SISÄLTÖ

Edellisessä esimerkissä työssäkäyvien osuus 43,9 % on **kunnittaisten osuuksien painotettu keskiarvo** painoina väkiluvut. Kun painoja merkitään yleisesti w_i:llä ja lukuja, joiden keskiarvo lasketaan x_i:llä, niin painotettu keskiarvo on

$$\bar{x} = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + \ldots + w_n \cdot x_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum w_i x_i}{\sum w_i}$$

MEDIAANI

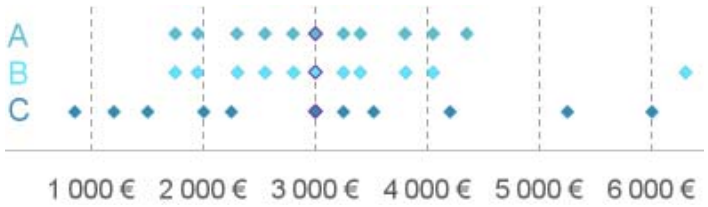
Mediaani, **Md**, on **suuruusjärjestykseen** järjestettyjen havaintoarvojen **keskimmäinen** arvo, jos arvoja on pariton määrä. Jos arvoja on parillinen määrä, mediaanina pidetään yleensä kahden lähinnä keskimmäisen arvon keskiarvoa, joskus toista kahdesta lähinnä keskimmäisestä arvosta. **Mediaani on siis arvo, jota pienempiä tai yhtä suuria on 50 %** havainnoista. Vastaavasti puolet havainnoista on suurempia (tai yhtä suuria) kuin mediaani. Mediaani voidaan määrittää, mikäli muuttujan arvot voidaan panna suuruusjärjestykseen eli muuttuja on **vähintään järjestysasteikon muuttuja**.

Esim. 3.3 Määritetään tammikuun 2010 viimeisen viikon lämpötilojen (esim. 3.1) mediaani.
Lämpötilat (C) ovat järjestyksessä pienimmästä suurimpaan

–19,3, –17,2, –16,9, **–13,6**, –10,6, –8,8, –3,7

Havaintoja on kaikkiaan seitsemän ja keskimmäinen arvo on –13,6, joten lämpötilan mediaani on –13,6 C.

Alla on kuvattu kolmen ryhmän palkkoja. Siitä käy ilmi, että mediaani ei huomioi sitä, miten suuria tai pieniä sen eri puolille sijoittuvat havaintoarvot ovat. Ryhmiä A ja B verrattaessa näkyy, miten yksi havainto vaikuttaa keskiarvoon, mutta ei mediaaniin.

