Aki Taanila

TILASTOLLINEN PÄÄTTELY

12.5.2016

SISÄLLYS

0 JOHDANTO 1
1 TILASTOLLINEN PÄÄTTELY 2
2 YHTÄ MUUTTUJAA KOSKEVA PÄÄTTELY 7
2.1 Normaalijakautuneisuuden testaaminen 7
2.2 Keskiarvon luottamusväli 8
2.3 Keskiarvon testaus (t-testi) 9
2.4 Prosenttiluvun luottamusväli 10
2.5 Prosenttiluvun testaus (binomitesti) 11
2.6 Khiin neliö -yhteensopivuustesti 12
3 KAHDEN RYHMÄN VERTAILU 14
3.1 Riippumattomien otosten t-testi 15
3.2 Riippuvien otosten t-testi 16
3.3 Mann-Whitney U -testi 18
3.4 Wilcoxonin merkittyjen sijalukujen testi 19
3.5 Khiin neliö -riippumattomuustesti 21
3.6 McNemar-testi 21
4 USEAMMAN RYHMÄN VERTAILU 23
4.1 Yksisuuntainen varianssianalyysi 23
4.2 Kruskall-Wallis -testi 26
4.3 Khiin neliö -riippumattomuustesti 27
5 KAHDEN MUUTTUJAN VÄLINEN RIIPPUVUUS 30
5.1 Korrelaatiokertoimen testaus 30
5.2 Khiin neliö -riippumattomuustesti 31
6 TÄRKEITÄ HUOMIOITA 33
6.1 p-arvon tulkinta 33
6.2 Tilastollinen merkitsevyys ja käytännön merkitsevyys 33
6.3 Normaalijakautuneisuus ja otoskoko 30 34

0 JOHDANTO

Löydät viimeisimmän version tästä monisteesta Akin menetelmäblogista http://tilastoapu.wordpress.com

Käsittelen tässä monisteessa tilastollista päättelyä. Tilastollinen päättely tarkoittaa otoksesta laskettujen tulosten yleistämistä laajempaan perusjoukkoon, josta otos on poimittu. Esitän asiat käytännön soveltajan näkökulmasta. Käsiteltävien menetelmien taustalla olevaa matematiikkaa (todennäköisyysjakaumia jne.) en käsittele. Lähtötietoina edellytän aineistojen esittämiseen ja kuvailuun käytettävien menetelmien hallinnan (keskiarvo, keskihajonta, mediaani, korrelaatio, ristiintaulukointi).

Tilastolliseen päättelyyn liittyvät laskutoimitukset on helpointa suorittaa tilasto-ohjelmalla. Tämä moniste sisältää SPSS tilasto-ohjelman ohjeet esiteltyjen menetelmien osalta. Lisätietoa voit hakea SPSS:n sisäänrakennetuista ohjeista (SPSS:n valintaikkunoissa on **Help-**painike, jota napsauttamalla pääset lukemaan kyseiseen toimintoon liittyviä ohjeita).

Virhemarginaaleja ja joitain testejä voidaan melko helposti laskea myös Excelillä. Tästä löydät lisätietoa menetelmäblogistani http://tilastoapu.wordpress.com.

Jos tilastollinen päättely ei ole sinulle entuudestaan tuttua, niin opiskele huolellisesti ensimmäinen luku ennen muiden lukujen lukemista.

Aki Taanilan nettimateriaaleja

Tutustu myös nettimateriaaleihini:

Akin menetelmäblogi http://tilastoapu.wordpress.com

Olennaiset Excel-taidot http://excelapu.wordpress.com

Datojen analysointi http://analysointi.wordpress.com

1 TILASTOLLINEN PÄÄTTELY

Tilastollinen päättely tarkoittaa perusjoukkoa koskevien päätelmien tekemistä perusjoukosta poimitun otoksen perusteella. Otoksesta laskettuja tuloksia ei voida suoraan yleistää laajempaa perusjoukkoa koskeviksi, vaan päättelyssä täytyy huomioida otantavirheestä aiheutuva epävarmuus. Tilastollinen päättely voi sisältää

- virhemarginaalien/luottamusvälien laskemista
- hypoteesien testausta / merkitsevyystestausta.

Tilastollisen päättelyn käyttöedellytyksenä on, että otos on valittu satunnaisesti asianmukaista otantamenetelmää käyttäen.

Otantavirhe

Otoksesta lasketut taulukot ja tunnusluvut kuvailevat otosta. Otoksen perusteella voidaan tehdä päätelmiä perusjoukosta jos otos on satunnaisesti valittu. Jos otosta ei ole valittu satunnaisesti, niin sitä on syytä kutsua näytteeksi.

Otoksen perusteella tehtyihin päätelmiin liittyy otantavirheen aiheuttamaa epävarmuutta. Otantavirhe seuraa siitä, että otoksen kokoonpano riippuu sattumasta ja näin ollen otoksesta lasketut tulokset vaihtelevat satunnaisesti otoksesta toiseen. Otantavirhe on luonnollisesti sitä pienempi mitä suurempaa otosta käytetään.

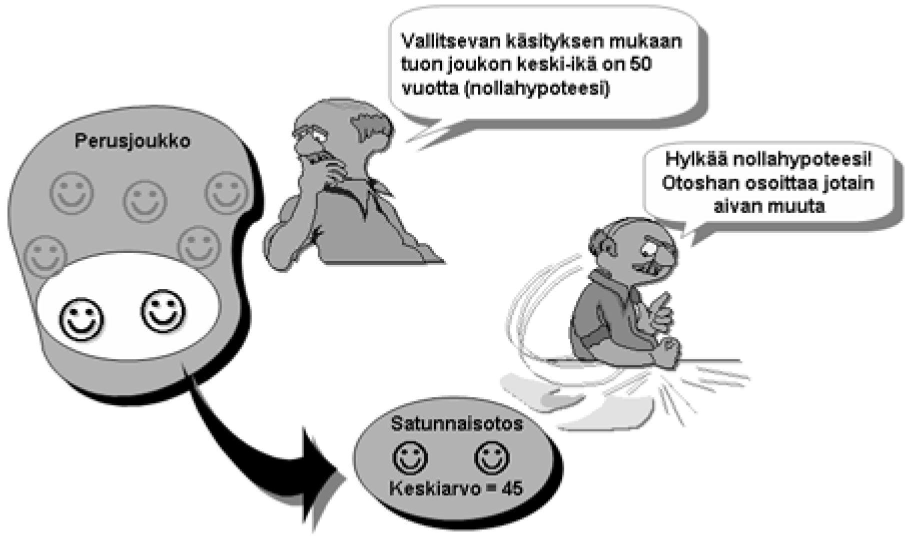
Virhemarginaali/Luottamusväli

Jos haluat tietää perusjoukon tunnusluvun arvon ja käytössäsi on otos perusjoukosta, niin paras arvaus perusjoukon tunnusluvun arvoksi on otoksesta laskettu tunnusluku. Otantavirheen aiheuttaman epävarmuuden voit ilmaista virhemarginaalina. Yleensä ilmoitetaan 95 % virhemarginaali. Luottamusväliksi kutsutaan otoksesta lasketun tunnusluvun ympärille muodostettua väliä tunnusluku ± virhemarginaali.

Esimerkki. Mielipidekyselyn mukaan 40,8 % suomalaisista kannattaa uuden ydinvoimalan rakentamista. Virhemarginaali on 3,4 prosenttiyksikköä. 95 % luottamusväli on siis 37,4 % - 44,2 %. Tämä tarkoittaa sitä, että väli 37,4 % - 44,2 % sisältää ydinvoimalan kannattajien todellisen osuuden 95 % varmuudella.

Hypoteesin testaus / merkitsevyystestaus

Hypoteesin testauksen lähtökohtana on nollahypoteesi, joka oletetaan oikeaksi, ellei otoksesta löydy todisteita sitä vastaan. Nollahypoteesi voi koskea esimerkiksi keskiarvon tai prosenttiluvun suuruutta. Tällöin nollahypoteesin lähteenä voi olla vallitseva käsitys, teoria, aikaisempi tutkimus, valmistajan ilmoitus jne. Tavallisimmin hypoteesi koskee ryhmien välistä eroa tai muuttujien välistä riippuvuutta. Tällöin nollahypoteesina on ’ei eroa’ tai ’ei riippuvuutta’. Jos otos antaa riittävät todisteet nollahypoteesia vastaan, niin nollahypoteesi hylätään.



Joissain yhteyksissä puhutaan hypoteesin testauksen sijasta merkitsevyystestauksesta ja jätetään muodollinen nollahypoteesin ja vaihtoehtoisen hypoteesin kirjaaminen tekemättä. Sen sijaan puhutaan eron tai riippuvuuden merkitsevyydestä.

Koska hypoteesin testaus perustuu otokseen, niin virhepäätelmän mahdollisuus on läsnä. Hypoteesin testauksessa toteutuu yksi seuraavista neljästä vaihtoehdosta:

- Nollahypoteesi on oikeasti totta ja testauksen tuloksena nollahypoteesi jää voimaan (oikea päätös).
- Nollahypoteesi ei oikeasti ole totta, mutta testauksen tuloksena nollahypoteesi jää voimaan (hyväksymisvirhe).
- Nollahypoteesi on oikeasti totta, mutta testauksen tuloksena nollahypoteesi päätetään hylätä (hylkäämisvirhe).
- Nollahypoteesi ei ole oikeasti totta ja testauksen tuloksena nollahypoteesi päätetään hylätä (oikea päätös).

	Todellinen tilanne	
Testauksen tulos	**Nollahypoteesi totta**	**Nollahypoteesi ei ole totta**
Hyväksy nollahypoteesi	Oikea päätös	Hyväksymisvirhe
Hylkää nollahypoteesi	Hylkäämisvirhe	Oikea päätös

Nollahypoteesi on perusolettamus ja se on syytä jättää voimaan ellei ole riittäviä todisteita sitä vastaan. Tämän vuoksi hylkäämisvirhettä (nollahypoteesi on oikeasti totta, mutta testauksen tuloksena se päätetään hylätä) pidetään vakavana virheenä, jota ei mielellään tehdä. Hylkäämisvirheen mahdollisuus on seurausta otantavirheestä ja hylkäämisvirheen todennäköisyys voidaan laskea. Hylkäämisvirheen todennäköisyyttä kutsutaan **p-arvo**ksi tai havaituksi merkitsevyystasoksi.

Toinen tapa tulkita p-arvo on seuraava: p-arvo on todennäköisyys sille, että havaittu poikkeama nollahypoteesista on sattuman (otantavirheen) aiheuttama.

Yleisesti käytetty päättelysääntö on seuraavanlainen: Jos p-arvo on alle 0,050 (5 %), niin nollahypoteesi hylätään. Muussa tapauksessa nollahypoteesi jää voimaan. Jos hypoteeseista ei haluta puhua, niin alle 5 % p-arvo tarkoittaa merkitsevää eroa/riippuvuutta. Jos p-arvo on yli 5 %, niin ero/riippuvuus ei ole merkitsevä.

Hyväksymisvirheen todennäköisyyden laskeminen on vaikeampaa kuin hylkäämisvirheen todennäköisyyden. Kannattaa kuitenkin pitää mielessä, että mitä pienempää hylkäämisvirheen todennäköisyyttä vaaditaan nollahypoteesin hylkäämiseksi sitä suuremmaksi kasvaa hyväksymisvirheen todennäköisyys. Päättelysäännössä käytetty 5 % raja onkin kompromissi hylkäämisvirheen ja hyväksymisvirheen välillä. Käytännön tilanteesta riippuen voidaan raja asettaa muunkin suuruiseksi. Jos hylkäämisvirhe koetaan erityisen kohtalokkaaksi, niin rajaksi voidaan asettaa esimerkiksi 1 % tai 0,1 %. Jos taas halutaan helpommin erottaa poikkeamia nollahypoteesista, niin rajaksi voidaan asettaa esimerkiksi 10 %.

Jos sekä hylkäämisvirheen että hyväksymisvirheen todennäköisyyttä halutaan yhtä aikaa pienemmäksi, niin on turvauduttava isompaan otokseen.

Hypoteesin testaukseen liittyy tyypillisesti seuraavia vaiheita:

- Muotoile nollahypoteesi ja vaihtoehtoinen hypoteesi.
- Kerää havainnot (satunnaisesti valittu otos!).
- Laske hylkäämisvirheen todennäköisyys eli p-arvo.
- Päättelysääntö: Hylkää nollahypoteesi, jos p-arvo on pienempi kuin 0,050 (5 %). Muussa tapauksessa nollahypoteesi jää voimaan.

Esimerkki. Laakerien valmistaja vastaanottaa alihankkijalta ison erän laakerinkuulia, joiden halkaisijan pitäisi olla 5,30 millimetriä. Laakerien valmistaja haluaa tarkistaa, että vastaanotetut laakerinkuulat ovat sopivan kokoisia. Tätä varten asetetaan hypoteesit:

Nollahypoteesi: Kuulien halkaisijan keskiarvo on 5,30 millimetriä.
Vaihtoehtoinen hypoteesi: Kuulien halkaisijan keskiarvo on eri kuin 5,30 millimetriä.

Vastaanotetusta erästä valitaan 100 kappaleen otos. Otoksesta lasketaan halkaisijan keskiarvoksi 5,31 millimetriä ja keskihajonnaksi 0,10 millimetriä. Hylkäämisvirheen todennäköisyydeksi eli p-arvoksi saadaan noin 0,320. Näin suuri p-arvo (suurempi kuin 0,050) merkitsee sitä, että nollahypoteesi jää voimaan ja vastaanotettu laakerinkuulaerä voidaan hyväksyä.

Hylkäämisvirheen todennäköisyyden eli p-arvon laskentatapa vaihtelee käyttötilanteen (testattavan hypoteesin mukaan). Tilasto-ohjelmasta löytyy valmiit toiminnot p-arvojen laskentaan.

