SE_{b2} = \sqrt{\frac{S_{y.12}^2}{\sum[(X_2 - \bar{X}_2)^2(1 - r_{12}^2)]}}$$

Kaavoissa r_{12} on selittävien muuttujien välinen korrelaatio ja $S_{y.12}^2$ ennustevirheiden eli residuaalien varianssi, joka lasketaan kaavalla

$$S_{y.12}^2 = \frac{\sum e^2}{n - 3}$$

Esimerkiksi jälleen aikaisemman esimerkin arvoja hyödyntäen on laskettu ennustetut arvot ja residuaalit

	X1	X2	Y	Ennuste	e	e^2	(X1-X)	(X1-X)^2	(X1-X)^2*(1-0.667^2)
1	4	6	8	7,0471	0,95294	0,908	0	0	0
2	5	7	7	7,7971	-0,7971	0,635	1	1	0,555111
3	3	3	4	6,1088	-2,1088	4,447	-1	1	0,555111
4	1	0	7	4,5147	2,4853	6,177	-3	9	4,995999
5	5	4	6	7,5147	-1,5147	2,294	1	1	0,555111
6	6	5	9	8,2647	0,7353	0,541	2	4	2,220444
7	7	9	10	9,2971	0,70294	0,494	3	9	4,995999
8	5	10	9	8,0794	0,92058	0,847	1	1	0,555111
9	3	7	6	6,4853	-0,4853	0,236	-1	1	0,555111
10	1	4	4	4,8912	-0,8912	0,794	-3	9	4,995999
Summa	40	55	70	70	0	17,37	0	36	19,983996
Keskiarvo	4	5,5	7	7	0	1,737	0	3,6	1,9983996

Jolloin $S_{y.12}^2 = \frac{17{,}37}{7} \approx 2{,}48$, muuttujien X1 ja X2 välinen korrelaatio (laskutoimitusta ei esitetty) on 0,667 ja tätä tietoa hyödyntäen on laskettu yllä olevan taulukon viimeisen sarakkeen arvot. Keskivirhe regressiokertoimelle b_1 on siis

$$SE_{b1} = \sqrt{\frac{S_{y.12}^2}{\sum[(X_1 - \bar{X}_1)^2(1 - r_{12}^2)]}} = \sqrt{\frac{2{,}48}{19{,}98}} \approx 0{,}3524$$

Varsinainen testisuure t lasketaan kahden selittäjän tapauksessa samalla tavalla kuin yhden selittäjän regressioanalyysissä, eli jaetaan regressiokerroin vastaavalla keskivirheellä. Tämä testisuure noudattaa kahden selittäjän regressioanalyysin kohdalla t-jakaumaa vapausasteilla $n - 3$.

Esimerkki: Aikaisempien laskutoimitusten perusteella tiedetään, että regressiokerroin $b_1 = 0{,}6559$ ja $SE_{b1} = 0{,}3524$, joten testisuure on

$$t = \frac{b_1}{SE_{b1}} = \frac{0{,}6559}{0{,}3524} \approx 1{,}86$$

Testisuure noudattaa t-jakaumaa vapausasteilla 7, jonka 5 % merkitsevyystason kriittinen arvo t-jakaumataulukon perusteella on likimäärin 2,36. Koska havaittu testisuure on kriittistä arvoa pienempi, ei regressiokerroin poikkea tilastollisesti merkitsevästi nollasta 5 %:n riskitasolla.

Standardoidut regressiokertoimet

Mikäli kaikki muuttujat on standardoitu siten, että niiden keskiarvo on 0 ja keskihajonta on 1, niin regressiokertoimet, joita kutsutaan myös standardoiduiksi regressiokertoimiksi, voidaan laskea suoraviivaisesti muuttujien välisten korrelaatioiden avulla. Tällöin

$$b_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}$$

$$b_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2}$$

Kaavoissa r_{y1} on muuttujan Y ja muuttujan X_1 välinen korrelaatio, r_{y2} on muuttujan Y ja muuttujan X_2 välinen korrelaatio ja r_{12} ennustavien muuttujien X_1 ja X_2 välinen korrelaatio.

Esimerkiksi jos muuttujien väliset korrelaatiot ovat alla esitettyjä

	Y	X_1	X_2
Y	1		
X_1	0.77	1	
X_2	0.72	0.68	1

niin

$$b_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2} = \frac{0{,}77 - 0{,}72 \cdot 0{,}68}{1 - 0{,}68^2} \approx 0{,}522$$

$$b_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} = \frac{0{,}72 - 0{,}77 \cdot 0{,}68}{1 - 0{,}68^2} \approx 0{,}365$$

Mallin kokonaisselitysosuus

Usein kiinnostuksen kohteena on, kuinka suuri osuus ennustettavan muuttujan vaihtelusta saadaan selitettyä ennustavilla muuttujilla. Kahden selittäjän regressioanalyysin tapauksessa tämä selitysosuus saadaan laskettua muuttujien välisten korrelaatioiden avulla seuraavasti. Selitysosuuden neliöjuurta kutsutaan myös mallin yhteiskorrelaatiokertoimeksi.

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}$$

Esimerkiksi jos muuttujien väliset korrelaatiot ovat

	Y	X_1	X_2
Y	1		
X_1	0.77	1	
X_2	0.72	0.68	1

jolloin selitysosuus mallissa, jossa muuttujaa Y ennustetaan muuttujilla X_1 ja X_2, on

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2} = \frac{0{,}77^2 \cdot 0{,}72^2 - 2 \cdot 0{,}77 \cdot 0{,}72 \cdot 0{,}68}{1 - 0{,}68^2} \approx 0{,}66$$

Eli muuttujan Y vaihtelusta saadaan selitettyä kahdella muuttujalla X_1 ja X_2 noin 66 %.

Tietokonetulosteiden tulkinta

Regressiomallin arvioimista ja mallin sopivuuteen liittyvää diagnostiikkaa ei tässä monisteessa käsitellä ja valintakokeessa voi olettaa mallin oletusten olevan voimassa. Tässä tuodaan kuitenkin esiin muutama näkökohta tietokonetulosteisiin, koska kahden ja useamman selittäjän regressioanalyysi tehdään tyypillisesti tietokoneella. Regressiomallin tilastollisen merkitsevyyden selvittäminen aloitetaan tutkimalla koko mallin tilastollista merkitsevyyttä selityskertoimen lisäksi. Mallin tilastollinen testaaminen tapahtuu F-testisuureen avulla. F-jakauma on jatkuva todennäköisyysjakauma, jolla on kaksi vapausastelukua, jotka vaikuttavat jakauman muotoon. F-jakauman käyttöä ei käydä läpi tarkemmin, ja valintakoetta ajatellen jakaumasta ei tarvitse tietää enempää. Regressioanalyysin yhteydessä F-testisuuretta käytetään testaamaan nollahypoteesia H_0: R = 0 eli mallin yhteiskorrelaatiokerroin on nolla. H_1-hypoteesi on puolestaan R > 0. Yhtä hyvin nollahypoteesin voi esittää muodossa $R^2 = 0$ eli mallin selitysaste on nolla. Jos tämä nollahypoteesi jää voimaan, niin silloin ei voida sanoa, että mallin selittävät muuttujat ennustaisivat tai selittäisivät selitettävää muuttujaa tilastollisesti merkitsevästi. Yhteiskorrelaatiokerroin on regressiomallin ennustearvojen $(\hat{y}_i)$ ja selitettävän muuttujan havaittujen arvojen (y_i) välinen korrelaatiokerroin.