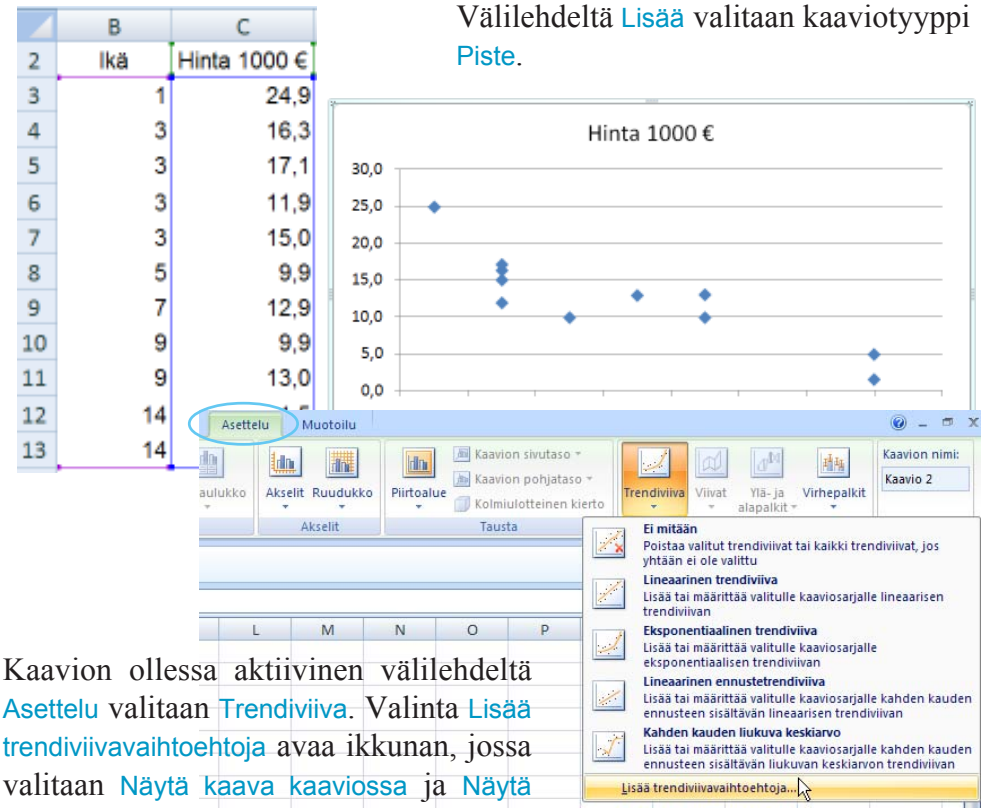


A: Md = 3 000 $\bar{x} \approx 3\,018$
B: Md = 3 000 $\bar{x} \approx 3\,195$
C: Md = 3 000 $\bar{x} \approx 3\,002$

Regressiomallin muodostaminen Excelillä

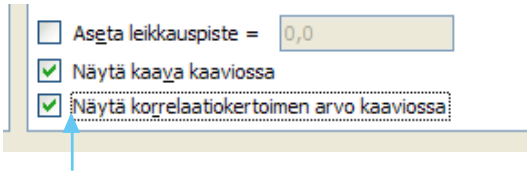
Regressiosuoran kertoimet voidaan laskea eri tavoin. Yhden selittävän muuttujan malli on havainnollisinta ja helpointa muodostaa hajontakaavion (Pistekaavio) yhteyteen. Tällöin kuvioon saadaan suoran kuvaaja, sen yhtälö ja selitysaste. Hajontakaaviossa (Piste-kaaviossa) selittävän muuttujan tulee sijaita x-akselilla ja selitettävän (riippuvan). muuttujan y-akselilla.

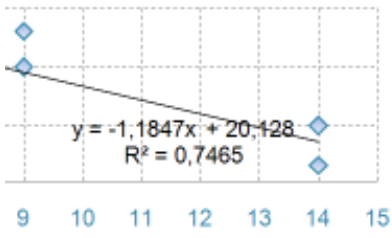
Esim. 4.14 Muodostetaan käytettyjen autojen hinnan riippuvuutta iästä kuvaavan regressiosuoran yhtälö hajontakaavioon. Kaaviossa selittävä muuttuja auton ikä on x-akselilla ja selitettävä eli riippuva muuttuja hinta y-akselilla. Valmiiseen kaavioon lisätään regressiosuora ja sen yhtälö.



Välilehdeltä Lisää valitaan kaaviotyyppi Piste.

Kaavion ollessa aktiivinen välilehdeltä Asettelu valitaan Trendiviiva. Valinta Lisää trendiviivavaihtoehtoja avaa ikkunan, jossa valitaan Näytä kaava kaaviossa ja Näytä korrelaatiokertoimen arvo kaaviossa.





Antaa nimityksestä huolimatta selitysasteen r^2.

3-26 Suomalaisten arvoja mittaavassa otantatutkimuksessa kysyttiin mm. miten hyväksyttävinä he pitävät eri asioita (asteikko 1–10, 1 = ei koskaan hyväksyttävä, ..., 10 = aina hyväksyttävä). Aineistosta saatiin seuraavat keskiarvot ja luottamusvälit (Lähde: Yhteiskuntatieteellinen tietoarkisto):

	Keskiarvo	95 %:n luottamusväli
Abortti	5,11	keskiarvo $\pm$ 0,19
Avioero	6,52	keskiarvo $\pm$ 0,17
Eutanasia	6,04	keskiarvo $\pm$ 0,20
Homoseksuaalisuus	4,54	keskiarvo $\pm$ 0,22
Prostituutio	3,54	keskiarvo $\pm$ 0,18

Havainnollista keskiarvot luottamusväleineen samassa kaaviossa.

3-27 Kuvaile tyypillinen Jyu:n opiskelija.

3-28 Laske Opiskelija-aineiston (Jyu) muuttujan Opintotuki
a) kvartiiliväli
b) kvartiilipoikkeama
ja selitä niiden antama informaatio.

3-29 Tee tunnuslukujen yhteenveto (tunnuslukutiivistelmä) Opiskelija-aineiston (Jyu) muuttujista Opintotuki ja Ansiotulot
a) ottamalla mukaan kaikki havainnot
c) ottamalla mukaan vai nollasta poikkeavat havainnot.
Vertaile jakaumia.