Määrälliset muuttujat

Yhtä muuttujaa koskeva päättely

Otantavirheestä johtuen otoskeskiarvot vaihtelevat satunnaisesti otoksesta toiseen. Keskiarvoon liittyvissä menetelmissä oletetaan, että otoskeskiarvojen jakauma on likimain normaalijakauma.

Riittävän suurilla otoksilla normaalijakaumaoletusta ei tarvitse erikseen tarkistaa, koska keskeisen raja-arvolauseen (central limit theorem) mukaan suurissa otoksissa otoskeskiarvot noudattavat normaalijakaumaa. Otosta voidaan pitää normaalijakaumaoletuksen kannalta riittävän suurena, jos otoskoko on vähintään 30.

Pienissä otoksissa normaalijakaumaoletus voidaan tarkistaa silmämäärin histogrammista ja laskennallisesti Kolmogorov-Smirnov -testillä tai Shapiro-Wilk -testillä. Tällöin tarkistetaan tarkkaan ottaen muuttujan arvojen normaalijakautuneisuus. Jos muuttujan arvot ovat normaalijakautuneet, niin sitä suuremmalla syyllä muuttujan arvoista lasketut otokeskiarvot ovat normaalijakautuneet.

Yhden muuttujan keskiarvolle voidaan laskea luottamusväli ja keskiarvoa voidaan testata yhden otoksen t-testillä.

Kahden ryhmän vertailu

Kahden ryhmän vertailussa on tärkeää tietää ovatko ryhmistä otetut otokset toisistaan riippumattomat vai pareittain toisiaan vastaavat (riippuvat).

Jos kyseessä ovat toisistaan riippumattomat otokset, niin kyseeseen tulee riippumattomien otosten t-testi. Riippumattomien otosten t-testillä testataan ovatko ryhmien keskiarvot yhtäsuuret. Tällöin kummankin ryhmän (otoksen) osalta oletetaan otoskeskiarvojen noudattavan normaalijakaumaa. Suurilla otoksilla (kummastakin ryhmästä vähintään 30) tätä ei tarvitse erikseen tarkistaa. Pienillä otoksilla normaalijakaumaoletus voidaan tarkistaa silmämäärin histogrammista ja laskennallisesti Kolmogorov-Smirnov –testillä tai Shapiro-Wilk –testillä. Jos normaalijakautuneisuutta ei voida olettaa, niin testaamiseen voidaan käyttää Mann-Whitney U-testiä.

Jos kyseessä ovat toisiaan pareittain vastaavat otokset, niin kyseeseen tulee riippuvien otosten t-testi. Riippuvien otosten t-testillä testataan onko toisiaan vastaavien parien erojen keskiarvo nolla. Tällöin oletetaan erojen keskiarvon otantajakaumaksi normaalijakauma. Suurilla otoksilla (pareja vähintään 30 kappaletta) tätä ei tarvitse erikseen tarkistaa. Pienillä otoksilla normaalijakaumaoletus voidaan tarkistaa silmämäärin erojen histogrammista ja laskennallisesti Kolmogorov-Smirnov –testillä tai Shapiro-Wilk –testillä. Jos normaalijakautuneisuutta ei voida olettaa, niin t-testin sijasta voidaan käyttää Wilcoxonin merkittyjen sijalukujen testiä.

Useamman ryhmän vertailu

Useamman toisistaan riippumattoman ryhmän vertailuun voidaan käyttää yksisuuntaista varianssianalyysiä. Jos ryhmistä otettujen otosten otoskeskiarvojen normaalijakautuneisuutta ei voida olettaa, niin varianssianalyysin sijasta voidaan käyttää Kruskal-Wallis –testiä.

Kahden muuttujan välinen riippuvuus

Kahden muuttujan riippuvuuden testaamiseen voidaan käyttää Pearsonin korrelaatiokertoimen testausta tai Spearmanin korrelaatiokertoimen testausta.

Kategoriset muuttujat

Kategoristen muuttujien kohdalla kyseeseen tulevat tilastollisen päättelyn menetelmät liittyvät lukumääriin ja niiden perusteella laskettuihin prosenttilukuihin.

Yhtä muuttujaa koskeva päättely

Yhtä muuttujaa tarkasteltaessa voidaan laskea prosenttiluvulle luottamusväli. Muuttujan sisällä voidaan tarkastella eri luokkiin kuuluvien prosenttiosuuksia ja testata noudattavatko ne jotain oletettua jakaumaa. Jos havainnot jaetaan kahteen luokkaan, niin kyseeseen tulee binomitesti. Useamman luokan kohdalla käytetään khiin neliö -yhteensopivuustestiä.

Ryhmien vertailu

Ryhmittelevän muuttujan määräämissä ryhmissä prosenttijakaumia voidaan vertailla khiin neliö -riippumattomuustestin avulla. McNemarin testillä voidaan tarkastellaan kaksiarvoista muuttujaa ennen ja jälkeen jonkin käsittelyn (esimerkiksi presidenttiehdokkaan kannatus ennen ja jälkeen vaalitentin). Tällöin testataan onko tapahtunut muutosta.

Kahden muuttujan välinen riippuvuus

Kahden kategorisen muuttujan riippuvuutta voidaan testata khiin neliö -riippumattomuustestin avulla. Vaikka testi lasketaan samalla tavalla kuin ryhmiä vertailtaessa, niin lähtökohta on erilainen. Ryhmien vertailussa ryhmittelevä muuttuja on yleensä selittävän muuttujan asemassa. Riippuvuutta tarkasteltaessa asetelma on korrelatiivinen ja molemmat muuttujat ovat keskenään samanlaisessa asemassa.

Mielipideasteikot

Melko yleisesti tasavälisiksi oletettujen mielipideasteikoiden kohdalla käytetään määrällisten muuttujien menetelmiä. Monien mielestä Mann-Whitney U-testi ja Wilcoxon merkittyjen sijalukujen testi sopivat t-testejä paremmin mielipideasteikolle. Samoin Kruskal-Wallis –testi sopii mielipideasteikolle varianssianalyysiä paremmin.

Tiekartta

Taulukossa on esitetty menetelmät, joita käsitellään tässä monisteessa. Mukaan on valittu vain laajasti käytettyjä ja vakiintuneita menetelmiä. Useamman muuttujan menetelmiä (useampia selitettäviä tai selittäviä muuttujia) ei käsitellä tässä monisteessa.

	Muuttujan mitta-asteikko	
Tarkoitus	Määrällinen	Kategorinen
Yhtä muuttujaa koskeva päättely	Normaalijakautuneisuuden testaaminen (Kolmogorov-Smirnov -testi ja Shapiro-Wilk -testi) Keskiarvon luottamusväli Keskiarvon testaus (t-testi)	Prosenttiluvun luottamusväli Prosenttiluvun testaus (binomitesti) Khiin neliö -yhteensopivuustesti
Kahden ryhmän vertailu	Kaksi riippumatonta otosta: -Riippumattomien otosten t-testi -Mann-Whitney U-testi Kaksi riippuvaa otosta: -Riippuvien otosten t-testi -Wilcoxon merkittyjen sijalukujen testi	Khiin neliö -riippumattomuustesti McNemar-testi (kaksiarvoinen muuttuja, riippuvat otokset)
Useamman ryhmän vertailu	Yksisuuntainen varianssianalyysi Kruskal-Wallis -testi	Khiin neliö -riippumattomuustesti
Kahden muuttujan välinen riippuvuus	Korrelaatiokertoimen testaus	Khiin neliö -riippumattomuustesti

2 YHTÄ MUUTTUJAA KOSKEVA PÄÄTTELY

2.1 Normaalijakautuneisuuden testaaminen

Normaalijakaumassa suurin osa arvoista sijoittuu keskiarvon läheisyyteen, symmetrisesti keskiarvon molemmille puolille.

Normaalijakauma määräytyy keskiarvon ja keskihajonnan perusteella. Keskiarvo määrää jakauman keskikohdan sijainnin ja keskihajonta määrää jakauman leveyden. Normaalijakaumassa noin 95 % tapauksista sijaitsee korkeintaan kahden keskihajonnan päässä keskiarvosta.

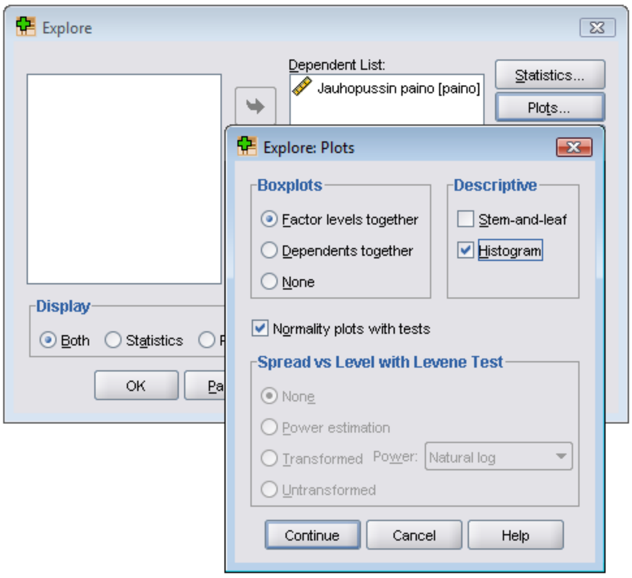
Hypoteesit

Normaalijakautuneisuuden testauksessa asetetaan hypoteesit:

Nollahypoteesi: Muuttuja noudattaa normaalijakaumaa.
Vaihtoehtoinen hypoteesi: Muuttuja ei noudata normaalijakaumaa.

SPSS ja normaalijakautuneisuuden testaaminen

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/jauho.sav.



1. Valitse **Analyze - Descriptive Statistics - Explore**.
2. Siirrä muuttujat, joiden normaalijakautuneisuutta haluat tarkastella, **Dependent List** -ruutuun.
3. Tarvittaessa voit siirtää **Factor List** -ruutuun kategorisen muuttujan, jonka mukaan jaat aineiston ryhmiin. Tällöin testaat normaalijakautuneisuutta erikseen kussakin ryhmässä.
4. Napsauta **Plots**-painiketta, jolloin aukenee **Explore: Plots** -valintaikkuna.
5. Valitse **Normality plots with tests** -ruutu. Kannattaa valita myös **Histogram**-ruutu histogrammin tulostamiseksi.
6. **Continue**.

Monien muiden tulosteiden ohella saat **Tests of Normality** -taulukon:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Jauhopussin paino	,156	20	,200*	,914	20	,076

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Sekä Kolmogorov-Smirnov -testi että Shapiro-Wilk -testi testaavat normaalijakautuneisuutta. Kuten nähdään yllä olevasta taulukosta, testien p-arvot voivat poiketa toisistaan. Jos molemmat testit johtavat samaan päätelmään nollahypoteesin suhteen, niin testien eroja ei tarvitse pohdiskella. Esimerkkimme tapauksessa nollahypoteesi jää molempien testien perusteella voimaan, koska p-arvot (**Sig.**) 0,200 ja 0,076 ovat suurempia kuin 0,050. Jos testit johtavat eri päätelmiin, niin asiaa kannattaa tarkastella kuvioiden avulla (esimerkiksi histogrammi). Kuvioiden tarkastelu on toki aina paikallaan. Epäselvissä tilanteissa kannattaa hylätä nollahypoteesi ja käyttää varmuuden vuoksi sellaisia testimenetelmiä, joissa normaalijakautuneisuutta ei tarvitse olettaa.

2.2 Keskiarvon luottamusväli

Pienillä otoksilla muuttujan normaalijakautuneisuus on syytä tarkistaa (katso 2.1 Normaalijakautuneisuuden testaaminen). Jos otoskoko on vähintään 30, niin tarkistusta ei tarvita.

Jos käytössä ei ole muuta tietoa kuin otoksesta laskettu keskiarvo, niin se on paras arvaus perusjoukon keskiarvoksi. Kun perusjoukon keskiarvo arvioidaan otoskeskiarvon suuruiseksi, niin arvioon liittyy epävarmuus. Epävarmuus ilmoitetaan virhemarginaalina. Yleensä ilmoitetaan 95 % virhemarginaali.

Esimerkki. Annostelukoneen pitäisi pussittaa 500 gramman pusseja. Pussin painon keskiarvo 20 pussin otoksessa on 480,3 grammaa ja keskihajonta 20,0 grammaa. Laskemalla saadaan 95 % virhemarginaaliksi noin 9,4 grammaa. Tavoitearvo 500 grammaa ei mahdu luottamusvälin 471 g - 490 g sisään, joten annostelukone on luultavasti väärin säädetty.

SPSS ja keskiarvon luottamusväli

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/jauho.sav.

1. Valitse **Analyze - Descriptive Statistics - Explore**.
2. Siirrä muuttujat, joista lasket keskiarvoja **Dependent List** -ruutuun.
3. Valitse **Display**-asetuksista **Statistics**.

Descriptives

			Statistic	Std. Error
Jauhopussin paino	Mean		480,30	4,483
	95% Confidence Interval for Mean	Lower Bound	470,92	
		Upper Bound	489,68	

Tulostaulukosta löydät muiden tunnuslukujen ohella 95 % luottamusvälin alarajan (**Lower Bound**) ja ylärajan (**Upper Bound**). Esimerkkitulosteessa 95 % luottamusväli jauhopussien painolle gramman tarkkuudella on 471 – 490 grammaa.

2.3 Keskiarvon testaus (t-testi)

Pienillä otoksilla muuttujan normaalijakautuneisuus on syytä tarkistaa (katso 2.1 Normaalijakautuneisuuden testaaminen). Jos otoskoko on vähintään 30, niin tarkistusta ei tarvita.

Hypoteesit

Jos käytössä on ennakko-oletus (nollahypoteesi), perusjoukon keskiarvosta, niin otoksen keskiarvoa voidaan verrata nollahypoteesin mukaiseen arvoon. Nollahypoteesi voi pohjautua esimerkiksi vallitsevaan käsitykseen, teoriaan, aikaisempaan tutkimukseen, valmistajan ilmoitukseen jne. Nollahypoteesin rinnalle asetetaan vaihtoehtoinen hypoteesi. Keskiarvon kaksisuuntaisessa testauksessa asetetaan hypoteesit (x_0 on jokin luku):

Nollahypoteesi: Perusjoukon keskiarvo on yhtä suuri kuin x_0.
Vaihtoehtoinen hypoteesi: Perusjoukon keskiarvo on eri suuri kuin x_0.