Alle on koottu taulukkoja erilaisista malleista, joissa samaa y-muuttujaa ennustetaan hieman eri x-muuttujilla.

Malli	R	R^2	F	df_1	df_2	p-arvo
1	.878	.772	8.45	4	10	.003
2	.877	.769	12.19	3	11	.001
3	.869	.755	18.53	2	12	<.001

Huomataan, että kaikki mallit selittävät Y-muuttujaa tilastollisesti merkitsevästi (p-arvot ovat pieniä). Yhteiskorrelaatiokerroin ja selityskerroin ovat lähes yhtä suuria malleissa 1–3, mutta muuttujien jättäminen pois parantaa suhteutettua selityskerrointa. Kun tarkastellaan yhdessä edellä olevan taulukon kanssa mallien regressiokertoimia, voidaan malleja verrata tarkemmin. Seuraavan sivun taulukossa on esitetty esimerkin regressiokertoimet.

Malli	Selittäjät	Regressio-kerroin	t-arvo	p-arvo
1	Vakiotermi	72.858	18.82	<.001
	Henkilöä/lääkäri	0.0014	0.36	.727
	Imeväiskuolleisuus	-0.257	-1.59	.143
	Hedelmällisyysluku	2.150	0.81	.439
	BKT/henkilö (1000 USD)	0.064	0.95	.365
2	Vakiotermi	72.800	19.61	<.001
	Imeväiskuolleisuus	-0.213	-2.15	.055
	Hedelmällisyysluku	2.020	0.80	.443
	BKT/henkilö (1000 USD)	0.077	1.38	.194
3	Vakiotermi	75.114	33.07	<.001
	Imeväiskuolleisuus	-0.156	-2.29	.041
	BKT/henkilö (1000 USD)	0.098	2.04	.064

Mallissa 1 huomataan, että vaikka koko malli selitti elinajanodotetta tilastollisesti merkitsevästi, niin yhdenkään muuttujan p-arvo ei ole tilastollisesti merkitsevä. Tämä ristiriitaiselta tuntuva tulos johtuu siitä, että selittävillä muuttujilla on paljon ”päällekkäistä” selitysosuutta. Voidaan olettaa, että imeväiskuolleisuus korreloi voimakkaasti muiden selittävien muuttujien (erityisesti henkilöä/lääkäri –muuttujan) kanssa.

Mallissa 2 imeväiskuolleisuuden p-arvo lähestyy perinteistä tilastollisen merkitsevyyden rajaa (0,05). Hedelmällisyysluvun p-arvo on suuri (mitä se oli myös mallissa 1), joten sen mukana olo mallissa ei ole tilastollisen merkitsevyyden valossa kovin hyödyllistä. Mallissa 3 imeväiskuolleisuus on tilastollisesti merkitsevä ja BKT:kin lähestyy perinteistä tilastollisen merkitsevyyden rajaa (0,05). Jos BKT jätettäisiin pois ja selitettäisiin elinajanodotetta ainoastaan imeväiskuolleisuudella, niin mallin selityskerroin olisi $(-0{,}819)^2 \approx 0.67$, mikä on selvästi pienempi kuin malleissa 1–3. Tämän esimerkin tapauksessa olisi järkevää pitää myös BKT selittäjänä. Todettakoon kuitenkin, että tässä esimerkissä aineisto on kovin pieni, mikä vaikuttaa tuloksiin.

Regressiokertoimet sisältävän taulukon merkinnöistä on hyvä huomata, että on oikeampaa ilmoittaa p-arvo muodossa $p < .001$, vaikka tietokoneohjelmien tulostus olisikin esimerkiksi .00.

Vanhoja valintakoetehtäviä

Tässä osassa on joitakin vanhoja valintakoetehtäviä. Ratkaisut niihin löytyvät materiaalin 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 C°.

Aikaisemmissa tutkimuksissa oli todettu, että kylmemmällä säällä useammat käyttävät päähinettä, mikä kuulostaakin ihan luontevalta. Jos lasket päähineen käyttäjien lukumäärät 0 C° ja -20 C° 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 ratkaisuissa 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

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

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

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 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 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 tulon. Painokertoimien tuloa sanotaan kokonaispainokertoimeksi. Siis jos valitaan kohteet 1, 4 ja 5, ja kaikista valitaan vaihtoehto 2, niin kokonaispainokertoimeksi tulee $2{,}30 \times 1{,}95 \times 1{,}50 = 6{,}7275$. Mikäli pelaaja sijoittaa peliin 100 mk ja veikkaa oikein kaikki kolme kohdetta, niin hän saa 100 mk $\times 6{,}7275 = 672{,}75$ 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 $2 \times 2 \times 2$ 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 c^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 luokkanaan ja yhdistä luokat 1,5–1,95 ja 2–2,3 yhdeksi luokaksi nimeltään ”vähintään 1,5”. Voit pyöristää laskutoimituksissa syntyvät luvut kokonaisluvuiksi.

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

		Pelikohteen lopputulos		
		Pienin paino-kerroin ei toteutunut	Pienin paino-kerroin toteutui	Yhteensä
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=riviprosentit, T%F=solufrekvenssien prosenttiosuus kokonaisfrekvensseistä.

a) Kirjoita tämän tehtävän H_0 –hypoteesi ja H_1 –hypoteesi.

b) Laske c^2 -riippumattomuustestin tulos ja testaa tuloksen merkitsevyys 1 % riskitasolla. Pane välitulokset näkyviin, merkitse vapausasteet, testin tulos, 1 % kriittinen raja c^2 -testissä. Mikä on johtopäätöksesi testituloksen perusteella?

c) Kun tarkastelet c^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 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, kuitenkaan 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, kuitenkaan 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 neliöjuurilausekkeita 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 ennätyksensä 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), huipukkuus (kurtosis), alakvartiili (lower_Q), mediaani (median), yläkvartiili (upper_Q) sekä desiilit (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 esitetyssä otoksessa.

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

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

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

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

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?

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 voittosi tai tappiosi 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.

Anna esimerkkinä kahden parhaan kaavan mukaisten pistemäärien (pm_i) jakaumien desiilit 50 pelin sarjassa (Huom esim. Fractile(.1) tarkoittaa jakauman 10 % desiiliä).

	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 kesäjuhlille, 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 kesäjuhlille?

Vuosi 1999, tehtävä 2

Eräässä 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 ”heittokertojen 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). Tunnusluvuista 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 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 ovenviereltä 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 täyttyminen 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 tiettynä iltana tekemää tilastoa, jossa on 8 lukua kertoen jokaisen säiliön täyttöasteen kyseisenä iltana. Jokaisella havaintonäytteellä on oma järjestysnumeronsa (muuttuja NR).