3-30 Laske Opiskelija-aineiston (Jyu) muuttujan Nukkuminen keskiarvot ja keskihajonnat
a) sukupuolittain
b) sukupuolittain parisuhteittain

3-31 Laske Opiskelija-aineiston (Jyu) muuttujien Opiskelu ja Liikunta keskiarvot, keskihajonnat ja variaatiokertoimet
a) sukupuolittain
b) tiedekunnittain.
Tulkitse tunnuslukujen antama informaatio.

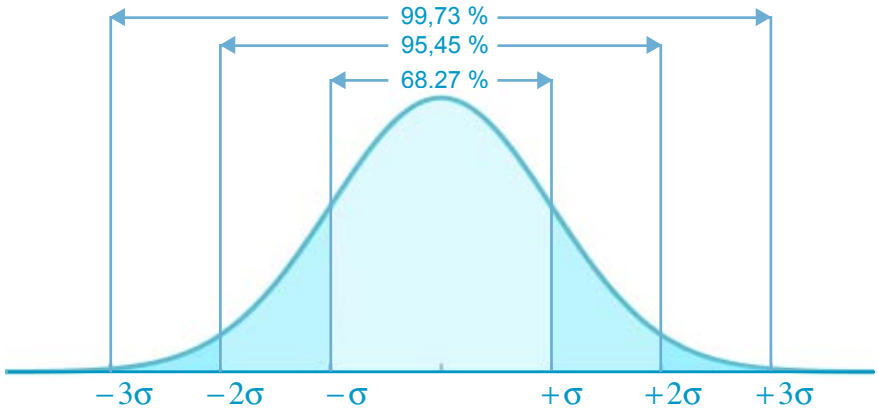
3-32 Havainnollista graafisesti Opiskelija-aineiston (Jyu) muuttujan Opiskelu keskiarvot tiedekunnittain. Piirrä kaavioon myös 95 %:n luottamusvälit.

3-33 Millaisia eroja mies- ja naisopiskelijoiden välillä näyttäisi olevan Opiskelija-aineiston (Jyu) perusteella
a) välivuosien lukumäärissä
b) suhtautumisessa hyviin harrastusmahdollisuuksiin?

Koska tiheysfunktion kuvaaja on symmetrinen odotusarvon suhteen, niin

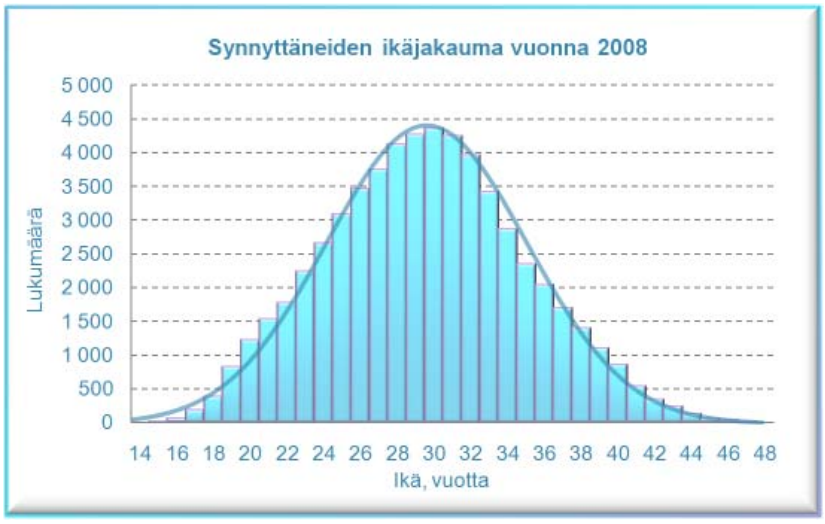
$$P(\underline{x} \leq \ ) = 0{,}5 \quad \text{ja} \quad P(\underline{x} \geq \ ) = 0{,}5$$

Todennäköisyysmassa on keskittynyt odotusarvon ympärille seuraavan kuvion mukaisesti:



Kuvion mukaan 68,27 % muuttujan arvoista poikkeaa odotusarvosta korkeintaan keskihajonnan verran suuntaan tai toiseen ja 99,73 % muuttujan arvoista on korkeintaan 3 keskihajonnan mitan päässä odotusarvosta.