Jos ollaan kiinnostuneita vain poikkeamasta jompaankumpaan suuntaan, niin käytetään yksisuuntaista testiä. Tällöin vaihtoehtoinen hypoteesi on:

Vaihtoehtoinen hypoteesi: Perusjoukon keskiarvo on pienempi kuin x_0 (tai suurempi kuin x_0).

Esimerkki. Pullotuskone pitäisi olla säädetty siten, että se pullottaa 1/3 litran pulloja. Entuudestaan tiedetään, että pullojen sisältö vaihtelee normaalijakaumaa noudattaen. Laadun valvoja testaa toistuvilla otoksilla hypoteeseja:

Nollahypoteesi: Pullojen sisällön keskiarvo 1/3 litraa.
Vaihtoehtoinen hypoteesi: Pullojen sisällön keskiarvo eri suuri kuin 1/3 litraa.

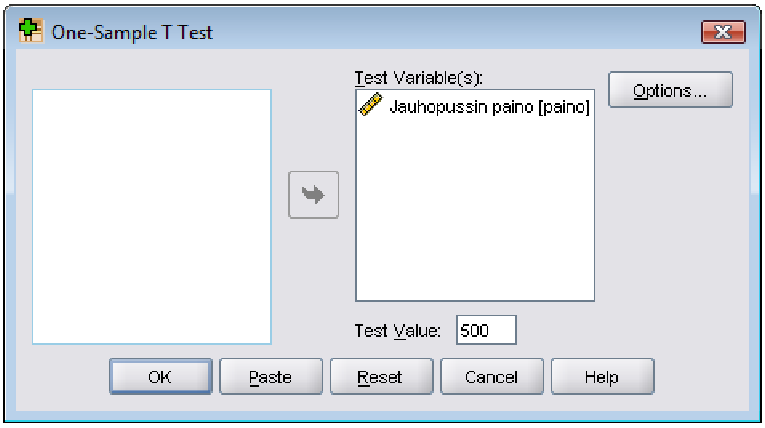
Laadun valvoja ottaa pullotuslinjalta 15 pullon otoksen ja saa sisältöjen keskiarvoksi 0,3420 litraa ja keskihajonnaksi 0,0115 litraa. Kaksisuuntaisen t-testin p-arvoksi saadaan noin 0,011. Nollahypoteesi hylätään, koska p-arvo on pienempi kuin 0,050. P-arvo voidaan tulkita myös riskiksi sille, että mahdollinen pullotuslinjan pysäyttäminen ja säätöjen korjaaminen tehdään turhaan.

SPSS ja keskiarvon testaus

Esimerkkinä käytetään SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/jauho.sav. Hypoteesit ovat:

Nollahypoteesi: Jauhopussien painon keskiarvo on 500 grammaa.
Vaihtoehtoinen hypoteesi: Jauhopussien painon keskiarvo on eri kuin 500 grammaa.

1. Valitse **Analyze - Compare Means - One-Sample T Test**.
2. Siirrä muuttuja, jonka keskiarvosta olet kiinnostunut, **Test Variable(s)**-ruutuun.
3. Kirjoita **Test Value** -ruutuun nollahypoteesin mukainen keskiarvo.



One-Sample Test

	Test Value = 500					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Jauhopussin paino	-4,395	19	,000	-19,700	-29,08	-10,32

Tulostaulukoista löydät kaksisuuntaisen testin p-arvon (**Sig**.). Yllä kaksisuuntaisen t-testin p-arvo kolmen desimaalin tarkkuudella on 0,000 eli selvästi pienempi kuin 0,050. Tässä tapauksessa nollahypoteesi (Jauhopussien painon keskiarvo 500 grammaa) hylätään.

Taulukosta löytyy myös 95 % luottamusväli jauhopussien painon keskiarvon ja nollahypoteesin erolle: -29,08 - -10,32.

Jos käytät yksisuuntaista testiä, niin p-arvo on puolet ilmoitetusta kaksisuuntaisen testin p-arvosta.

2.4 Prosenttiluvun luottamusväli

Prosenttiluvun luottamusväliä laskettaessa edellytyksenä on, että np ja n(1-p) ovat molemmat suuruudeltaan vähintään 10. Tämä tarkoittaa käytännössä sitä, että aineistossa esiintyy vähintään 10 onnistumista (jos onnistumisella tarkoitetaan sitä vaihtoehtoa, jonka prosenttiosuutta p tarkastellaan) ja vähintään 10 epäonnistumista.

Jos käytössä ei ole muuta tietoa kuin otoksesta laskettu prosenttiluku, niin se on paras arvaus perusjoukon prosenttiluvuksi. Kun perusjoukon prosenttiluku arvioidaan otoksesta lasketun prosenttiluvun suuruiseksi, niin arvioon liittyy epävarmuus. Epävarmuus ilmoitetaan virhemarginaalina. Yleensä ilmoitetaan 95 % virhemarginaali.

Esimerkki. Otoksesta (n=1800) laskettu viallisten tuotteiden osuus on 5,0 % ja virhemarginaali 1,0 prosenttiyksikköä. 95 % luottamusväli viallisten osuudelle on 4,0 % - 6,0 %.

Voit laskea virhemarginaalin helposti laskimella. Käytännön sovelluksissa saat 95 % virhemarginaalin riittävällä tarkkuudella laskemalla (p on otoksesta laskettu prosenttiluku desimaalimuodossa, n on otoskoko):

$$2 \times \sqrt{\frac{p(1-p)}{n}}$$

Esimerkki. Kyselytutkimuksen otoskoko n=800 henkilöä. Otoksesta 40,8 % (p=0,408) on uuden ydinvoimalan kannalla. Laskemalla saadaan virhemarginaaliksi noin 3,4 prosenttiyksikköä. Näin ollen 95 % luottamusväli ydinvoiman kannattajien osuudelle on 37,4 % - 44,2 %.

Otos on huomattava osa perusjoukosta

Jos otoskoko on yli 5 % perusjoukon koosta, niin voit kertoa virhemarginaalin äärellisen perusjoukon korjauskertoimella (N on perusjoukon koko, n on otoskoko):

$$\sqrt{\frac{N-n}{N-1}}$$

Korjauskertoimen käyttö tuottaa pienemmän virhemarginaalin.

2.5 Prosenttiluvun testaus (binomitesti)

Hypoteesit

Jos käytössä on ennakko-oletus (nollahypoteesi) perusjoukon prosenttiluvusta, niin otoksen prosenttilukua voidaan verrata nollahypoteesin mukaiseen arvoon. Nollahypoteesi voi pohjautua esimerkiksi olemassa olevaan teoriaan, vallitsevaan käsitykseen, aikaisempaan tutkimukseen, valmistajan ilmoitukseen jne. Nollahypoteesin rinnalle asetetaan vaihtoehtoinen hypoteesi. Prosenttiluvun kaksisuuntaisessa testauksessa asetetaan hypoteesit (P_0 on luku väliltä 0-100):

Nollahypoteesi: Perusjoukon prosenttiluku on yhtä suuri kuin P_0 %.
Vaihtoehtoinen hypoteesi: Perusjoukon prosenttiluku on eri suuri kuin P_0 %.

Jos ollaan kiinnostuneita vain poikkeamasta jompaankumpaan suuntaan, niin käytetään yksisuuntaista testiä. Tällöin vaihtoehtoinen hypoteesi on:

Vaihtoehtoinen hypoteesi: Perusjoukon prosenttiluku on pienempi (tai suurempi) kuin P_0 %.

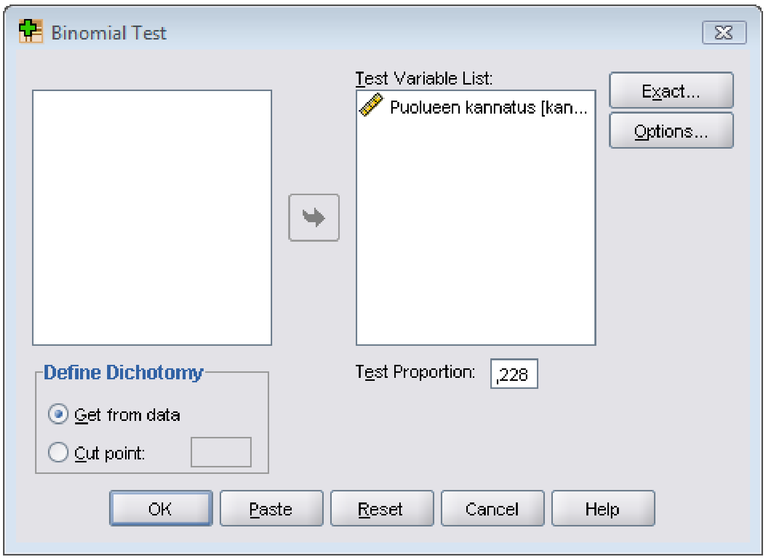
Esimerkki. Puolueen kannatus oli aiemmin 22,8 %. Tutkimuslaitos asetti seuraavat hypoteesit:

Nollahypoteesi: Puolueen kannatus on 22,8 %.
Vaihtoehtoinen hypoteesi: Puolueen kannatus on laskenut aiemmasta (pienempi kuin 22,8 %).

Satunnaisesti valitussa 800 henkilön otoksessa puolueen kannattajia oli 165. Yksisuuntaisen binomitestin p-arvoksi saadaan 0,076, joka on suurempi kuin 0,050. Nollahypoteesi jää voimaan Huomaa kuitenkin, että p-arvo on lähellä arvoa 0,050.

SPSS ja prosenttiluvun testaus (binomitesti)

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/kannatus.sav. Hypoteesit ovat edellisen esimerkin mukaiset.



1. Valitse **Analyze - Non-parametric Tests – Legacy Dialogs - Binomial**.
2. Siirrä **Test Variable List** -ruutuun muuttuja, josta olet kiinnostunut.
3. Kirjoita **Test Proportion** kohtaan nollahypoteesin mukainen prosenttiluku (,228 tarkoittaa 22,8 %).
4. Jos muuttuja on kaksiarvoinen (esimerkiksi kannattaa puoluetta/ei kannata puoluetta), niin valitse **Define Dichotomy** kohdasta **Get from data**. Jos muuttujalla on useampia mahdollisia arvoja, niin valitse **Cut point** ja kirjoita rajakohta, jota suurempia arvoja ei enää huomioida laskettaessa prosenttilukua.

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
Puolueen kannatus	Group 1	Kannattaa puoluetta	165	,206250	,228000	,076[a,b]
	Group 2	Ei kannata puoluetta	635	,793750		
	Total		800	1,000000		

a. Alternative hypothesis states that the proportion of cases in the first group < ,228000.

b. Based on Z Approximation.

Yllä olevan perusteella nollahypoteesi jää voimaan, koska yksisuuntaisen binomitestin p-arvo (**Sig.**) on 0,076 > 0,050.

Tärkeitä huomioita binomitestin käytöstä:

- SPSS ottaa kaksiarvoisen muuttujan tapauksessa testattavaksi luokaksi (Group 1) sen, jota esiintyy aineistossa ensimmäisenä. Jos haluat vaihtaa testattavaa luokkaa, niin järjestä aineisto (**Data – Sort Cases…**) sopivalla tavalla ennen binomitestin laskemista.
- Jos nollahypoteesissa esiintyy prosenttiluku 50 % (0,50), niin SPSS tulostaa kaksisuuntaisen testin p-arvon. Vastaavan yksisuuntaisen testi p-arvo on puolet kaksisuuntaisen testin p-arvosta.
- Jos nollahypoteesina esiintyy muu kuin 50 % (0,50), niin SPSS tulostaa yksisuuntaisen testin p-arvon (Vaihtoehtoinen hypoteesi: Testattavan ryhmän prosenttiluku pienempi kuin…). Vastaavan kaksisuuntaisen testi p-arvo on likimain kaksi kertaa yksisuuntaisen testin p-arvo (tämä pitää paikkansa sitä tarkemmin mitä isommasta otoksesta on kyse).

2.6 Khiin neliö -yhteensopivuustesti

Khiin neliö -yhteensopivuustestillä voidaan testata vastaako luokkien lukumäärien jakauma jotain oletettua jakaumaa. Khiin neliö -yhteensopivuustestin käyttöedellytyksenä on

- korkeintaan 20 % nollahypoteesin mukaisen jakauman lukumääristä on pienempiä kuin 5.
- nollahypoteesin mukaisen jakauman lukumäärät ovat suuruudeltaan vähintään 1.

Hypoteesit

Khiin neliö -yhteensopivuustestiin liittyvät hypoteesit ovat:

Nollahypoteesi: Jakauma on oletetun jakauman mukainen.
Vaihtoehtoinen hypoteesi: Jakauma ei ole oletetun jakauman mukainen.

Esimerkki. Yrityksen työntekijöiden koulutuksen tiedetään jakautuneen seuraavasti: peruskoulu 30 %, ammatillinen koulutus 30 %, korkeakoulututkinto 30 % ja ylempi korkeakoulututkinto 10 %. Jos halutaan tietää vastaako valitun otoksen koulutusjakauma kaikkien työntekijöiden koulutusjakaumaa, niin voidaan suorittaa khiin neliö -yhteensopivuustesti koulutusmuuttujalle.