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

Roskisten täyttöasteet

pvm	viikon päivä	NR	R1	R2	R3	R4	R5	R6	R7	R8	perustelut:
2.3.	to	**1**	3	5	4	2	3	0	0	2	
3.3.	pe	**2**	4	7	7	3	6	3	0	2	
6.3.	ma	**3**	6	9	11	8	10	9	1	1	
10.3.	pe	**4**	1	1	7	1	3	3	0	2	
11.3.	la	**5**	1	1	9	2	4	6	1	3	
12.3.	su	**6**	3	3	10	5	9	9	1	3	
13.3.	ma	**7**	5	8	10	5	9	9	1	3	
14.3.	ti	**8**	6	11	10	6	9	10	3	3	
15.3.	ke	**9**	1	1	2	1	1	1	0	0	
16.3.	to	**10**	1	3	6	1	1	3	0	0	
17.3.	pe	**11**	2	6	6	1	2	9	0	0	
18.3.	la	**12**	2	8	9	3	3	5	1	1	
19.3.	su	**13**	3	11	10	4	5	7	1	1	
20.3.	ma	**14**	4	10	11	6	6	10	1	1	
21.3.	ti	**15**	4	11	11	6	9	10	3	3	
22.3.	ke	**16**	0	1	2	0	1	0	0	1	
23.3.	to	**17**	1	1	4	0	2	1	0	1	
24.3.	pe	**18**	1	6	5	2	3	2	1	1	
25.3.	la	**19**	6	9	8	4	4	4	1	1	
26.3.	su	**20**	7	11	9	6	5	6	2	1	
27.3.	ma	**21**	7	11	9	6	5	6	2	1	
28.3.	ti	**22**	7	11	9	7	7	7	3	3	
9.4.	su	**23**	5	8	3	8	10	6	2	0	
10.4.	ma	**24**	6	11	5	8	10	6	2	0	
11.4.	ti	**25**	9	9	7	10	10	8	3	1	
12.4.	ke	**26**	2	1	0	0	0	1	0	0	
25.4.	ti	**27**	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 tervehtimisseremonioita eri vuosilta. Tutkijoilla oli kiinnostuksen kohteena se, onko naisten niiaaminen 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: niiaajat ja ne jotka eivät niianneet.

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	Niiaajat	Kaikki kättelijät	Niiaajien %-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 tarkoitetulla testillä, onko presidentti Kekkosen 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 niiata on 0.8. Kuvitellaan, että kolme nuorta naista tulee peräkkäin kättelytilanteeseen ja ennen heitä on tullut useita miehiä.

a) Mikä on todennäköisyys sille, että vähintään yksi näistä kolmesta niiaa?

b) Mikä on todennäköisyys sille, että keskimmäinen niiaa?

c) Mikä on todennäköisyys sille, että ensimmäisenä tullut on niiannut, jos tiedetään että kolmantena tullut on niiannut.

Vuosi 2002, tehtävä 2

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

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

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 $x_3 = 2$.

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, ..., 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 kuusitahkoinen 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.

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						
		1	*2*	*3*	*4*	*5*	*6*	*Yhteensä*
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.

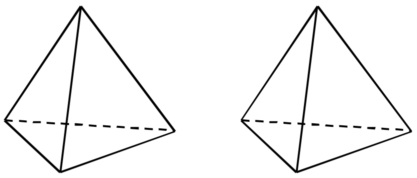
Tutkijan poika huomasi, että vanhan pelin muovikuomu oli niin kulunut, ettei lukuja tahtonut nähdä kuomun läpi. Niinpä hän osti vanhemmilleen joululahjaksi uuden ”Nalle Puh Kimble” -pelin. Tutkijaa kiinnosti nähdä oliko uuden pelin ja vanhan pelin välillä eroa. Seuraavassa on ”Nalle Puh Kimble” -pelin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

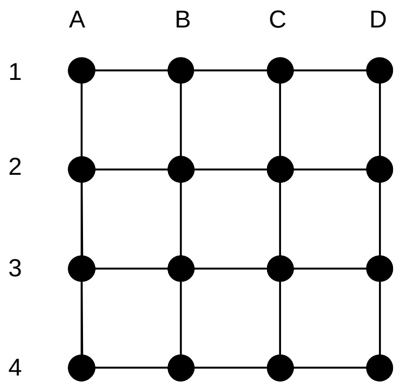
"Nalle Puh Kimble" -pelin nopan tulosten frekvenssit. Taulukossa on ristiintaulukoitu nopanheiton tulosten jakauma sen mukaan, mikä oli ensimmäinen luku ja mikä sitä seuraava luku.

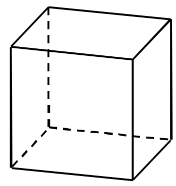
		Seuraava luku						
		1	*2*	*3*	*4*	*5*	*6*	*Yhteensä*
Ensimmäinen luku	*1*	10	14	16	11	7	9	67
	2	13	6	5	7	13	10	54
	3	10	13	5	13	11	9	61
	4	13	8	13	6	9	9	58
	5	6	5	19	11	13	13	67
	6	14	8	4	10	14	3	53
	Yhteensä	66	54	62	58	67	53	360

Vuosi 2004, tehtävä 3

Aarteenetsintä-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.

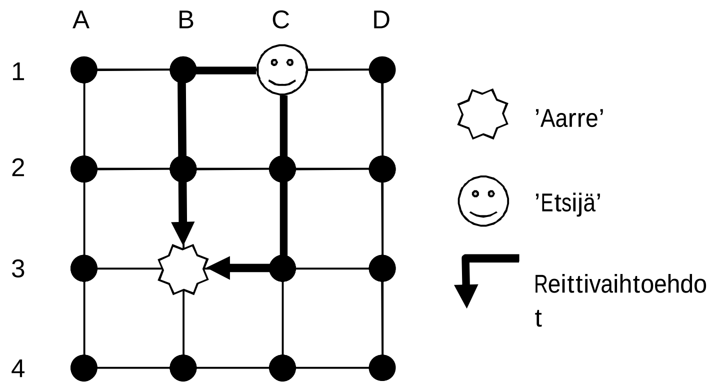
Heitetään molempia tetraedrinoppia uudelleen. Heiton tulos määrittää 'etsijän' paikan pelilaudalla.

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 tetraedri- noppien 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 esimerkissä 3 polkua.