Esim. 6.35 Synnyttäjien ikäjakauma vuonna 2008 (lähde: THL/ Syntymärekisteri) oli oheisen kaavion mukainen. Jakauma noudattaa likimain normaalijakaumaa siten, että odotusarvo on 30 vuotta ja keskihajonta 5,4 vuotta.



Histogrammiin on lisätty normaalijakauman kuvaaja.

Ensiksi tutkitaan varianssien yhtäsuuruutta. SPSS-tilasto-ohjelma antaa t-testin yhteydessä Levenen varianssien yhtäsuuruustestin tuloksen. Tässäkin testissä nollahypoteesina on, että varianssit ovat yhtäsuuret.

		Levene's Test for Equality of Variances	
		F	Sig.
Internet	Equal variances assumed	3,780	,052
	Equal variances not assumed		

Koska p-arvo (Sig.) on 0,052, niin nollahypoteesi jää voimaan (mikäli rajana pidetään 5 %:a) ja varianssit voidaan siis olettaa yhtä suuriksi (tällöin t-testin tulos luetaan ylemmältä riviltä).

T-testin tulosteet SPSS-tilasto-ohjelmalla:

t-test for Equality of Means						
					95% Confidence Interval of the Difference	
t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
2,565	473	,011	,4715	,1838	,1103	,8326
2,377	252,915	,018	,4715	,1983	,0809	,8620

Excelillä:

Kahden otoksen t-testi olettaen varianssit yhtäsuuriksi

	Mies	*Nainen*
Keskiarvo	3,006	2,535
Varianssi	4,649	2,986
Havainnot	155	320
t Tunnusluvut	2,565	
P(T<=t) kaksisuuntainen	0,011	
t-kriittinen kaksisuuntainen	1,965	

Koska p-arvo on 0,011, niin nollahypoteesi hylätään. Nais- ja miesopiskelijoiden internetin käytön keskiarvoissa esiintyvä ero on todellinen eikä ilmeisestikään johdu sattumasta. Koko opiskelijajoukossa miehet näyttävät käyttävän enemmän aikaa internetin selailuun kuin naiset.

Taulukosta käy lisäksi ilmi, että miesten ja naisten keskiarvojen erotus on noin 0,47 (Mean Difference). Erotuksen 95 %:n luottamusvälin alaraja on 0,11 ja yläraja 0,83, joten perusjoukossa miesten ja naisten keskiarvojen erotus on 95 %:n varmuudella välillä (0,11, 0,83).

Tavallisimpien testien valintataulukko

Verrataan ryhmien samanlaisuutta

Ryhmittelevä muuttuja luokitteluasteikollinen, 2 ryhmää

Verrattava muuttuja

- Luokitteluasteikollinen → χ^2-riippumattomuustesti
- Järjestysasteikollinen → χ^2-riippumattomuustesti; Mann-Whitneyn testi
- Välimatka- tai suhdeasteikollinen
  - Riippumattomat otokset
    - Normaali-jakautunut → Riippumattomien otosten t-testi*
    - Ei-normaali-jakautunut → Mann-Whitneyn U-testi
  - Riippuvat otokset
    - Normaali-jakautunut → Parittaisten otosten t-testi
    - Ei-normaali-jakautunut → Wilcoxonin testi

Ryhmittelevä muuttuja luokitteluasteikollinen, ryhmiä 3 tai enemmän

Verrattava muuttuja

- Luokitteluasteikollinen → χ^2-riippumattomuustesti
- Järjestysasteikollinen → χ^2-riippumattomuustesti; Kruskal-Wallisin testi
- Välimatka- tai suhdeasteikollinen
  - Riippumattomat otokset
    - Normaali-jakautunut → Varianssi-analyysi
    - Ei-normaali-jakautunut → Kruskal-Wallisin testi

Tarkastellaan kahden muuttujan riippuvuutta/korrelaatiota

- Ainakin toinen muuttuja luokitteluasteikollinen → χ^2-riippumattomuustesti
- Toinen muuttuja järjestysasteikollinen, toinen järjestys-, välimatka- tai suhdeast. → (Spearmanin) järjestyskorrelaatio-kertoimen merkitsevyystesti
- Molemmat muuttujat välimatka- tai suhdeast. → Pearsonin korrelaatiokertoimen merkitsevyystesti

* t-testi sen mukaan ovatko varianssit yhtä suuret vai eri suuret