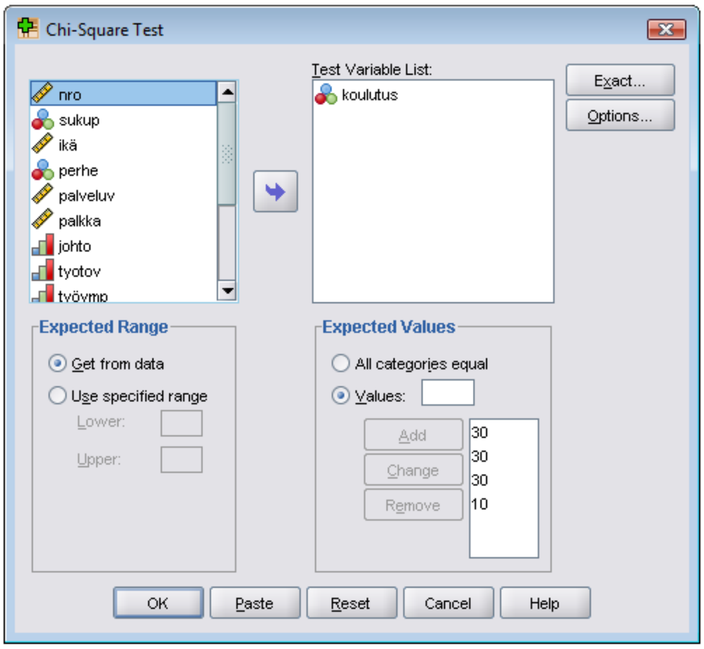
Nollahypoteesi: Koulutuksen jakauma on 30 %, 30 %, 30 %, 10 %.
Vaihtoehtoinen hypoteesi: Koulutuksen jakauma ei ole nollahypoteesin mukainen.

Khiin neliö -yhteensopivuustesti on kohdassa 2.5 kuvatun binomitestin yleistys tapaukseen, jossa luokkia on enemmän kuin kaksi.

SPSS ja khiin neliö -yhteesopivuustesti

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/data1.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

1. Valitse **Analyze – Nonparametric Tests – Legacy Dialogs – Chi-Square….**
2. Siirrä tarkasteltava muuttuja **Test Variable List:** -ruutuun.
3. Jos testaat sitä onko kaikkien luokkien prosenttiosuus sama, niin valitse **Expected Values** kohdasta **All categories equal**. Muussa tapauksessa valitse **Values:** ja kirjoita eri luokkien suhdeluvut yksi kerrallaan **Values:** ruutuun ja paina jokaisen suhdeluvun jälkeen **Add** painiketta.



Tuloksena saadaan kaksi taulukkoa. Ensimmäisessä taulukossa esitetään ryhmiin kuuluvien havaintojen lukumäärät (**Observed N**), nollahypoteesin mukaiset oletetut havaintojen lukumäärät (**Expected N**) sekä erot edellisten välillä (**Residual**). Varsinaisesta testitaulukosta löydät testin p-arvon (**Sig.**).

koulutus

	Observed N	Expected N	Residual
Peruskoulu	27	24,3	2,7
Ammatillinen	30	24,3	5,7
Korkeakoulututkinto	22	24,3	-2,3
Ylempi korkeakoulututkinto	2	8,1	-6,1
Total	81		

Test Statistics

	koulutus
Chi-Square	6,449[a]
df	3
Asymp. Sig.	,092

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 8,1.

Yllä khiin neliö -testin p-arvo on 0,092 on suurempi kuin 0,050, joten nollahypoteesi jää voimaan. Huomaa, että testin käyttöedellytykset ovat tarkistettavissa testitaulukon alaviitteestä:

- korkeintaan 20 % nollahypoteesin mukaisen jakauman lukumääristä on pienempiä kuin 5 (esimerkissämme 0 %).
- nollahypoteesin mukaisen jakauman lukumäärät ovat suuruudeltaan vähintään 1 (esimerkissämme 8,1).

3 KAHDEN RYHMÄN VERTAILU

Seuraavassa esitellään kuusi kahden ryhmän vertailuun soveltuvaa testiä:

- Riippumattomien otosten t-testi.
- Riippuvien otosten t-testi.
- Mann-Whitney U-testi, jota voidaan käyttää riippumattomien otosten t-testin sijasta, jos normaalijakautuneisuutta ei voida olettaa.
- Wilcoxon merkittyjen sijalukujen testi, jota voidaan käyttää riippuvien otosten t-testin sijasta, jos normaalijakautuneisuutta ei voida olettaa.
- Khiin neliö -riippumattomuustesti, joka soveltuu kahden riippumattoman otoksen prosenttilukujen vertailuun.
- McNemar-testi, jota käytetään dikotomisten muuttujien yhteydessä.

Riippumattomat vai riippuvat otokset?

Jos otetaan satunnaisotos kahdesta eri perusjoukosta, niin kyseessä on toisistaan riippumattomat otokset.

Esimerkki. Jos halutaan verrata kahdella eri menetelmällä valmistettujen lamppujen kestoikää, niin voidaan ottaa otos menetelmällä 1 valmistettuja lamppuja ja toinen otos menetelmällä 2 valmistettuja lamppuja.

Aineisto tallennetaan siten, että molemmilla menetelmillä valmistettujen lamppujen kestoiät ovat samassa sarakkeessa (muuttujassa). Ryhmittely toteutetaan kirjoittamalla valmistusmenetelmää kuvaava numero omaan sarakkeeseen (muuttujaan).

Myös saman satunnaisotoksen sisällä olevia ryhmiä voidaan pitää riippumattomina.

Esimerkki. Jos yrityksen työntekijöistä otetaan satunnaisotos, niin voimme pitää otokseen sisältyviä naisia ja miehiä toisistaan riippumattomina otoksina (otos naisista ja otos miehistä).

Jos esimerkiksi halutaan tehdä palkkavertailu, niin miesten ja naisten palkat ovat samassa sarakkeessa (muuttujassa). Ryhmittely miehiin ja naisiin toteutetaan kirjoittamalla sukupuolta kuvaava numero omaan sarakkeeseen (muuttujaan).

Jos toistetaan mittaus samoille tutkittaville, niin mittauskerrat muodostavat toistaan riippuvat otokset.

Esimerkki. Jos mitataan samojen kuluttajien asennetta tuotteeseen ennen ja jälkeen tuote-esittelyn, niin kyseessä ovat toisistaan riippuvat otokset. Asenne ennen tuote-esittelyä kirjoitetaan yhteen sarakkeeseen (muuttujaan) ja asenne tuote-esittelyn jälkeen toiseen sarakkeeseen (muuttujaan).

Toisistaan riippuvat otokset voidaan muodostaa myös käyttämällä toisiaan vastaavia pareja.

Esimerkki. Verrataan kahden akkutyypin kestoa matkapuhelimissa. Testiin valitaan useita matkapuhelinmalleja, kaksi kutakin. Kustakin matkapuhelinmallista muodostetaan pari, jotta päästän testaamaan kumpaakin akkutyyppiä kyseisessä matkapuhelinmallissa. Akkutyyppeihin liittyvät otokset ovat toisistaan riippuvat.

Ensimmäiseen akkutyyppiin liittyvät kestoiät kirjoitetaan omaan sarakkeeseen (muuttujaan) ja toiseen akkutyyppiin liittyvät kestoiät omaan sarakkeeseen (muuttujaan).

3.1 Riippumattomien otosten t-testi

Jos ryhmistä otetut otokset ovat pieniä, niin muuttujan normaalijakautuneisuus on syytä tarkistaa (katso 2.1 Normaalijakautuneisuuden testaaminen). Jos ryhmästä otettu otoskoko on vähintään 30, niin tarkistusta ei tarvita.

Hypoteesit

Kaksisuuntaisessa testissä asetetaan hypoteesit:

Nollahypoteesi: Ryhmien keskiarvot ovat samat.
Vaihtoehtoinen hypoteesi: Ryhmien keskiarvot poikkeavat toisistaan.

Jos ollaan kiinnostuneita vain poikkeamasta tiettyyn suuntaan, niin voidaan käyttää yksisuuntaista testiä. Tällöin vaihtoehtoinen hypoteesi on:

Vaihtoehtoinen hypoteesi: Toisen ryhmän keskiarvo on pienempi.

Esimerkki. Lamppujen valmistaja valmistaa samantyyppisiä lamppuja kahdella eri menetelmällä. Tutkimus- ja kehittämisosasto valitsee 40 lampun otoksen kummastakin menetelmästä ja mittaa lamppujen kestoiät. Hypoteeseina ovat:

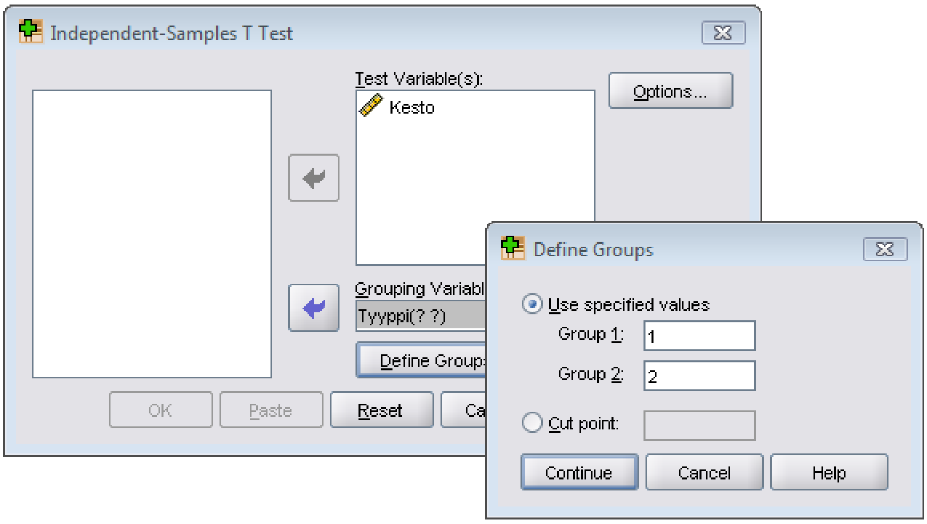
Nollahypoteesi: Kestoiän keskiarvo on sama molemmissa menetelmissä
Vaihtoehtoinen hypoteesi: Kestoiän keskiarvot ovat erisuuret eri menetelmissä.

Riippumattomien otosten kaksisuuntaisen t-testin p-arvoksi saadaan 0,006, joka on pienempi kuin 0,050. Tämän perusteella nollahypoteesi hylätään.

SPSS ja riippumattomien otosten t-testi

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/lamput.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

1. Valitse **Analyze - Compare Means - Independent-Samples T Test**.



2. Siirrä muuttujat, joiden keskiarvoista olet kiinnostunut, **Test Variable(s)** -ruutuun.
3. Siirrä muuttuja, jonka määräämissä ryhmissä haluat verrata keskiarvoja, **Grouping Variable** -ruutuun.
4. Ohjelma odottaa, että määrität **Define Groups** -painikkeella vertailtavat ryhmät. Jos esimerkiksi ryhmittelevä muuttuja on lampun valmistusmenetelmä ja olet merkinnyt menetelmiä numeroilla 1 ja 2, niin voit määrittä **Define Groups** -painikkeen takaa **Group 1** -ruutuun **1** ja **Group 2** -ruutuun **2**. Palaa edelliseen ikkunaan **Continue**-painikkeella.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Kesto	Equal variances assumed	1,772	,187	-2,846	78	,006	-170,0250	59,7506	-288,9792	-51,0708
	Equal variances not assumed			-2,846	75,210	,006	-170,0250	59,7506	-289,0488	-51,0012

Ennen t-testin p-arvon lukemista täytyy päättää, kumpaa taulukon riviä luetaan. Valinta tehdään Levene-testin avulla. Levene-testi on kaksisuuntainen testi, jonka hypoteesit ovat:

Nollahypoteesi: Perusjoukkojen varianssit yhtä suuret.
Vaihtoehtoinen hypoteesi: Perusjoukkojen varianssit erisuuret.

Jos Levene-testin p-arvo (**Sig**.) on alle 0,050, niin käytä '**Equal variances not assumed**' -riviä, muussa tapauksessa '**Equal variances assumed**' -riviä. Jos olet epävarma, niin voit käyttää '**Equal variances not assumed**' -riviä, koska sen käyttäminen ei koskaan ole väärin.

Esimerkkitulosteessa kaksisuuntaisen t-testin p-arvo 0,006 on pienempi kuin 0,050, joten nollahypoteesi hylätään. Jos käytät yksisuuntaista testiä, niin p-arvo on puolet kaksisuuntaisen testin p-arvosta.

3.2 Riippuvien otosten t-testi

Jos ryhmistä otetut otokset ovat pieniä, niin havaintoparien erotusten normaalijakautuneisuus on syytä tarkistaa (katso 2.1 Normaalijakautuneisuuden testaaminen). Jos otoskoko on vähintään 30, niin tarkistusta ei tarvita.

Hypoteesit

Kaksisuuntaisessa testissä asetetaan hypoteesit:

Nollahypoteesi: Ryhmien keskiarvot ovat samat.
Vaihtoehtoinen hypoteesi: Ryhmien keskiarvot poikkeavat toisistaan.

Jos ollaan kiinnostuneita vain poikkeamasta tiettyyn suuntaan, niin vaihtoehtoinen hypoteesi on:

Vaihtoehtoinen hypoteesi: Toisen ryhmän keskiarvo on pienempi.

Esimerkki. Lääkäri testaa erikoisruokavaliota potilaille, joiden suvussa esiintyy perinnöllistä taipumusta sydänsairauksiin. Erityisruokavalion tarkoituksena on alentaa painoa ja sydänsairauksien kannalta haitallisten triglyseridien määrää elimistössä. Potilaiden paino ja triglyseridi-arvot tutkitaan ennen ja jälkeen erityisruokavalion. Lääkäri asettaa hypoteesit:

Nollahypoteesi: Painon keskiarvo on erityisruokavalion jälkeen sama kuin aiemmin.
Vaihtoehtoinen hypoteesi: Painon keskiarvo on erityisruokavalion jälkeen pienempi.

Nollahypoteesi: Triglyseridi-arvojen keskiarvo on erityisruokavalion jälkeen sama kuin aiemmin.
Vaihtoehtoinen hypoteesi: Triglyseridi-arvojen keskiarvo ont erityisruokavalion jälkeen pienempi.