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

Tarkastellaan yhtä ainoata pelikierrosta.

a) **Laske todennäköisyys sille, että ’etsijä’ pääsee ’aarteen’ luo, vaikka kuutionopalla heitettäisiin luku 0? Merkitse näkyviin laskun vaiheet.**

b) **Laske todennäköisyys sille, että ’etsijä’ ei pääse ’aarteen’ luo, vaikka kuutionopalla heitettäisiin luku 5? Merkitse näkyviin laskun vaiheet.**

c) **Selvitä ’etsijän’ kaikkien eripituisten reittien, jotka pelissä ovat mahdollisia, todennäköisyysjakauma ja esitä se taulukkona. Merkitse näkyviin, miten olet laskenut todennäköisyydet.**

d) **Laske todennäköisyys sille, että ’etsijä’ pääsee ’aarteen’ luo. Merkitse näkyviin laskun vaiheet.**

Vuosi 2006, tehtävä 7

Oletetaan, että olet kulttuurin muutoksen tutkija, joka tahtoo selvittää, onko naisen kehon kauneusihanne muuttunut viime vuosikymmeninä, ja jos on, niin millä tavoin. Sinulla on käytettävissäsi asian selvittämiseen aineistoa vuosien 1977 – 2006 Miss Suomi kisoista. Tiedot on koottu Finnartist Oy:n kotisivuilta niiltä vuosilta, joilta ne olivat saatavissa. Tehtävässä oletetaan, että aineisto on edustava. Käytössä on tiedot loppukilpailuun osallistuneiden sijoituksesta: ensimmäinen, toinen, kolmas tai jokin muu sija. Muu sija tarkoittaa loppukilpailuun pääsyä, mutta ei sijoittumista kolmen ensimmäisen joukkoon. Käytössä ovat tiedot osallistumisvuodesta sekä pituus ja paino ja näistä johdettu painoindeksi. Painoindeksi lasketaan seuraavasti PAINOINDEKSI = PAINO / (PITUUDEN NELIÖ), kun paino on mitattu kilogrammoina ja pituus metreinä. Siis esim., jos henkilön paino on 63 kg ja pituus 175 cm, niin hänen painoindeksinsä on $63 / 1{,}75^2 = 20{,}6$. Osallistumisvuodesta on kaksi muuttujaa VUOSI: Vuodet 1977–2006 ja VUOSI2: joka on saatu yhtälöllä VUOSI-1976, eli ensimmäistä tutkittua vuotta 1977 merkitään 1:llä.

Kansanterveyslaitoksen FINRISKI -tutkimuksessa normaalipainon rajoina pidetään nykyään painoindeksin lukuja väliltä 18,5–25. Tutkija selvitti, että FINRISKI -tutkimuksessa 25–34 -vuotiailta pääkaupunkiseudulla asuvilta naisilta vuonna 2002 kerätyssä aineistossa, jossa havaintoja oli 156, painoindeksin minimiarvo oli 17,2, maksimiarvo 36,9, keskiarvo 23,1 ja keskihajonta 3,7. Tässä oletetaan, että tämä olisi edustava aineisto nuorista pääkaupunkiseudulla asuvista suomalaisnaisista.

Taulukko 1. Miss Suomi voittajien pituus, paino, painoindeksi, vuosi ja vuosi2. Vuosi2 = vuosi -1976, eli ensimmäistä tutkittua vuotta 1977 on merkitty 1:llä.

Pituus (cm)	Paino (kg)	Paino-indeksi	Vuosi	Vuosi 2
172	51	17,239	1977	1
172	53	17,915	1978	2
170	50	17,301	1979	3
172	52	17,577	1981	5
175	55	17,959	1983	7
176	55	17,756	1987	11
173	53	17,709	1990	14
173	50	16,706	1993	17
175	56	18,286	1994	18
175	57	18,612	1995	19

Pituus (cm)	Paino (kg)	Paino-indeksi	Vuosi	Vuosi 2
174	60	19,818	1997	21
175	60	19,592	1998	22
171	55	18,809	1999	23
176	59	19,047	2001	25
178	59	18,621	2002	26
176	59	19,047	2003	27
173	53	17,709	2004	28
172	58	19,605	2005	29
175	63	20,571	2006	30

Taulukko 2. Miss-Suomi kilpailijoiden painoindeksit. Sijoitus on luokiteltu seuraavasti: 1, 2, 3, muut = muut loppukilpailuun osallistuneet.
k.a. = keskiarvo, s = keskihajonta, f = havaintojen lkm.

Sijoitus loppukilpailussa		Vuosi									
		1977	1978	1979	1981	1983	1987	1990	1993	1994	1995
1	k.a.	17,239	17,915	17,301	17,577	17,959	17,756	17,709	16,706	18,286	18,612
	s	.	.	.	.	.	.	.	.	.	.
	f	1	1	1	1	1	1	1	1	1	1
2	k.a.	18,809	17,099	17,040	17,836	17,709	18,832	17,239	19,045	17,433	19,265
	s	.	.	.	.	.	.	.	.	.	.
	f	1	1	1	1	1	1	1	1	1	1
3	k.a.	17,915	17,836	18,377	17,374	17,374	17,990	18,424	17,836	19,152	18,724
	s	.	.	.	.	.	.	.	.	.	.
	f	1	1	1	1	1	1	1	1	1	1
muut	k.a.	18,009	18,065	18,276	17,851	18,078	18,188	18,255	18,045	18,097	18,646
	s	,444	,531	,316	,477	,370	,496	,876	,620	,598	1,222
	f	7	7	7	7	7	7	7	9	9	7
Yht.	**k.a.**	**18,003**	**17,931**	**18,065**	**17,774**	**17,959**	**18,189**	**18,116**	**18,000**	**18,145**	**18,712**
	s	**,519**	**,529**	**,542**	**,423**	**,384**	**,485**	**,800**	**,731**	**,634**	**1,017**
	f	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**12**	**12**	**10**

Sijoitus loppukilpailussa		Vuosi									
		1997	1998	1999	2001	2002	2003	2004	2005	2006	**Yht.**
1	k.a.	19,818	19,592	18,809	19,047	18,621	19,047	17,709	19,605	20,571	**18,415**
	s	.	.	.	.	.	.	.	.	.	**1,017**
	f	1	1	1	1	1	1	1	1	1	**19**

2	k.a.	18,207	18,711	19,151	19,377	18,832	18,809	20,478	19,487	18,827	**18,536**
	s	.	.	.	.	.	.	.	.	.	**,927**
	f	1	1	1	1	1	1	1	1	1	**19**
3	k.a.	19,713	17,359	16,652	18,557	19,379	19,157	18,685	18,724	17,577	**18,253**
	s	.	.	.	.	.	.	.	.	.	**,806**
	f	1	1	1	1	1	1	1	1	1	**19**
muut	k.a.	18,359	18,147	19,132	18,343	18,746	19,155	19,417	19,280	18,719	**18,452**
	s	,506	,826	1,416	1,023	,927	,239	,761	1,250	,902	**,872**
	f	7	7	7	7	7	7	7	7	7	**137**
Yht.	**k.a.**	**18,625**	**18,269**	**18,854**	**18,539**	**18,805**	**19,110**	**19,279**	**19,277**	**18,801**	**18,437**
	s	**,731**	**,880**	**1,395**	**,913**	**,785**	**,225**	**,935**	**1,045**	**1,030**	**,882**
	f	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**194**

Taulukko 3. Miss-Suomi kilpailijoiden painoindeksit luokiteltu kolmeen ajanjaksoon ja kahteen painoindeksiluokkaan.
k.a. = keskiarvo, s = keskihajonta, f = havaintojen lkm.

Painoindeksi		Vuosi			
		1977-1989	1990-1999	2000-2006	**Yht.**
alle 18,5	k.a.	17,838	17,755	17,779	**17,797**
	s	0,396	0,565	0,389	**0,468**
	f	49	42	13	**104**
18,5 tai enemmän	k.a.	18,651	19,181	19,298	**19,177**
	s	0,108	0,618	0,659	**0,633**
	f	11	32	47	**90**
Yht.	**k.a.**	**17,987**	**18,372**	**18,969**	**18,437**
	s	**0,480**	**0,920**	**0,876**	**0,882**
	f	**60**	**74**	**60**	**194**

Taulukko 4. Regressioanalyysin tulokset. Selitettävä muuttuja Painoindeksi, selittävä muuttuja VUOSI2. Aineistona missikilpailun voittajat.

