

[MTTTP1] TILASTOTIETEEN JOHDANTOKURSSI, Kevät 2019

<https://coursepages.uta.fi/mtttp1/kevat-2019/>

HARJOITUS 3 viikko 13

Ryhmät:

ke	08.30–10.00	LS C8	Härkönen	to	12.15–13.45	LS C6	Paajanen
ke	10.15–11.45	LS C8	Harju	to	14.15–15.45	LS C6	Paajanen
ke	12.15–13.45	LS C8	Harju	pe	08.30–10.00	LS C8	Korhonen
to	08.30–10.00	LS C8	Härkönen	pe	10.15–11.45	LS C8	Korhonen

Aiheet: tunnusluvut, ehdolliset tunnusluvut, riippuvuus, korrelaatiokerroin

- Viiden kynttilän palamisajat tunteina ovat 8, 9, 6, 7, 10. Standardoi palamisajat ja laske standardoidun muuttujan keskiarvo ja keskihajonta. Voit halutessasi käyttää keskiarvojen ja keskihajontojen laskennassa esim. sivulta <http://vassarstats.net> löytyvää laskuria <http://vassarstats.net/basic.html> tai SPSS-ohjelmaa.
- Tarkastellaan harjoitusten 1 tehtävän 4 a) aineistoa, josta tehdyn analyysin tuloksia on ohessa. Aineistossa on nainen, joka on osallistunut tavalliseen luento-opetukseen ja jonka testipisteiden erotus on -9,4. Aineistossa on miesopiskelija, joka on osallistunut TV:n kautta tapahtuvaan opetukseen ja hänen testipisteiden erotus on -7,8. Kumpi opiskelija on menestynyt suhteellisesti huonommin ryhmäläisiinsä verrattuna?

Descriptive Statistics

Dependent Variable: Piste-erotus

Opetustapa	Sukupuoli	Mean	Std. Deviation	N
Tavallinen	Nainen	14,4583	11,82505	12
	Mies	13,2471	15,20671	17
	Total	13,7483	13,69094	29
TV	Nainen	17,0583	8,44915	12
	Mies	17,1000	11,66660	37
	Total	17,0898	10,88311	49
Total	Nainen	15,7583	10,13813	24
	Mies	15,8870	12,86560	54
	Total	15,8474	12,02649	78

- Harjoitusten 2 tehtävässä 4 on hirmumyrskyjen lukumäärän jakauma. Kuinka monta prosenttia havainnoista on korkeintaan yhden hajonnan päässä keskiarvosta?

Voit halutessasi käyttää keskihajonnan laskussa jotain ohjelmaa tai esim. tehtävässä 1 mainittua laskuria. Aineisto saatavana

<http://www.sis.uta.fi/tilasto/mtttp1/syksy2018/hirmumyrskyt.xlsx>

4. Tutkitaan kolmen autotyypin polttoaineen kulutusta (kulutus=mailit/gallona) huomioiden kuljettajan ikä (ikäryhmiä 5) ja saadaan oheiset kulutuksen keskiarvot. Piirrä tuloksia havainnollistava graafinen esitys ja tee johtopäätelmät sen perusteella. Sopivan graafisen esityksen voit mieltä itse tai soveltaa luentomonisteen esimerkin 5.1.22 kuviota (ks. myös <http://davidmlane.com/hyperstat/B111146.html>).

Ikäryhmä	Auto		
	A	B	C
1	25,2	24,1	25,7
2	24,7	23,7	25,2
3	26,2	24,8	25,7
4	24,3	23,9	23,7
5	23,9	24,3	25,2

5. Järjestettiin koetilanne, jossa professori halusi tutkia sitä, miten voisi saada selville onko opiskelija kirjoittanut tehtävänsä ratkaisun itse. Käytettiin kahta ryhmää, joissa toinen ryhmä suoritti tehtävän kirjoittamisen itse ja toinen ryhmä sai kopioida ratkaisujaan toiselta opiskelijalta. Tehtävien palautuksen jälkeen järjestettiin viikon kuluttua koetilanne, jossa jokaisen opiskelijan tuli täyttää omasta tehtävästään professorin peittämät kohdat. Tämän jälkeen kirjattiin jokaisesta vastauksesta virheiden lukumäärät, jotka olivat ryhmittäin:

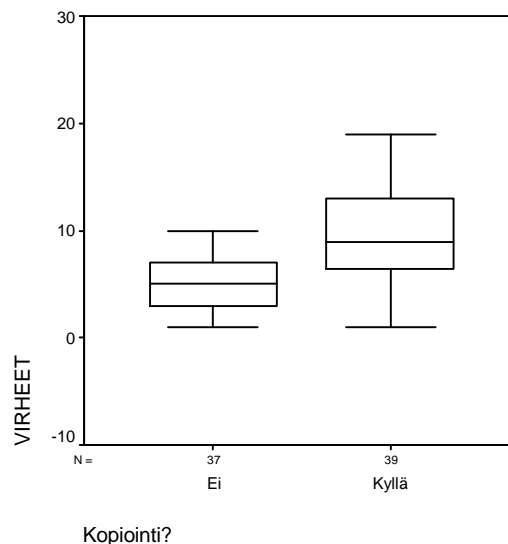
ei käytetty kopiointi

1, 1, 1, 1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 8, 8, 9, 9, 9, 10, 10, 10

käytetty kopiointia

1, 2, 3, 4, 4, 4, 5, 6, 6, 6, 7, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10, 10, 10, 11, 12, 13, 13, 13, 13, 14, 14, 15, 17, 17, 18, 19, 19.