SPSS ja riippuvien otosten t-testi

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/dieetti.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

Ennen riippuvien otosten t-testin suorittamista on hyvä selvittää tarkasteltavien muuttujien normaalijakautuneisuus (**Analyze – Descriptive Statistics – Explore**, katso luku 2.1).

Tests of Normality

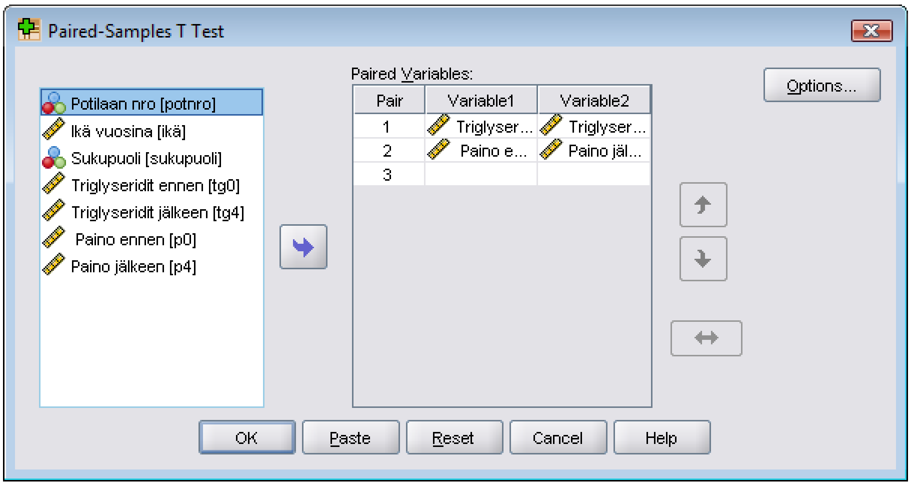
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Triglyseridit ennen	,194	16	,111	,933	16	,277
Triglyseridit jälkeen	,162	16	,200*	,941	16	,359
Paino ennen	,156	16	,200*	,938	16	,320
Paino jälkeen	,135	16	,200*	,949	16	,480

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Kolmogorov-Smirnov -testin ja Shapiro-Wilk -testin mukaan muuttujat voidaan olettaa normaalijakautuneiksi, koska kaikkien p-arvot (**Sig.**) ovat yli 0,050.

1. Valitse **Analyze - Compare Means - Paired-Samples T Test**.



2. Valitse vertailtava pari (ensimmäisen muuttujan valitset normaalisti ja toisen **ctrl-**näppäin alhaalla) ja siirrä pari **Paired variables** -ruutuun. Toista menettely kaikkien tarkasteltavien parien kohdalla.

Löydät testin p-arvon **Paired Samples Test** -taulukon **Sig**-sarakkeesta.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Triglyseridit ennen - Triglyseridit jälkeen	14,062	46,875	11,719	-10,915	39,040	1,200	15	,249
Pair 2	Paino ennen - Paino jälkeen	8,062	2,886	,722	6,525	9,600	11,175	15	,000

Triglyseridi arvot eivät ole alentuneet merkitsevästi (yksisuuntaisen testin p-arvo puolet kaksisuuntaisen testin p-arvosta eli 0,124 > 0,050), joten nollahypoteesi jää voimaan.

Paino sen sijaan on pudonnut merkitsevästi ($p = 0,000 < 0,050$), joten nollahypoteesi hylätään.

3.3 Mann-Whitney U -testi

Mann-Whitney U-testin käyttöedellytyksenä on, että muuttujan arvot ovat peräisin likimain samanmuotoisista jakaumista. Voit käyttää Mann-Whitney U-testiä riippumattomien otosten t-testin sijasta, jos epäilet t-testin käyttöedellytysten (normaalijakautuneisuus) toteutumista.

Hypoteesit

Oletetaan, että ollaan kiinnostuneita kahden ryhmän, esimerkiksi miesten ja naisten, erosta perusjoukossa. Mann-Whitney U-testissä tarkastellaan ryhmien jakaumien eroa. Kaksisuuntaisessa testissä asetetaan hypoteesit:

Nollahypoteesi: Ryhmien jakaumat ovat samat.
Vaihtoehtoinen hypoteesi: Ryhmien jakaumat poikkeavat toisistaan.

Jos ollaan kiinnostuneita vain poikkeamasta tiettyyn suuntaan, niin voidaan käyttää yksisuuntaista testiä. Tällöin vaihtoehtoinen hypoteesi on:

Vaihtoehtoinen hypoteesi: Toisen ryhmän jakauma sisältää suurempia arvoja.

Esimerkki. Yrityksen työntekijöistä kerätyn otoksen perusteella halutaan selvittää, onko naisten ja miesten palkoissa eroa.

Nollahypoteesi: Miesten ja naisten palkkajakaumat ovat samat.
Vaihtoehtoinen hypoteesi: Miesten ja naisten palkkajakaumien välillä on eroa.

SPSS ja Mann Whitneyn U-testi

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/data1.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

Riippumattomien otosten t-testin käyttö ei tule kyseeseen, koska naisten otos on pieni (n=19) ja normaalijakautuneisuuden testauksen perusteella naisten palkkajakaumaa ei kiistatta voi olettaa normaaliksi. (**Analyze – Descriptive Statistics – Explore**, katso luku 2.1).

Tests of Normality

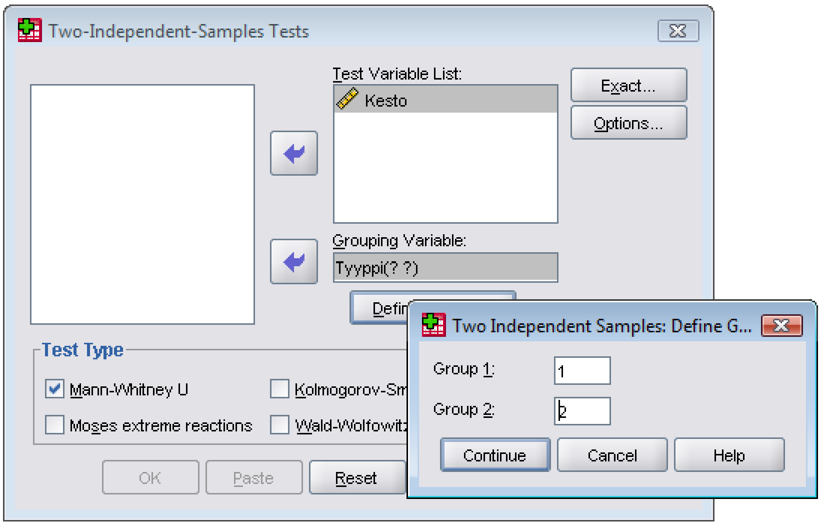
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	sukup	Statistic	df	Sig.	Statistic	df	Sig.
palkka	Mies	,168	63	,000	,838	63	,000
	Nainen	,174	19	,133	,894	19	,037

a. Lilliefors Significance Correction

Shapiro-Wilk -testi johtaa normaalijakautuneisuuden hylkäämiseen ($p = 0,037 < 0,050$).

1. Valitse **Analyze - Nonparametric Tests – Legacy Dialogs - 2 Independent Samples**.
2. Valitse **Test Variable List** -ruutuun muuttujat, joiden mediaaneja vertaat.
3. Valitse **Grouping Variable** -ruutuun muuttuja, jonka arvoista määräytyy vertailtavat ryhmät.

4. Ohjelma odottaa, että määrität **Define Groups** -painikkeella vertailtavat ryhmät. Jos esimerkiksi ryhmittelevä muuttuja on sukupuoli ja olet merkinnyt miehiä numerolla 1 ja naisia numerolla 2, niin voit määrittä **Define Groups** -painikkeen takaa **Group 1** -ruutuun **1** ja **Group 2** -ruutuun **2**. Palaa edelliseen ikkunaan **Continue**-painikkeella.



Ensimmäisestä tulostaulukosta voidaan lukea miesten ja naisten lukumäärät (**N**), palkkojen suuruusjärjestykseen perustuvien sijalukujen keskiarvot (**Mean Rank**) ja sijalukujen summat (**Sum of Ranks**). Sijalukujen keskiarvoista nähdään, että miehillä on keskimäärin isompia sijalukuja.

Ranks

	sukup	N	Mean Rank	Sum of Ranks
palkka	Mies	63	44,48	2802,50
	Nainen	19	31,61	600,50
	Total	82		

Test Statistics[a]

	palkka
Mann-Whitney U	410,500
Wilcoxon W	600,500
Z	-2,067
Asymp. Sig. (2-tailed)	,039

a. Grouping Variable: sukup

Toisesta tulostaulukosta voidaan lukea Mann-Whitney U-testin p-arvon (**Sig.**). Esimerkkimme tapauksessa nollahypoteesi hylätään (p = 0,039 < 0,050).

Jos käytät yksisuuntaista testiä, niin p-arvo on puolet kaksisuuntaisen testin p-arvosta.

SPSS:n versiosta 18 lähtien on kätevämpikin tapa testin laskemiseen. Tästä lisätietoa artikkelissani http://tilastoapu.wordpress.com/2012/03/08/mann-whitney-u-testi/

3.4 Wilcoxonin merkittyjen sijalukujen testi

Wilcoxonin käyttöedellytyksenä on, että parien väliset erot ovat jakautuneet likimain symmetrisesti. Voit käyttää Wilcoxonin merkittyjen sijalukujen testiä riippuvien otosten t-testin sijasta, jos epäilet t-testin käyttöedellytysten (normaalijakautuneisuus) toteutumista.

Hypoteesit

Kaksisuuntaisessa testissä asetetaan hypoteesit:

Nollahypoteesi: Parien välisten erojen mediaani on 0.
Vaihtoehtoinen hypoteesi: Parien välisten erojen mediaani on erisuuri kuin 0.

Jos ollaan kiinnostuneita vain poikkeamasta tiettyyn suuntaan, niin voidaan käyttää yksisuuntaista testiä. Tällöin vaihtoehtoinen hypoteesit on:

Vaihtoehtoinen hypoteesi: Parien välisten erojen mediaani on suurempi (tai pienempi) kuin 0.

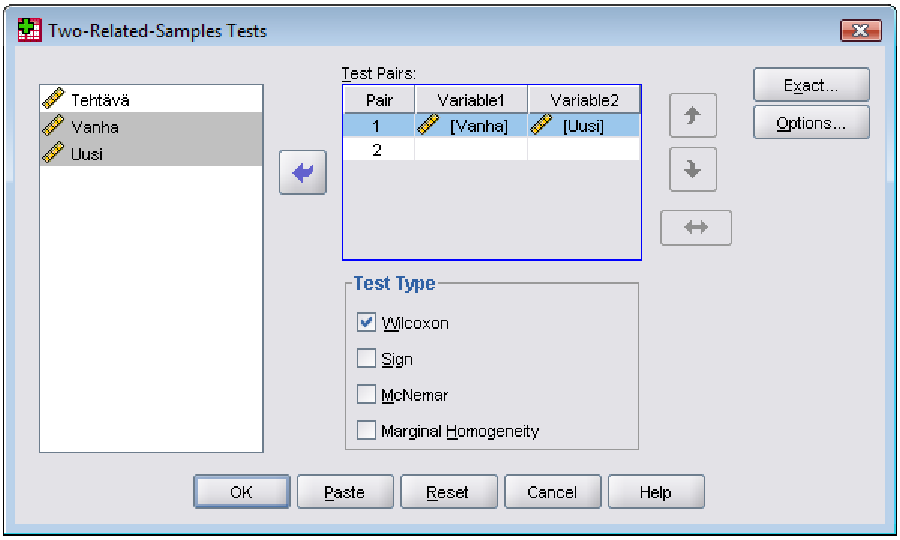
Esimerkki. Tietokoneohjelmien testaaja halusi tutkia onko uusi ohjelma nopeampi kuin vanha. Koska tietokoneohjelmalla suoritetaan erilaisia tehtäviä, niin testaaja arpoi ohjelmalle tyypillisten tehtävien joukosta 10 tehtävää. Kyseiset tehtävät suoritettiin kummallakin ohjelmalla ja suoritusajat mitattiin. Testaaja asetti hypoteesit:

Nollahypoteesi: Uuden ja vanhan ohjelman suoritusaikojen erojen mediaani on 0.
Vaihtoehtoinen hypoteesi: Uuden ja vanhan ohjelman suoritusaikojen erojen mediaani on pienempi kuin 0 (uusi ohjelma nopeampi).

SPSS ja Wilcoxonin merkittyjen sijalukujen testi

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/ohjelmat.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

1. Valitse **Analyze – Nonparametric Tests – Legacy Dialogs - 2 Related Samples….**
2. Valitse vertailtava pari (ensimmäisen muuttujan valitset normaalisti ja toisen **ctrl-**näppäin alhaalla) ja siirrä pari **Test Pair(s) List:** -ruutuun. Toista menettely kaikkien tarkasteltavien parien kohdalla (esimerkissämme ei ole kuin yksi pari).
3. Valitse testin tyypiksi **Wilcoxon**



Ensimmäisestä tulostaulukosta voidaan lukea kuinka monessa parissa uuden ohjelman suoritusaika oli pienempi kuin vanhan (8), uuden ohjelman suoritusaika oli suurempi kuin vanhan (1) ja kuinka monessa tapauksessa suoritusajat olivat samat (1).

Ranks

		N	Mean Rank	Sum of Ranks
Uusi - Vanha	Negative Ranks	8[a]	5,50	44,00
	Positive Ranks	1[b]	1,00	1,00
	Ties	1[c]		
	Total	10		

a. Uusi < Vanha

b. Uusi > Vanha

c. Uusi = Vanha

Test Statistics[b]

	Uusi - Vanha
Z	-2,549[a]
Asymp. Sig. (2-tailed)	,011

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Toisesta taulukosta löydetään Wilcoxon -testin kaksisuuntainen p-arvo. Vastaava yksisuuntaisen testin p-arvo on puolet kaksisuuntaisen testin p-arvosta. Esimerkissämme yksisuuntaisen testin p-arvo 0,006 on pienempi kuin 0,050, joten nollahypoteesi hylätään.