TULOKSET	
R	0,695
R^2	0,483
R_a^2	0,452
Keskivirhe	0,752
Havaintoja	19

ANOVA	*df*	*SS*	*MS*	*F*	*Merkitsevyys F*
Regressio	1	8,982	8,982	15,866	< 0,001
Residuaali	17	9,624	0,566		
Kokonais	18	18,606			

	Kertoimet	Keskivirhe	t-arvo	p-arvo	
Vakio	17,169	0,357	48,072	< 0,001	
Vuosi2	0,072	0,018	3,983	< 0,001	

Taulukko 5. Regressioanalyysin tulokset. Selitettävä muuttuja Painoindeksi, selittävä muuttuja VUOSI2. Aineistona missikilpailun 2 ja 3 sijoille sijoittuneet..

TULOKSET	
R	0,504
R^2	0,254
R_a^2	0,233
Keskivirhe	0,761
Havaintoja	38

ANOVA	*df*	*SS*	*MS*	*F*	*Merkitsevyys F*
Regressio	1	7,083	7,083	12,235	0,001
Residuaali	36	20,842	0,579		
Kokonais	37	27,925			

	Kertoimet	Keskivirhe	t-arvo	P-arvo	
Vakio	17,612	0,255	68,962	< 0,001	
Vuosi2	0,045	0,013	3,498	0,001	

Taulukko 6. Regressioanalyysin tulokset. Selitettävä muuttuja Painoindeksi, selittävä muuttuja VUOSI2. Aineistona missikilpailun finalistit, jotka eivät päässeet kolmen parhaan joukkoon.

ANOVA	*df*	*SS*	*MS*	*F*	*Merkitsevyys F*

TULOKSET	
R	0,397
R^2	0,158
R_a^2	0,151
Keskivirhe	0,804
Havaintoja	137

Regressio	1	16,317	16,317	25,270	0,0000
Residuaali	135	87,173	0,646		
Kokonais	136	103,490			

	Kertoimet	Keskivirhe	t-arvo	P-arvo	
Vakio	17,817	0,144	123,975	0,00000	
Vuosi2	0,037	0,007	5,027	0,0000	

Taulukko 7. Kansanterveyslaitoksen Finriski -tutkimuksesta painoindeksin tunnuslukuja 25–34 -vuotiailta pääkaupunkiseudulla asuvilta suomalaisnaisilta vuodelta 2002.

Painoindeksi	alle 18,5	18,5–24,99	25,0–29,99	30 tai enemmän
Prosenttiosuus	5,1 %	73,1 %	16,7 %	5,1 %

a) Laske annetun aineiston perusteella regressioanalyysin antama ennuste vuoden 2006 Miss Suomen painoindeksiksi. Merkitse oikea lauseke ja sen arvo.

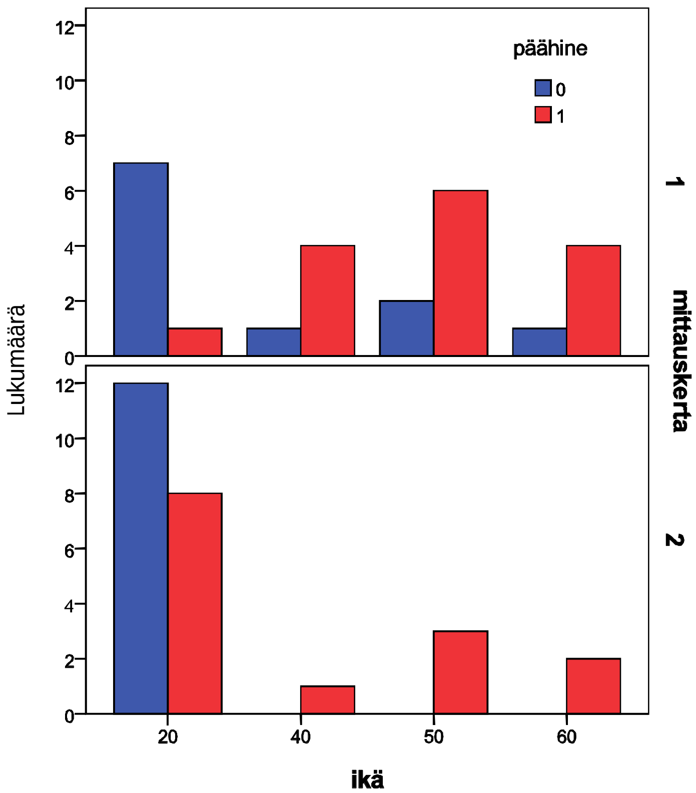
b) Laske Miss Suomen 2006 painoindeksin residuaali. Merkitse oikea lauseke ja sen arvo.

c) Seuraavassa halutaan tutkia, oliko painoindeksi normaalisti jakautunut 25–34 -vuotiaiden pääkaupunkiseudulla asuvien naisten vuoden 2002 aineistossa. Laske kuinka suuri osa jakaumasta (prosentin kymmenesosan tarkkuudella) kuuluisi painoindeksiluokkaan 18,5–24,99, mikäli painoindeksi olisi normaalijakautunut.

d) Kerro annetun tiedon valossa, miten kauneusihanne on muuttunut vuodesta 1977 vuoteen 2006. Kuvaa vain keskeiset tekijät ja kerro mihin materiaalin perustat vastauksesi. Virheellisistä perusteluista ja johtopäätöksistä vähennetään pisteitä.

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 tutkimusyksikköjä (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 kiintiöotantaa käyttäen iän ja sukupuolen perusteella.

	mittauskerta							
	1				2			
	sukupuoli				sukupuoli			
	M		N		M		N	
	päähine		päähine		päähine		päähine	
ikä	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) $P(X<5)=\frac{5-3,5}{6-3,5}=0.6$

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

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

Odotetut frekvenssit

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

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

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

Tarkka arvo (kun odotetut frekvenssit desimaalilukuja): c^2 =8,19172; df=(2-1)·(2-1)=1.

Kriittinen raja c^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 c^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 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 todennäköisyyssääntöä saadaan, että yhdessä messussa todennäköisyys, että tarvitaan kaikki 12 numero ykkösen laattaa on:

Normaalipyhä

pNM= (24/631)*(23/630)*(22/629)*(21/628)*(20/627)*(19/626)

pNM=0,00000000157233

Todennäköisyys, että ei tarvita: 1-pNM=0,99999999842767

Erikoispyhä

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

pEM=(24/631)^6=0,00000000302755

Todennäköisyys, että ei tarvita: 1-pEM=0,99999999697245

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

Lasketaan ensiksi todennäköisyys, että missään kirkossa ei normaalipyhinä 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 tn:ä käyttäen todennäköisyys, että ykköset eivät lopu missään kirkossa 10 vuoden aikana.

p(eivät lopu)=(1-pNM)^(400*51*10)*(1-pEM)^(400*10)

p(loppuvat)=1-p(eivät lopu)

p(loppuvat)=1-((1-pNM)^(400*51*10))*((1-pEM)^(400*10))

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

painottaa pieniä poikkeavia arvoja vähemmän kuin kaava 2)

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

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.