Aineistosta on muodostettu laatikko-jana-kuvio:



- a) Määritä virheiden lukumäärän ehdolliset keskiarvot ja mediaanit.
- b) Tee johtopäätelmät tulosten perusteella.

6. Tutkittiin erään aineiston (n = 315) perusteella miesten ja naisten tupakointia ja saatiin oheinen ristiintaulukko, jossa on selitettävän muuttujan ehdolliset prosenttijakaumat.

		sukupuoli		Total
		mies	nainen	
tupakointi	ei koskaan polttanut	30,95	52,75	49,84
	entinen tupakoija	52,38	34,07	36,51
	nykyinen tupakoija	16,67	13,19	13,65
Total		100,00	100,00	100,00

Tulkitse tulokset. Laske lisäksi frekvenssit ristiintaulukkoon.

7. Tutkitaan nuorten keskuudessa sitä, onko vanhempien tupakoinnilla vaikutusta lastensa tupakointiin. Saadaan oheinen ristiintaulukko.

	Nuori polttaa	Nuori ei polta	Yhteensä
Molemmat vanhemmat polttavat	400	1380	1780
Toinen vanhemmista polttaa	416	1823	2239
Vanhemmat eivät polta	188	1168	1356
Yhteensä	1004	4371	5375

Määritä selitettävän muuttujan ehdolliset prosenttijakaumat. Tee johtopäätelmät.

8. a)
Piirrä oheisesta aineistosta muuttujien välinen pisteparvi. Onko riippuvuus lineaarista? Jos on, niin laske myös korrelaatiokerroin. Jos olisit vahingossa käyttänyt ensimmäisen puun kohdalla viisivuotiskasvun arvona lukua 700, niin miten korrelaatiokerroin olisi muuttunut?

Voit käyttää piirtämisessä ja laskennassa esim. SPSS –ohjelmaa tai sivulla <http://www.socscistatistics.com/tests/pearson/Default.aspx> olevaa laskinta.

Aineisto:

puun ikä vuosina: 5, 9, 9, 10, 10, 11, 11, 12, 13, 13, 14, 14, 15, 15, 18, 18
viisivuotiskasvu: 70, 150, 260, 230, 255, 165, 225, 340, 305, 335, 290, 340, 225, 300, 380, 400

Aineisto myös osoitteessa <http://www.sis.uta.fi/tilasto/tiltp1/syksy2004/h5t3.xls>

b)

Riippuuko y lineaarisesta x:stä oheisessa aineistossa? Jos ei, niin riippuuko y jollain muulla tavalla x:stä?

Aineisto:

x: -1, -2, -7, -8, 1, 2, 3, 4, 5, 6, 7, 8
y: 2, 5, 100, 120, 2, 5, 20, 30, 50, 70, 100, 120

Aineisto myös osoitteessa <http://www.sis.uta.fi/tilasto/tiltp1/syksy2004/h5t3b.xls>

9. Tallenna harjoitustyöaineistosi havaintomatriisi (muokkaa aineistosi tarvittaessa ennen tallennusta havaintomatriisimuotoon). Tallenna analyyseissä käyttämälläsi ohjelmistolla tai sellaisessa muodossa, että ohjelmistosi pystyy lukemaan sen.
10. Muodosta harjoitustyöaineistostasi muuttujien jakaumat. Käytä tilanteeseen sopivasti joko taulukoita tai graafisia esityksiä. Huomaa mahdolliset tallennusvirheet ja korjaa ne. Laske tarpeelliset tunnusluvut. Kirjoita harjoitustyösi liitteeksi vaadittava muuttujaluettelo (ks. harjoitustyön teko-ohjeet <http://www.sis.uta.fi/tilasto/mtt1/syky2018/htyop118.pdf> , raportin laadinta, kohta 7. Voit myös kirjoittaa alustavasti aineiston esittelyosuuden.
11. Pohdi omaa oppimistasi esimerkiksi seuraavien kysymysten avulla. Osaatko keskihajonnan? Entä osaatko muuttujan standardoinnin? Osaatko tehdä riippuvuustarkasteluja ehdollisten keskiarvojen avulla? Entä Box-plot -kuvioiden perusteella? Osaatko ristiintaulukon käytön riippuvuustarkasteluissa? Osaatko tulkita pisteparven? Tiedätkö mitä korrelaatiokerroin mittaa?