SPSS:n versiosta 18 lähtien on kätevämpikin tapa testin laskemiseen. Tästä lisätietoa artikkelissani http://tilastoapu.wordpress.com/2012/03/18/wilcoxon-merkittyjen-sijalukujen-testi/

3.5 Khiin neliö -riippumattomuustesti

Khiin neliö -riippumattomuustestillä voidaan verrata kahden ryhmän prosenttilukuja. Khiin neliö -riippumattomuustesti soveltuu myös useamman ryhmän vertailuun. Tämän vuoksi testin käyttö esitellään luvussa 4.3.

3.6 McNemar-testi

McNemar-testi on riippuvien otosten testi, joka sopii käytettäväksi kaksiarvoisten (dikotomisten) muuttujien kanssa.

Esimerkki. Asiakkailta kysyttiin valitsisivatko he tietyn pesuainemerkin. Promootion jälkeen samoilta asiakkailta kysyttiin valitsivatko he esitellyn pesuainemerkin. McNemar-testillä voidaan testata, onko promootio saanut aikaan muutosta mielipiteissä.

Hypoteesit

Nollahypoteesi: Ennen ja jälkeen tilanteen välillä ei eroa.
Vaihtoehtoinen hypoteesi: Ennen ja jälkeen tilanteilla on eroa.

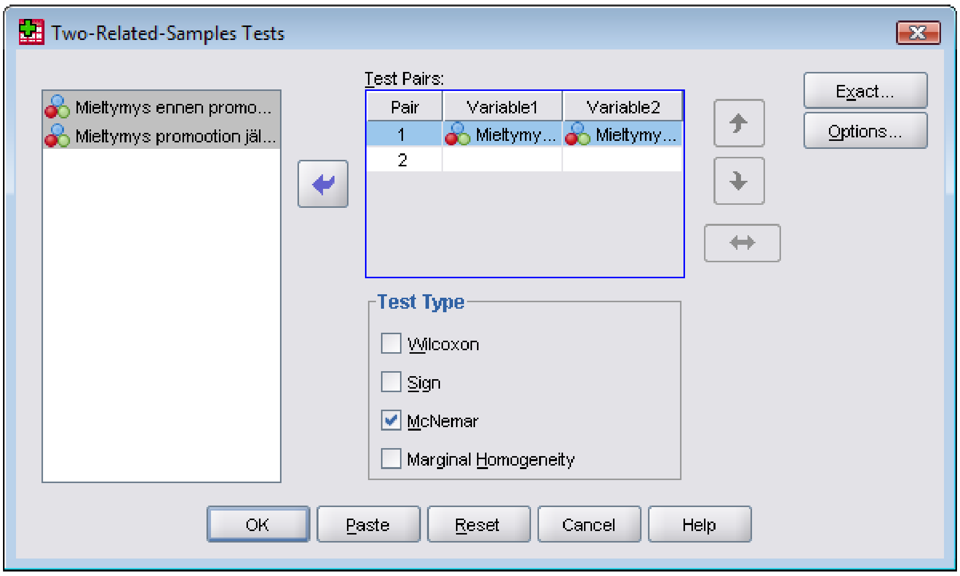
Esimerkki. Edellisessä esimerkissä hypoteeseina voisi olla:

Nollahypoteesi: Asiakkaiden mielipiteet eivät muuttuneet.
Vaihtoehtoinen hypoteesi: Asiakkaiden mielipiteet muuttuivat.

SPSS ja McNemar-testi

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/promootio.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

1. Valitse **Analyze – Nonparametric tests – Legacy Dialogs - 2- Related Samples…**
2. Valitse vertailtava pari (ensimmäisen muuttujan valitset normaalisti ja toisen **ctrl**-näppäin alhaalla) ja siirrä pari **Test Pair(s) List:** -ruutuun. Toista menettely kaikkien tarkasteltavien parien kohdalla (esimerkissämme ei ole kuin yksi pari).
3. Valitse testin tyypiksi **McNemar**.



Ensimmäisestä tulostaulukosta löydetään ostohalukkuudet muihin merkkeihin ja liikkeen omaan pesuainemerkkiin. Esimerkissämme ostohalukkuus näyttää kasvaneen ennen promootiota vallinneeseen tilanteeseen verrattuna. Taulukon mukaan 48 niistä, jotka ennen promootiota olisivat valinneet muun, valitseekin promootion jälkeen liikkeen oman merkin.

Mieltymys ennen promootiota & Mieltymys promootion jälkeen

Mieltymys ennen promootiota	Mieltymys promootion jälkeen: Muu	Mieltymys promootion jälkeen: Liikkeen oma merkki
Muu	64	48
Liikkeen oma merkki	26	59

Test Statistics[b]

	Mieltymys ennen promootiota & Mieltymys promootion jälkeen
N	197
Chi-Square[a]	5,959
Asymp. Sig.	,015

a. Continuity Corrected

b. McNemar Test

Testitaulukosta löydetään testin p-arvo (**Sig.**). Esimerkkimme tapauksessa McNemar testin p-arvo 0,015 on pienempi kuin 0,050, joten nollahypoteesi hylätään.

SPSS:n versiosta 18 lähtien on kätevämpikin tapa testin laskemiseen. Tästä lisätietoa artikkelissani http://tilastoapu.wordpress.com/2012/04/14/mcnemar-testi/

4 USEAMMAN RYHMÄN VERTAILU

4.1 Yksisuuntainen varianssianalyysi

Käyttöedellytykset

Yksisuuntaisen varianssianalyysin käyttöedellytykset ovat:

1. Otokset ovat toisistaan riippumattomia.
2. Ryhmistä otettujen otosten otoskeskiarvot noudattavat normaalijakaumaa.
3. Ryhmien varianssit ovat yhtä suuria.

Otokset ovat toisistaan riippumattomia

Jos kyseessä on asetelma, jossa vertailtavat ryhmät saavat tutkijan toimesta erilaiset käsittelyt, niin erilaisen käsittelyn saavat täytyy valita satunnaisesti samasta perusjoukosta.

Esimerkki. Jos kokeillaan kolmen eri oppimateriaalin vaikutusta oppimistuloksiin, niin kullekin oppimateriaalille valitaan käyttäjät satunnaisesti samasta perusjoukosta.

Jos kyseessä on asetelma, jossa verrataan ryhmiä, jotka ovat luonnostaan erilaisen "käsittelyn" saaneita (ilman tutkijan myötävaikutusta), niin tutkittavat täytyy valita satunnaisesti tietyn käsittelyn saaneista.

Esimerkki. Jos verrataan eri ikäluokkiin kuuluvien reaktionopeutta, niin kustakin ikäluokasta valitaan otokset satunnaisesti.

Ryhmistä otettujen otosten otoskeskiarvot noudattavat normaalijakaumaa

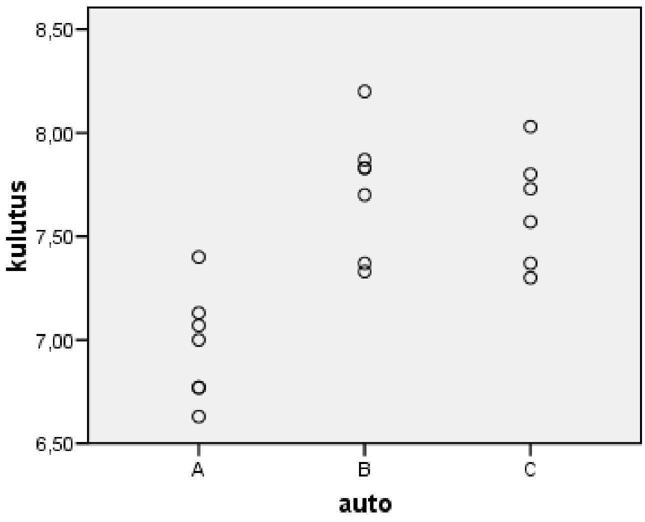
Jos ryhmistä otetut otokset ovat pieniä, niin muuttujan normaalijakautuneisuus on syytä tarkistaa (katso 2.1 Normaalijakautuneisuuden testaaminen). Jos ryhmästä otettu otoskoko on vähintään 30, niin tarkistusta ei tarvita.

Ryhmien varianssit ovat yhtä suuria

Riippuvan muuttujan täytyy omata likimain samansuuruiset varianssit (ja samalla keskihajonnat) kussakin tarkasteltavista ryhmistä. Jos kustakin ryhmästä valitaan samansuuruinen otos, niin pienet erot variansseissa eivät ole vakavia.

Esimerkki. Tarkastellaan kolmen eri automallin polttoaineenkulutusta. Selittävänä muuttujana on automalli. Arvotaan tietty määrä kuljettajia ajamaan kutakin automallia ja lasketaan kullekin automallille keskimääräinen polttoaineenkulutus.

A- ja B-autoilla oli kumpaisellakin 7 kuljettajaa ja C autolla 6 kuljettajaa. Polttoaineen kulutuksen vaihtelua voidaan havainnollistaa hajontakuviolla:



Kuviosta nähdään, että samallakin automallilla esiintyy kuljettajasta johtuvaa vaihtelua. Kuljettajasta johtuva vaihtelu on tässä tutkimusasetelmassa satunnaisvaihtelua, koska sitä ei ole millään tavalla kontrolloitu. Automallien erot ovat tässä tapauksessa niin suuria, että ne erottuvat kuljettajasta johtuvasta vaihtelusta huolimatta.

Yksisuuntaisella varianssianalyysilla pyritään tunnistamaan ryhmien välinen vaihtelu, joka erottuu satunnaisvaihtelusta. Ideana on kokonaisvarianssin jakaminen ryhmien väliseen varianssiin ja ryhmien sisäiseen varianssiin. Mitä suurempi ryhmien välinen varianssi on ryhmien sisäiseen varianssiin verrattuna, sitä todennäköisempää on, että riippumaton muuttuja on aiheuttanut vaihtelua.

Hypoteesit

Yksisuuntainen varianssianalyysin hypoteesit ovat:

Nollahypoteesi: Ryhmien keskiarvot ovat yhtä suuret.
Vaihtoehtoinen hypoteesi: Ainakin kahden ryhmän välillä on merkitsevä ero.

Esimerkki. Edellisen esimerkin hypoteesit voisivat olla:

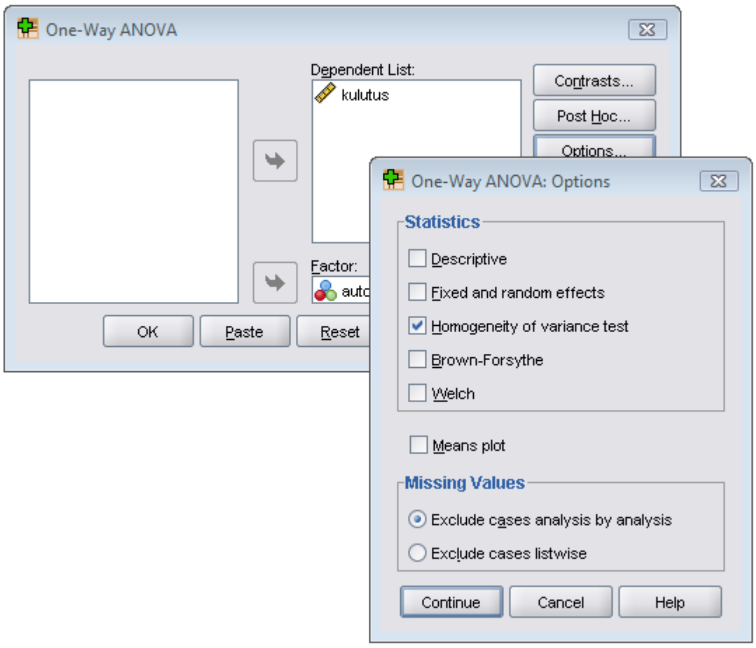
Nollahypoteesi: Autojen keskimääräisessä polttoaineen kulutuksessa ei ole eroja.
Vaihtoehtoinen hypoteesi: Vähintään yhden autoparin välillä on eroa keskimääräisessä polttoaineenkulutuksessa.

SPSS ja yksisuuntainen varianssianalyysi

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/kulutus.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

Oletetaan, että kullakin autolla polttoaineen kulutus noudattaa normaalijakaumaa (oletusta voi testata luvussa 2.1 kuvatulla testimenettelyllä).

1. Valitse **Analyze - Compare Means - One-Way ANOVA**:
2. Siirrä riippumaton muuttuja **Factor** ruutuun.
3. Siirrä riippuva muuttuja (muuttuja, josta lasketaan keskiarvot) **Dependent List:** ruutuun.
4. Napsauta **Options** painiketta ja valitse **Homogeneity of variance test**.



5. **Continue**.

Ensimmäisestä tulostaulukosta löydät Levene-testin varianssien yhtä suuruudelle (**Nollahypoteesi**: Varianssit ovat yhtä suuret). Varianssit voidaan testin perusteella olettaa yhtä suuriksi, koska p-arvo 0,972 on suurempi kuin 0,050.

Test of Homogeneity of Variances

kulutus

Levene Statistic	df1	df2	Sig.
,029	2	17	,972

ANOVA

kulutus

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2,389	2	1,195	15,079	,000
Within Groups	1,347	17	,079		
Total	3,736	19			

ANOVA taulukosta löydät testin p-arvon (**Sig.**). Yllä p-arvo on pienempi kuin 0,001, joten nollahypoteesi hylätään.

Vertailut

Edellä kuvattu testi ilmaisee ainoastaan onko joidenkin ryhmien välillä merkittävää eroa, mutta ei ilmaise minkä ryhmien välillä on merkittäviä eroja. Tarkastelua voidaan jatkaa vertailutesteillä, joiden avulla selvitetään minkä parien välillä on merkittäviä eroja. Jos ei tehdä etukäteisoletuksia eroja sisältävistä pareista, niin käytetään niin kutsuttuja Post Hoc -vertailuja. Tarjolla on useita vaihtoehtoisia menetelmiä Post Hoc -vertailujen tekemiseen. Tässä monisteessa ei oteta kantaa vaihtoehtoisten menetelmien vahvuuksiin ja heikkouksiin. Käytetään esimerkkinä yleisesti käytettyä Bonferroni-menetelmää (Napsauta yksisuuntaisen varianssianalyysin määrittelyikkunassa **Post Hoc…** painiketta ja valitse **Bonferroni**).