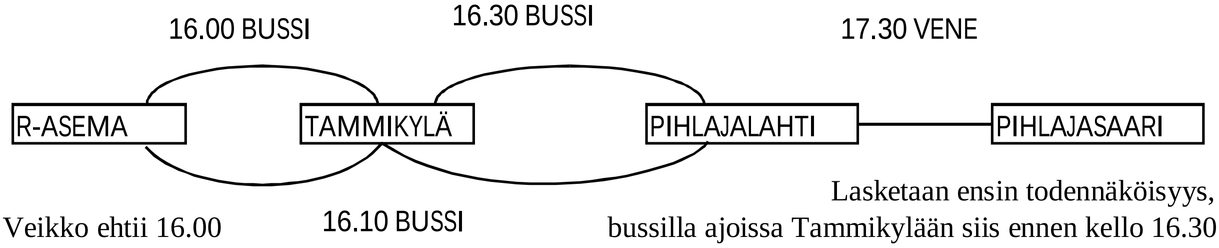
Kaavalla 2 kannattaa pelata puolet peleistä (tai 26 peliä, riippuen miten laskee mediaanin) siten, että niiden pistemääräksi tulee nolla ja loput mahdollisimman hyvin. Tällöin Mediaani on nolla eikä tulokseen tule lainkaan sitä laskevia negatiiviisia arvoja.

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).

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

Vuosi 1996, tehtävä 3

Veikon mahdolliset tavat kulkea Pihlajasaareeen ovat.



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

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

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

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

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

Lasketaan todennäköisyys, ehtiä 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

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

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

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

Jos vene lähtee, on sen todennäköisyys, olla ajoissa saaressa

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

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

(0,6915 ×0,9772 + 0,3085 ×0,5) ×0,9 ×1 » 0,747

Vuosi 1999, tehtävä 2

a)

1) P(kutonen tulee ensimmäisen kerran ensimmäisellä heittokerralla) = 1/6 » 0,167

2) P(kutonen tulee ensimmäisen kerran toisella heittokerralla) = (5/6)Ÿ(1/6) » 0,139

3) P(kutonen tulee ensimmäisen kerran kolmannella heittokerralla) = (5/6)Ÿ(5/6)Ÿ(1/6) » 0,116

b)

P(ensimmäinen kutonen tulee neljännellä heittokerralla tai myöhemmin) =
1-P(kutonen tulee ensimmäisen kerran ensimmäisellä, toisella tai kolmannella heittokerralla) = 1 - (1/6 + (5/6)Ÿ(1/6) + (5/6)Ÿ(5/6)Ÿ(1/6)) = 1 - 91/216 = 125/216 » 0,579

c)

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

x (heittokerta, jolla saadaan ensimmäinen kutonen)	P(kutonen tulee ensimmäisen kerran x:nnellä heittokerralla)	P:n Kertymäfunktio
1	1/6 » 0,167	1/6» 0,167
2	(5/6)Ÿ(1/6) » 0,139	1/6+(5/6)Ÿ(1/6) = 11/36 » 0,306
3	(5/6)Ÿ(5/6)Ÿ(1/6) » 0,116	1/6+(5/6)Ÿ(1/6)+(5/6)Ÿ(5/6)Ÿ(1/6)= 91/216 » 0,421
4	(5/6)Ÿ(5/6)Ÿ(5/6)Ÿ(1/6) » 0,096	1/6+(5/6)Ÿ(1/6)+(5/6)Ÿ(5/6)Ÿ(1/6)+ (5/6)Ÿ(5/6)Ÿ(5/6)Ÿ(1/6)= 671/1296 » 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(ensimmäisen kutonen tulee ensimmäisellä nostolla) = 2/12 » 0,167

2) P(ensimmäisen kutonen tulee toisella nostolla) = (10/12)Ÿ(2/11) » 0,152

3) P(ensimmäisen kutonen tulee kolmannella nostolla) = (10/12)Ÿ(9/11)Ÿ(2/10) » 0,136

f)

P(ensimmäinen kutonen tulee neljännellä nostolla tai myöhemmin) =

1-P(kutonen tulee ensimmäisen kerran ensimmäisellä, toisella tai kolmannella nostolla) =

1 - (2/12 + (10/12)Ÿ(2/11) + (10/12)Ÿ(9/11)Ÿ(2/10)) = 1 - 600/1320 = 720/1320 » 0,545

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	**2**	4	7	7	3	6	3	0	2
11.3.	la	**5**	1	1	9	2	4	6	1	3
17.3.	pe	**11**	2	6	6	1	2	9	0	0
25.3.	la	**19**	6	9	8	4	4	4	1	1
12.4.	ke	**26**	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 χ^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

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(14-15)^2}{15} + \frac{(24-15)^2}{15} + \frac{(30-15)^2}{15} + \frac{(11-15)^2}{15} + \frac{(15-15)^2}{15} + \frac{(17-15)^2}{15} + \frac{(3-15)^2}{15} + \frac{(6-15)^2}{15}$$

$$= \frac{1+81+225+16+0+4+144+81}{15} = \frac{552}{15} = 36{,}8$$

Vapausasteet: f=8-1-0=7, jolloin merkitsevyystasolla 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 χ^2-yhteensopivuustestin avulla. Lasketaan ensin odotetut frekvenssit tilanteessa, jossa niiaajien suhteellinen osuus olisi ollut yhtä suuri jokaisen presidentin aikana.

Niiaajien kokonaismäärä: 92+72+52+14 = 230

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

Näin ollen niiaajien odotettu suhteellinen osuus on 230/1313 ≈ 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: Niiaaminen on ollut yhtä yleistä kaikkien neljän presidentin aikana.

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

χ^2-testisuure on

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

Testisuure noudattaa χ^2-jakaumaa vapausasteilla k – 1 = 4 – 1 = 3.

0,1 % merkitsevyystason kriittinen arvo on χ^2-jakauman kertymäfunktiotaulukon perusteella 16.266. Koska havaittu testisuureen arvo on suurempi kuin 0,1 % merkitsevyystason 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 % merkitsevyystasolla kriittinen arvo on 1,96. Koska testisuure on pienempi, nollahypoteesi jää voimaan. Aineiston perusteella ei voi väittää, että Kekkosen viimeisenä vuonna 1980 ja ensimmäisenä aineistossa olevana Koiviston vuonna 1986 niiaamisessa olisi ollut eroa.

Vuosi 2001, tehtävä 2

a)

P(vähintään yksi niiaa) = 1-P(kukaan ei niiaa)

1-0,5·0,5·0,5 = 0,875

b)

P(N1 niiaa ja N2 niiaa)+P(N1 ei niiaa ja N2 niiaa)

0,5·0,8+0,5·0,5 = 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	0,5·0,8·0,8 = 0,32
niiaa	ei niiaa	niiaa	0,5	0,2	0,5	0,5·0,2·0,5 = 0,05
ei niiaa	niiaa	niiaa	0,5	0,5	0,8	0,5·0,5·0,8 = 0,2
ei niiaa	ei niiaa	niiaa	0,5	0,5	0,5	0,5·0,5·0,5 = 0,125

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

Todennäköisyys sille, että ensimmäisenä tullut on niiannut, kun tiedetään että kolmantena tullut on niiannut on (0,32+0,05)/0,695 ≈ 0,532.

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, H1 vai H2, saa suuremman arvon ennustemuuttujalla. Koska muuttujilla x_4, x_5, ..., 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

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

b)

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

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

c)

Ensin on selvitettävä 1. ja 2. lapseksi syntyneiden poikien ja tyttöjen lukumäärien odotusarvot. Koska perheitä on 120, niin ensimmäinen lapsi on poika 0,5·120=60 perheessä eikä näihin synny toista lastaa.

Ensimmäinen lapsi on tyttö 0,5·120=60 perheessä ja näistä 0,5·60=30 toinen lapsi on poika ja 0,5·60=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 2/3·60=40 P-poikaa valitsee vaimokseen TT-tytön ja 1/3·60=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, $e_i \neq 10$

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

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

$$f = 6 - 1 = 5$$

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

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.

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

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.

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

c)

Erilaisia ’etsijä’ – ’aarre’ tilanteita eli mahdollisia reittejä on yhteensä 16·16=256 kpl. Mahdolliset reitin pituudet ovat 0, 1, 2, 3, 4, 5 ja 6 polkua. Näistä itse asiassa 0 ja 6 pituisen polun tn:t laskettiin kohdissa 3a ja 3b. Tn: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	4·1+4·1+8·1= 16	16/256=1/16
1	2	4	3	4·2+4·4+8·3= 48	48/256=3/16
2	3	6	4	4·3+4·6+8·4= 68	68/256=17/64
3	4	4	4	4·4+4·4+8·4= 64	64/256=1/4
4	3	1	3	4·3+4·1+8·3= 40	40/256=5/32
5	2	0	1	4·2+4·0+8·1= 16	16/256=1/16
6	1	0	0	4·1+4·0+8·0= 4	4/256=1/64
				256	

d)

Kysytty todennäköisyys saadaan reitin pituuksien tn:ien ja kuutionopan tn:ien yhdistelmänä.

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

Vuosi 2006, tehtävä 7

a)

17,169 + 0,072 × 30 = 19,329

b)

20,571 - 19,329 = 1,242

c)

Painoindeksi X ~ N(23,1;3,7)

$$
\begin{aligned}
P(18{,}5 \le x \le 24{,}99) &= P\left(\frac{18{,}5 - 23{,}1}{3{,}7} \le z \le \frac{24{,}99 - 23{,}1}{3{,}7}\right) \\
&= P(-1{,}2432432 \le z \le 0{,}5108108) \\
&= P(z \le 0{,}51) - P(z \le -1{,}24) \\
&= \varphi(0{,}51) - \varphi(-1{,}24) \\
&= (1 - \varphi(-0{,}51)) - \varphi(-1{,}24) \\
&= (1 - 0{,}3050) - 0{,}1075 \\
&= 0{,}6950 - 0{,}1075 \\
&= 0{,}5875
\end{aligned}
$$

V: 58,8 %

d)

Kun kauneusihannetta tarkastellaan missien painoindeksinä, niin sekä keskiarvon että hajonnan havaitaan kasvaneen kyseisellä aikavälillä. Tämä nähdään taulukoista 2 ja 3. Myös se, että Vuosi2- muuttujan regressiokertoimet ovat positiivisia taulukoissa 4 – 6 kertoo keskiarvon kasvusta. Se, että voittaneiden vakiotermi on pienempi, mutta regressiokerroin on suurempi kuin muilla kertoo, että voittajien kohdalla muutos on ollut vahva. Voidaan siis sanoa, että tarkastellulla aikavälillä kauneusihanne on avartunut ja muuttunut FINNRISKI rajojen mukaisten normaaliarvojen suuntaan.

Lopuksi

Runsaasti intoa valintakokeeseen valmistautumiseen!

Psykologian tilastotieteen valintakoemateriaali 2019

Valintakoemateriaalissa havaitut virheet

Muokattu 12.4.2019

Sivu 10: Seuraava lause on virheellinen:

Jos kaksi tapahtumaa A ja B ovat erillisiä eli toisensa poissulkevia tapahtumia, niin todennäköisyys sille, että joko tapahtuma A tai tapahtuma B tapahtuu (myös molemmat voivat tapahtua samanaikaisesti) on

$$p(A \; tai \; B) = p(A) + p(B) - p(A \; ja \; B)$$

Pitäisi lukea:

Yleisesti todennäköisyys sille, että joko tapahtuma A tai tapahtuma B tapahtuu (myös molemmat voivat tapahtua samanaikaisesti) on

$$p(A \; tai \; B) = p(A) + p(B) - p(A \; ja \; B)$$

Sivu 15:

Yleisesti, jos tapahtumassa on n kappaletta alkioita ja näistä halutaan tarkastella sellaisia variaatioita, joissa esiintyy k kappaletta alkioita, niin mahdollisten tapahtumien lukumäärä on $n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots \cdot (n-k)$, joka voidaan laskea myös kaavalla $\frac{n!}{(n-k)!}$. Esimerkiksi esimerkissä mainitussa tilanteessa lukumäärä on siis $\frac{n!}{(n-k)!} = \frac{10!}{(10-5)!} = \frac{3628800}{120} = 30240$.

Pitäisi lukea:

Yleisesti, jos tapahtumassa on n kappaletta alkioita ja näistä halutaan tarkastella sellaisia variaatioita, joissa esiintyy k kappaletta alkioita, niin mahdollisten tapahtumien lukumäärä on $n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots \cdot (n-k+1)$, joka voidaan laskea myös kaavalla $\frac{n!}{(n-k)!}$. Esimerkiksi esimerkissä mainitussa tilanteessa lukumäärä on siis $\frac{n!}{(n-k)!} = \frac{10!}{(10-5)!} = \frac{3628800}{120} = 30240$.

Sivu 25:

Luottamusväliä koskevassa esimerkki laskussa on luottamusvälin kertoimessa pieni virhe. Jos otoskoko on n=100, luottamusvälin kerroin tulisi olla 1,98. Esimerkissä annettu kerroin 2,26 on laskettu otoskoolla 10.

Sivu 26:

Kahden riippumattoman otoksen t-testin esimerkissä on laskuvirhe, joka vaikuttaa myös tuloksen tulkintaan. Materiaalissa kirjoitettu virheellisesti:

$$S = \sqrt{\frac{(10-1) \cdot 16.06 + (10-1) \cdot 8.92}{(10-1)+(10-1)}} \approx 3{,}53, \text{ jolloin } t = \frac{110.50 - 102.20}{3{,}53 \cdot \sqrt{\left(\frac{1}{10}+\frac{1}{10}\right)}} \approx 5{,}258$$

Pitäisi olla:

$S = \sqrt{\frac{(10-1)\cdot 16.06^2+(10-1)\cdot 8.92^2}{(10-1)+(10-1)}} \approx$ 12,99, jolloin $t = \frac{110.50-102.20}{12,99\cdot\sqrt{\left(\frac{1}{10}+\frac{1}{10}\right)}} \approx$ 1,43, joka ei ole merkitsevä 5 % riskitasolla, eli nollahypoteesi jää voimaan, koska t-jakauman kriittinen arvo 5 % riskitasolla ja vapausasteilla 18 on 2,10.