Multiple Comparisons

Dependent Variable: kulutus
Bonferroni

(I) auto	(J) auto	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
A	B	-,76571*	,15045	,000	-1,1652	-,3663
	C	-,66619*	,15659	,002	-1,0819	-,2504
B	A	,76571*	,15045	,000	,3663	1,1652
	C	,09952	,15659	1,000	-,3162	,5153
C	A	,66619*	,15659	,002	,2504	1,0819
	B	-,09952	,15659	1,000	-,5153	,3162

*. The mean difference is significant at the .05 level.

Edellä olevasta taulukosta selviää, että parin A-B välillä on eroa (p-arvo 0,000 < 0,050) samoin parin A-C välillä (p-arvo 0,002 < 0,050). Sen sijaan parin B-C välillä ei ole eroa (p-arvo 1,000 > 0,050).

4.2 Kruskall-Wallis -testi

Kruskall Wallis -testi sopii useamman toisistaan riippumattoman satunnaisesti valitun ryhmän vertailuun. Testiä voidaan käyttää yksisuuntaisen varianssianalyysin sijasta, jos normaalijakautuneisuutta tai varianssien yhtä suuruutta on syytä epäillä.

Kruskall Wallis -testin käyttöedellytyksenä on, että otokset ovat peräisin likimain samanmuotoisista jakaumista.

Hypoteesit

Nollahypoteesi: Ryhmien jakaumat ovat samanlaiset.
Vaihtoehtoinen hypoteesi: Ainakin kahden ryhmän välillä on merkitsevä ero.

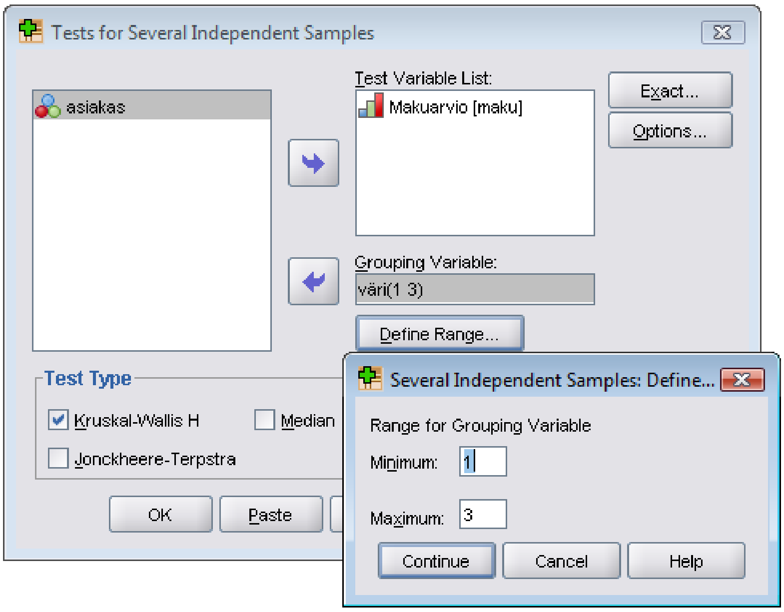
Esimerkki. Multa, jossa marjat kasvavat saattaa vaikuttaa marjojen makuun. Asiakkaita pyydettiin arvioimaan samaa lajiketta olevia marjoja, jotka olivat kasvaneet erilaisilla alustoilla (punainen, sininen ja musta multa). Marjojen makua arvioitiin 5-portaisella asteikolla.

Nollahypoteesi: Eri alustoilla kasvaneet marjat mielletään saman makuisiksi.
Vaihtoehtoinen hypoteesi: Ainakin kahden alustan välillä on makueroja.

SPSS ja Kruskall-Wallis -testi

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/maku.sav.
Hypoteesit ovat edellisen esimerkin mukaiset.

1. Valitse **Analyze – Nonparametric Tests – Legacy Dialogs - K Independent Samples….**
2. Siirrä ryhmittelevä muuttuja **Grouping Variable:** -ruutuun.
3. Ohjelma odottaa, että määrität **Define Range** -painikkeella vertailtavat ryhmät. Jos esimerkiksi vertailtavia ryhmiä on merkitty numeroilla 1-3, niin voit määrittä **Define Range** -painikkeen takaa **Minimum**-ruutuun **1** ja **Maximum**-ruutuun **3**. Palaa edelliseen ikkunaan **Continue**-painikkeella.
4. Siirrä tarkasteltavat muuttujat **Test Variable List:** -ruutuun.
5. Varmista, että testiksi on valittu **Kruskall-Wallis H**.



Ensimmäisestä taulukosta löytyy sijalukujen keskiarvo kullekin ryhmälle. Testitaulukosta löytyy p-arvo (**Sig.**). Esimerkkimme tapauksessa nollahypoteesi hylätään, koska Kruskall-Wallis testin p-arvo 0,008 on pienempi kuin 0,050.

Ranks

	Mullan ...	N	Mean Rank
Makuarvio	Punainen	10	9,05
	Sininen	10	16,75
	Musta	10	20,70
	Total	30	

Test Statistics[a,b]

	Makuarvio
Chi-Square	9,751
df	2
Asymp. Sig.	,008

a. Kruskal Wallis Test

b. Grouping Variable: Mullan väri

SPSS:n versiosta 18 lähtien on kätevämpikin tapa testin laskemiseen. Tästä lisätietoa artikkelissani http://tilastoapu.wordpress.com/2012/04/14/kruskal-wallis-testi/

4.3 Khiin neliö -riippumattomuustesti

Khiin neliö riippumattomuustestillä voidaan testata ryhmien välistä eroa kategorisilla muuttujilla. Khiin neliö -riippumattomuustestin käyttöedellytyksenä on,

- korkeintaan 20 % nollahypoteesin mukaisen jakauman lukumääristä on pienempiä kuin 5.
- nollahypoteesin mukaisen jakauman lukumäärät ovat suuruudeltaan vähintään 1.

Nollahypoteesin mukainen jakauma tarkoittaa teoreettista jakaumaa, jossa eroja ryhmien välillä ei esiinny.

Hypoteesit

Käytettäessä khiin neliö -riippumattomuustestiä ryhmien vertailuun hypoteesit ovat:

Nollahypoteesi: Ryhmien välillä ei ole eroa.
Vaihtoehtoinen hypoteesi: Ryhmien välillä on eroa.

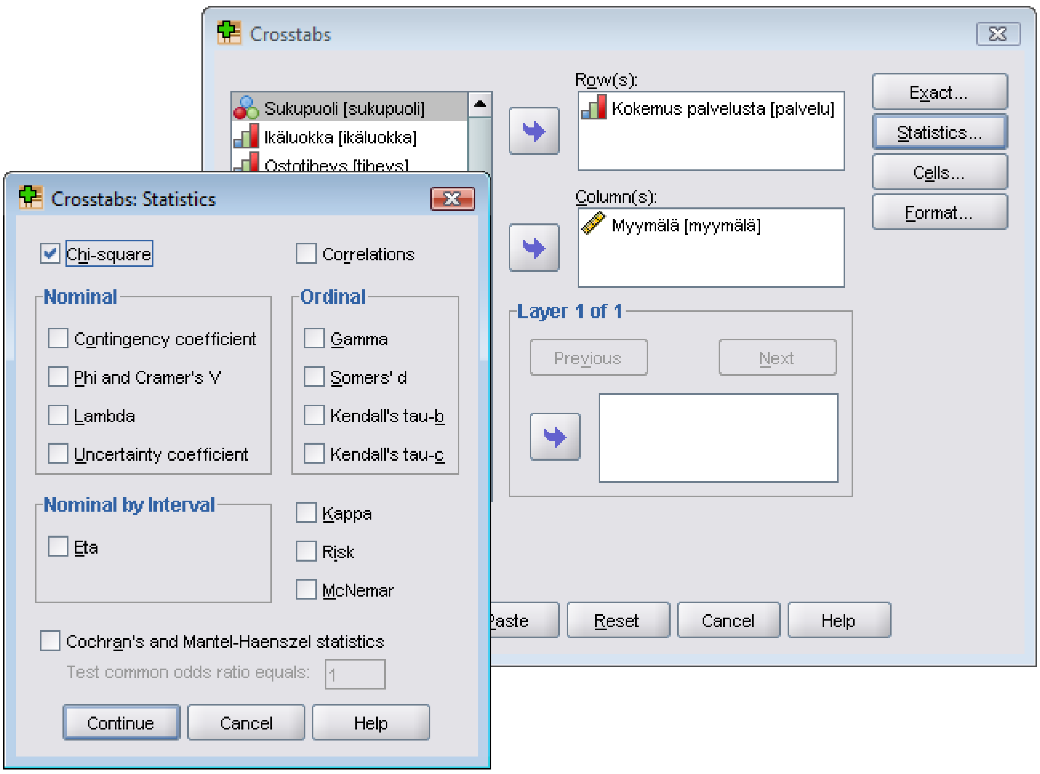
Esimerkki. Asiakkaat arvioivat neljän myymälän osalta kokemuksiaan palvelusta 5-portaisella asteikolla. Khiin neliö -riippumattomuustestillä voidaan testata, onko eri myymälöiden palvelu koettu erilaiseksi. Hypoteesit ovat:

Nollahypoteesi: Myymälöiden välillä ei eroja.
Vaihtoehtoinen hypoteesi: Myymälöiden välillä on eroja.

SPSS ja khiin neliö riippumattomuustesti

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/asiakas.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

1. Valitse **Analyze – Descriptive Statistics – Crosstabs…**.
2. Siirrä rivi- ja sarakemuuttujat paikoilleen.
3. Napsauta **Statistics** painiketta.
4. Valitse **Chi-square**.



5. **Continue**.

Varsinaisen ristiintaulukoinnin jälkeisestä khiin neliö testitaulukon alareunasta voimme tarkistaa edeltävyysehdot.

Kokemus palvelusta * Myymälä Crosstabulation

Count

		Myymälä				Total
		Myymälä 1	Myymälä 2	Myymälä 3	Myymälä 4	
Kokemus palvelusta	Erittäin negatiivinen	25	26	15	27	93
	Jossain määrin negatiivinen	20	30	20	35	105
	Neutraali	38	34	41	44	157
	Jossain määrin positiivinen	30	27	33	22	112
	Erittäin positiivinen	33	19	29	34	115
Total		146	136	138	162	582

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16,293[a]	12	,178
Likelihood Ratio	17,012	12	,149
Linear-by-Linear Association	,084	1	,772
N of Valid Cases	582		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 21,73.

Nyt testin käyttöedellytykset ovat voimassa:

- korkeintaan 20 % nollahypoteesin olettamista lukumääristä on pienempiä kuin 5 (esimerkissämme 0 %).
- nollahypoteesin mukaisen oletetun jakauman lukumäärät ovat suuruudeltaan vähintään 1 (esimerkissämme 21,73).

Khiin neliö -testin p-arvo 0,178 on suurempi kuin 0,050, joten nollahypoteesi jää voimaan.

Huomautus: Jos toinen muuttujista on mielipideasteikollinen kuten yllä olevassa esimerkissä, niin ryhmien välisen eron testaamiseen voidaan käyttää myös Kruskal-Wallis –testiä tai kahden ryhmän tapauksessa Mann-Whitney U –testiä. Kyseisten testien kohdalla ei niin helposti tule ongelmia käyttöedellytysten kanssa.

5 KAHDEN MUUTTUJAN VÄLINEN RIIPPUVUUS

5.1 Korrelaatiokertoimen testaus

Jos testaat Pearsonin korrelaatiokerrointa, niin pienillä otoksilla muuttujien normaalijakautuneisuus on syytä tarkistaa (katso 2.1 Normaalijakautuneisuuden testaaminen). Jos otoskoko on yli 30, niin tarkistusta ei tarvita.

Spearmanin korrelaatiokertoimen kohdalla normaalijakautuneisuus ei kuulu käyttöedellytyksiin.

Hypoteesit

Korrelaatiokertoimen kaksisuuntaisessa testauksessa asetetaan hypoteesit:

Nollahypoteesi: Perusjoukon korrelaatiokerroin on nolla.
Vaihtoehtoinen hypoteesi: Perusjoukon korrelaatiokerroin on nollasta poikkeava.

Jos korrelaatiokertoimen etumerkistä (+ vai -) on vahva ennakko-oletus, niin voidaan käyttää yksisuuntaista testausta. Tällöin vaihtoehtoinen hypoteesi on:

Vaihtoehtoinen hypoteesi: Perusjoukon korrelaatiokerroin on positiivinen (tai negatiivinen).

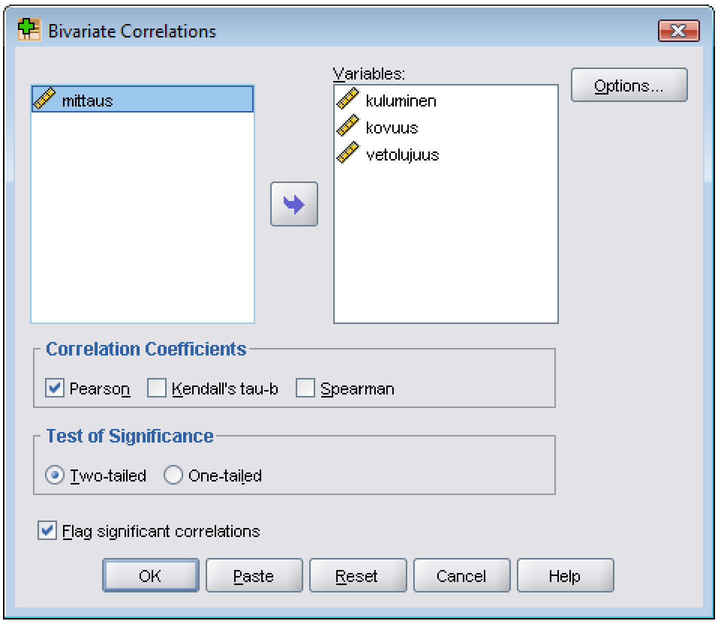
Esimerkki. Tutkittiin kumiseoksen vetolujuuden, kovuuden ja kulumisen välistä riippuvuutta. Mittaukset suoritettiin 30 kumiseokselle. Esimerkiksi vetolujuuden ja kulumisen osalta esitettiin hypoteesit:

Nollahypoteesi: Vetolujuuden ja kulumisen välinen korrelaatiokerroin on 0.
Vaihtoehtoinen hypoteesi: Vetolujuuden ja kulumisen välinen korrelaatiokerroin on eri suuri kuin 0.