Sivu 30:

$R = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}$, *pitäisi lukea* $R = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$

Sivu 32:

$t = \frac{0,5\cdot\sqrt{100-2}}{\sqrt{1-r0,5^2}} = \frac{0,5\cdot 10}{\sqrt{1-0.25}} = \frac{5}{\sqrt{0,75}} \approx$ 5,77 , *pitäisi olla* $t = \frac{0,5\cdot\sqrt{102-2}}{\sqrt{1-0,5^2}} = \frac{0,5\cdot 10}{\sqrt{1-0.25}} = \frac{5}{\sqrt{0,75}} \approx$ 5,77

Sivu 33:

Lause: Vaihtoehtoisesti regressiosuoran kulmakerroin $b_1$voidaan laskea myös kaavalla $b_1 = \frac{S_{xy}}{S_x}$, jossa $S_{xy} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1}$, eli muuttujien x ja y välinen kovarianssi ja $S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$, eli muuttujan x keskihajonta.

Pitäisi olla:

Vaihtoehtoisesti regressiosuoran kulmakerroin $b_1$voidaan laskea myös kaavalla $b_1 = \frac{S_{xy}}{{S_x}^2}$, jossa $S_{xy} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1}$, eli muuttujien x ja y välinen kovarianssi ja $S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$, eli muuttujan x keskihajonta.

Tai

Vaihtoehtoisesti regressiosuoran kulmakerroin $b_1$voidaan laskea myös kaavalla $b_1 = \frac{S_{xy}}{{S_x}^2}$, jossa $S_{xy} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1}$, eli muuttujien x ja y välinen kovarianssi ja ${S_x}^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$, eli muuttujan x varianssi.

Sivu 36:

Vastaavasti suorana yleistyksenä yhden selittäjän tapauksesta voidaan johtaa vakiotermille b_0 kaava, joka on

$b_0 = \bar{Y} - b_1\bar{X}_1 - b_1\bar{X}_2$ *pitäisi olla* $b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$

Sivu 37: (viimeinen kaava)

$\sum(x_1x_2) = \sum(X_1-\bar{X}_1)(X_2-\bar{X}_2) = \sum X_2Y - \frac{\sum X_2\sum Y}{N}$, *pitäisi olla* $\sum(x_1x_2) = \sum(X_1-\bar{X}_1)(X_2-\bar{X}_2) = \sum X_1X_2 - \frac{\sum X_1\sum X_2}{N}$

Korjattu 28.6.2019 tehtävän 2.1 kohdan 2 osalta
Täsmennetty 29.6.2019 tehtävän 2.2 kohdan 2 pisteytystietoa

Mallivastaukset psykologia – Helsingin, Itä-Suomen, Jyväskylän, Tampereen ja Turun yliopistojen valintakoeyhteistyö 2019

Tehtävä 1.1

Tehtävä 1.1.1

a) A, B, D ja E

b) A, C, D ja E

c) A, D ja E

d) A ja E

e) D

Pisteytys: Jokaisesta täysin oikeasta kohdasta 2.3 pistettä.

Tehtävä 1.1.2

a) Kaava numero 4 tai 5 (0,7 pistettä)

b) 0,583 (4,9 pistettä)

c) 1,926 (3,1 pistettä)

d) 13 (1,1 pistettä)

e) iv) $p > 0,05$ (1,7 pistettä)

f) Kaava numero 11 (0,7 pistettä)

g) 0,457 (4,9 pistettä)

h) 2,410 (3,1 pistettä)

i) 12 (1,1 pistettä)

j) iii) $p < 0,05$ (1,7 pistettä)

k) kaava 12 (0,7 pistettä)

l) 0,674 (4,9 pistettä)

m) 4,403 (3,1 pistettä)

n) 12 (1,1 pistettä)

Korjattu 28.6.2019 tehtävän 2.1 kohdan 2 osalta
Täsmennetty 29.6.2019 tehtävän 2.2 kohdan 2 pisteytystietoa

o) i) $p<0,001$ (1,7 pistettä)

p) iii) Ei havaittavissa mediaatio efektiä (2,0 pistettä)

Tehtävä 1.2

1) c

2) a ja c

3) a

4) b

5) d

6) d

7) b ja d

8) a

9) c

10) c ja d

Pisteytys: Jokaisesta täysin oikeasta kohdasta 2.1 pistettä.

Tehtävä 2.1

1) b

2) c (Korjattu 28.6.2019, virhe alkuperäisessä mallivastauksissa. Tarkistus skriptissä oikein, eikä vaikuta tuloksiin.)

3) b ja c

4) b

5) b ja c

6) a ja c

7) d

8) b ja d

9) a, c ja d

Korjattu 28.6.2019 tehtävän 2.1 kohdan 2 osalta
Täsmennetty 29.6.2019 tehtävän 2.2 kohdan 2 pisteytystietoa

10) c

11) c

12) b

13) a, b, c ja d

14) a, c ja d

15) a, b ja c

16) a, b, c ja d

17) a ja b

18) b, c ja d

19) a, c ja d

20) d

21) a, b ja d

22) b ja d

23) b ja c

24) a ja d

25) b ja d

26) a

27) b, c ja d

28) a ja c

29) b, c ja d

30) a, b ja c

Pisteytys: Täysin oikea vastaus 1,7 pistettä. Mikäli tehtävässä oli useampi kuin yksi vastausvaihtoehto oikea sai yhden oikean vaihtoehdon valitsematta jättämisestä 0,7 pistettä, jos muut kohdat olivat oikein.

Korjattu 28.6.2019 tehtävän 2.1 kohdan 2 osalta
Täsmennetty 29.6.2019 tehtävän 2.2 kohdan 2 pisteytystietoa

Tehtävä 2.2

1) a

2) a

3) c

4) e

5) d

6) d

7) b

8) d

9) e

10) c

11) d

12) e

13) b

14) e

15) c

Pisteytys: Täysin oikea vastaus 1,7 pistettä (lukuun ottamatta kohta 2, josta maksimi 1.2 pistettä, huomautus lisätty 29.6.2019). Osio-vaste analyysin perusteella osassa tehtävistä arvaamisen todennäköisyys oli erittäin korkea, joten tämän huomioiminen laski haastavampien tehtävien pisteytystä, jolloin kaikista kohdista annettiin sama pistemäärä.

Valintakoetehtävistä saatava yhteenlaskettu raakapistemäärä muutetaan valintakoepisteiksi välille 0,000 – 120,000 seuraavasti:

• Kaikista psykologian alan valintakoeyhteistyönä järjestettävään valintakokeeseen osallistuneista hakijoista ne hakijat, jotka kuuluvat 1 % parhaiten vastanneiden joukkoon, saavat lopulliseksi valintakoepistemääräkseen 120,000 pistettä.