SPSS ja korrelaatiokertoimen testaus

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/kumi.sav. Tutkimusasetelma on edellisen esimerkin mukainen.

1. Valitse **Analyze - Correlate – Bivariate**.
2. Siirrä muuttujat, joista lasket korrelaatioita **Variables:**-ruutuun.
3. Varmista, että valittuna on tilanteeseen sopiva korrelaatiokerroin (Pearson tai Spearman) ja testin suuntaisuus (1- vai 2-suuntainen).



Tulostaulukosta löydät korrelaatiokertoimien ohella p-arvot (**Sig.**).

Correlations

		kuluminen	kovuus	vetolujuus
kuluminen	Pearson Correlation	1,000	-,738**	-,298
	Sig. (2-tailed)		,000	,109
	N	30	30	30
kovuus	Pearson Correlation	-,738**	1,000	-,299
	Sig. (2-tailed)	,000		,108
	N	30	30	30
vetolujuus	Pearson Correlation	-,298	-,299	1,000
	Sig. (2-tailed)	,109	,108	
	N	30	30	30

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

Yllä p-arvo on alle 0,050 parissa kuluminen-kovuus (nollahypoteesi hylätään) ja yli 0,050 pareissa kuluminen-vetolujuus ja kovuus-vetolujuus (nollahypoteesi jää voimaan).

5.2 Khiin neliö -riippumattomuustesti

Khiin neliö -riippumattomuustestillä voidaan testata kahden muuttujan välistä riippuvuutta. Khiin neliö -riippumattomuustestin käyttöedellytyksenä on,

- korkeintaan 20 % nollahypoteesin mukaisen jakauman lukumääristä on pienempiä kuin 5.
- nollahypoteesin mukaisen jakauman lukumäärät ovat suuruudeltaan vähintään 1.

Nollahypoteesin mukainen jakauma tarkoittaa teoreettista jakaumaa, jossa riippuvuutta ei esiinny.

Hypoteesit

Khiin neliö -riippumattomuustesti lasketaan samalla tavalla kuin käytettäessä khiin neliö -riippumattomuustestiä ryhmien vertailuun. Ero on lähtökohdassa, joka näkyy hypoteeseissa. Riippuvuutta tarkasteltaessa hypoteesit ovat:

Nollahypoteesi: Muuttujien välillä ei ole riippuvuutta.
Vaihtoehtoinen hypoteesi: Muuttujien välillä on riippuvuutta.

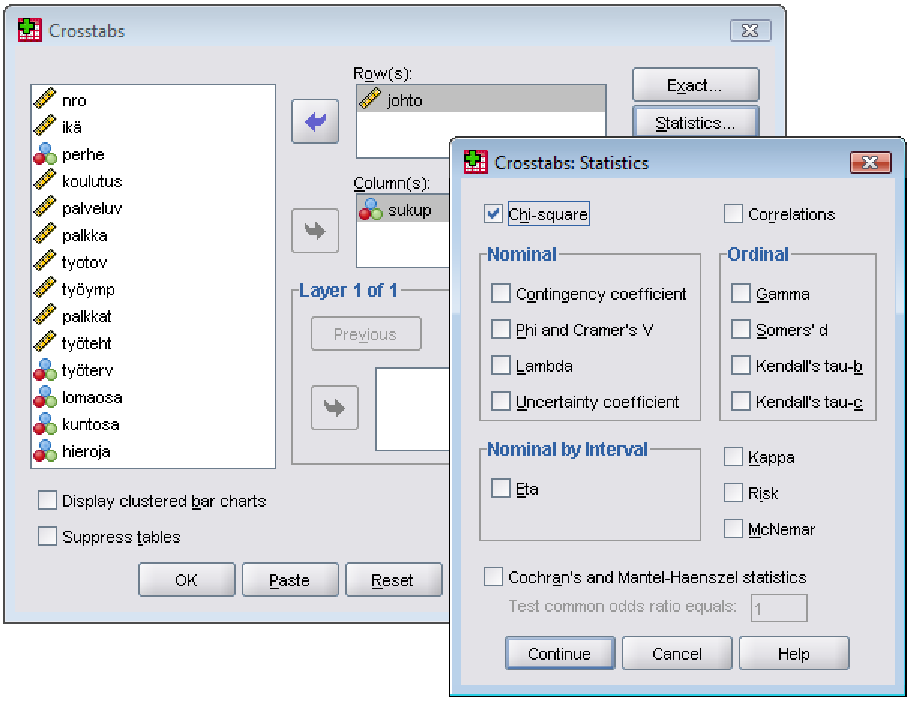
Esimerkki. Työntekijöistä otettiin satunnainen otos ja suoritettiin kyselytutkimus. Kyselyssä selvitettiin mm. vastaajan sukupuoli ja tyytyväisyys johtoon 5-portaisella tyytyväisyysasteikolla.

Nollahypoteesi: Sukupuolen ja tyytyväisyyden välillä ei ole riippuvuutta.
Vaihtoehtoinen hypoteesi: Sukupuolen ja tyytyväisyyden välillä on riippuvuutta.

SPSS ja khiin neliö -riippumattomuustesti

Esimerkkinä käytän SPSS-aineistoa http://myy.haaga-helia.fi/~taaak/p/data1.sav. Hypoteesit ovat edellisen esimerkin mukaiset.

1. Valitse **Analyze – Descriptive Statistics – Crosstabs…**.
2. Siirrä rivi- ja sarakemuuttujat paikoilleen.
3. Napsauta **Statistics** painiketta.
4. Valitse **Chi-square**.



5. **Continue**.

Varsinaisen ristiintaulukoinnin jälkeisestä khiin neliö -testitaulukon alareunasta voimme tarkistaa edeltävyysehdot.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8,853[a]	4	,065
Likelihood Ratio	10,663	4	,031
Linear-by-Linear Association	8,579	1	,003
N of Valid Cases	82		

a. 4 cells (40,0%) have expected count less than 5. The minimum expected count is 1,39.

Tässä tapauksessa huomataan, että testin käyttöedellytykset eivät täyty, koska 40 % (> 20 %) nollahypoteesin mukaisista lukumääristä on pienempiä kuin 5. Tyytyväisyyttä johtoon on mitattu 5-portaisella asteikolla (erittäin tyytymätön, tyytymätön, neutraali, tyytyväinen, erittäin tyytyväinen). Tyytyväisyysasteikko voidaan tiivistää 3-portaiseksi (tyytymätön, neutraali, tyytyväinen) käyttämällä SPSS:n uudelleenkoodaus toimintoa **Transform-Recode** tai toimintoa **Transform-Visual Binning**. Laskemalla khiin neliö testi uudelleenkoodauksen jälkeen saadaan seuraava testitaulukko.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8,127[a]	2	,017
Likelihood Ratio	9,459	2	,009
Linear-by-Linear Association	7,977	1	,005
N of Valid Cases	82		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 5,33.

Nyt testin käyttöedellytykset ovat voimassa:

- korkeintaan 20 % nollahypoteesin olettamista lukumääristä on pienempiä kuin 5 (esimerkissämme 0 %).
- nollahypoteesin mukaisen oletetun jakauman lukumäärät ovat suuruudeltaan vähintään 1 (esimerkissämme 5,33).

Khiin neliö -testin p-arvo 0,017 on pienempi kuin 0,050, joten nollahypoteesi hylätään.

6 TÄRKEITÄ HUOMIOITA

6.1 p-arvon tulkinta

Nollahypoteesin hylkäämisraja (esimerkiksi 5 %) on mielivaltainen. Ei ole käytännössä juurikaan eroa jos p-arvo on 4,8 % tai 5,1 %. Kuitenkin edellä esitetyn mekaanisen (ja mielivaltaisen) päättelysäännön mukaan p-arvo 4,8 % johtaa nollahypoteesin hylkäämiseen ja 5,1 % ei johda nollahypoteesin hylkäämiseen. Asia kannattaakin ymmärtää seuraavasti:

Mitä pienempi p-arvo sitä enemmän vaihtoehtoinen hypoteesi saa tukea.

Käytännön tilanne, johon testaamien liittyy, täytyy myös aina huomioida.

Esimerkki. Puolueen kannatuksen muutoksia selvitettäessä voidaan hyvinkin arvioida kannatuksen muuttuneen vaikka p-arvo olisikin suurempi kuin 5 %. Jos p-arvo on 5 % ja 10 % välillä, niin tulosta voidaan hyvinkin pitää suuntaa antavana sen puolesta, että kannatus on muuttunut.

Esimerkki. On tavallista, että valmistettujen tuotteiden laatua seurataan jatkuvasti otosten avulla. Tällöin nollahypoteesina on, että valmistusprosessi toimii kuten pitääkin ja tuotteet ovat ominaisuuksiltaan tavoitearvojen mukaisia. Jos otos antaa todisteita nollahypoteesia vastaan , niin riippuu toimintaympäristöstä miten tähän suhtaudutaan. Kyseessä on itse asiassa tärkeä päätöksentekotilanne, jonka vaihtoehtoina ovat:

	Todellinen tilanne	
Testauksen tulos	**Prosessi OK**	**Prosessissa jotain vialla**
Jatka tuotantoa	Oikea päätös	Hyväksymisvirhe
Pysäytä tuotanto	Hylkäämisvirhe	Oikea päätös

Jos tehdään hyväksymisvirhe, niin virheellinen tuotanto saa jatkua, mistä on tietenkin haitallisia seurauksia. Jos tehdään hylkäämisvirhe, niin tuotanto pysäytetään turhaan vian etsimistä varten ja tämä maksaa rahaa.

Päätöksentekijän täytyy löytää toimintaympäristöön sopiva p-arvo, jonka alittamien johtaa tuotannon pysäyttämiseen. Mitä kalliimpi hylkäämisvirhe on verrattuna hyväksymisvirheeseen sitä pienempää p-arvoa edellytetään nollahypoteesin hylkäämiseksi.

6.2 Tilastollinen merkitsevyys ja käytännön merkitsevyys

Hypoteesin testauksessa on tapana puhua tilastollisesta merkitsevyydestä. Yhden muuttujan testeissä kyse on esimerkiksi keskiarvon tai prosenttiluvun merkitsevästä erosta nollahypoteesiin verrattuna. Ryhmien vertailussa kyse on ryhmien välisten erojen merkitsevyydestä. Riippuvuuden testaamisesta kyse on riippuvuuden merkitsevyydestä. Tilastollisen merkitsevyyden ohella on syytä miettiä myös käytännön merkitsevyyttä.

Esimerkki. Oletetaan, että älykkyystestin maksimipistemäärä on 200.

Nollahypoteesi: Miehillä ja naisilla on sama keskiarvo.

Valitaan satunnaisesti otos miehiä ja otos naisia suorittamaan kyseinen älykkyystesti. Miesten ja naisten keksiarvopistemäärän eroksi saadaan 0,5 pistettä ja p-arvoksi saadaan

3 %. Tällöin nollahypoteesi hylätään ja miesten ja naisten ero on näin osoitettu tilastollisesti merkitseväksi. Voidaan kuitenkin oikeutetusti kysyä, onko 0,5 pisteen ero tässä asiassa käytännössä millään tavalla merkityksellinen?

Erityisesti isojen otosten kohdalla tilastollinen merkitsevyys saadaan usein osoitettua vaikka käytännön merkitsevyys on kyseenalainen. Myös toisin päin voi käydä. Käytännöllinen merkitsevyys voi vaikuttaa ilmeiseltä vaikka tilastollista merkitsevyyttä ei saada osoitettua esimerkiksi pienen otoskoon takia.

Järki siis täytyy aina säilyttää päässä ja lopullisen tilannearvion ja päätöksen tekee ihminen.

6.3 Normaalijakautuneisuus ja otoskoko 30

Monissa testeissä edellytetään otoskeskiarvojen normaalijakautuneisuutta. Pienillä otoksilla normaalijakautuneisuus on syytä tarkistaa. Otoskeskiarvojen jakaumaa ei päästä suoraan tarkastelemaan. Sen sijaan testataan otoksen perusteella muuttujan arvojen normaalijakautuneisuutta. Jos muuttujan arvot noudattava normaalijakaumaa niin myös otoskeskiarvojen voidaan olettaa noudattavan normaalijakaumaa.

Edellä on esitetty, että otoskoosta 30 ylöspäin testin käyttöedellytyksiin kuuluvaa normaalijakautuneisuutta ei tarvitse erikseen tarkistaa. Rajana pidetty otoskoko 30 ei ole mikään maaginen raja, vaan tilanteen mukaista harkintaa kannattaa käyttää.

Keskeisen raja-arvolauseen mukaan eri otoksista saatavien otoskeskiarvojen jakauma lähenee normaalijakaumaa otoskoon kasvaessa, riippumatta siitä minkälainen jakauma muuttujalla on perusjoukossa. Käytännössä on havaittu, että otoskoosta 30 ylöspäin ollaan useimmissa tapauksissa jo riittävän lähellä normaalijakaumaa. Jos muuttujan jakauma perusjoukossa on epätavallinen (erittäin vino, monihuippuinen, jne.), niin tarvitaan isompi otos normaalijakautuneisuuden takaamiseksi.