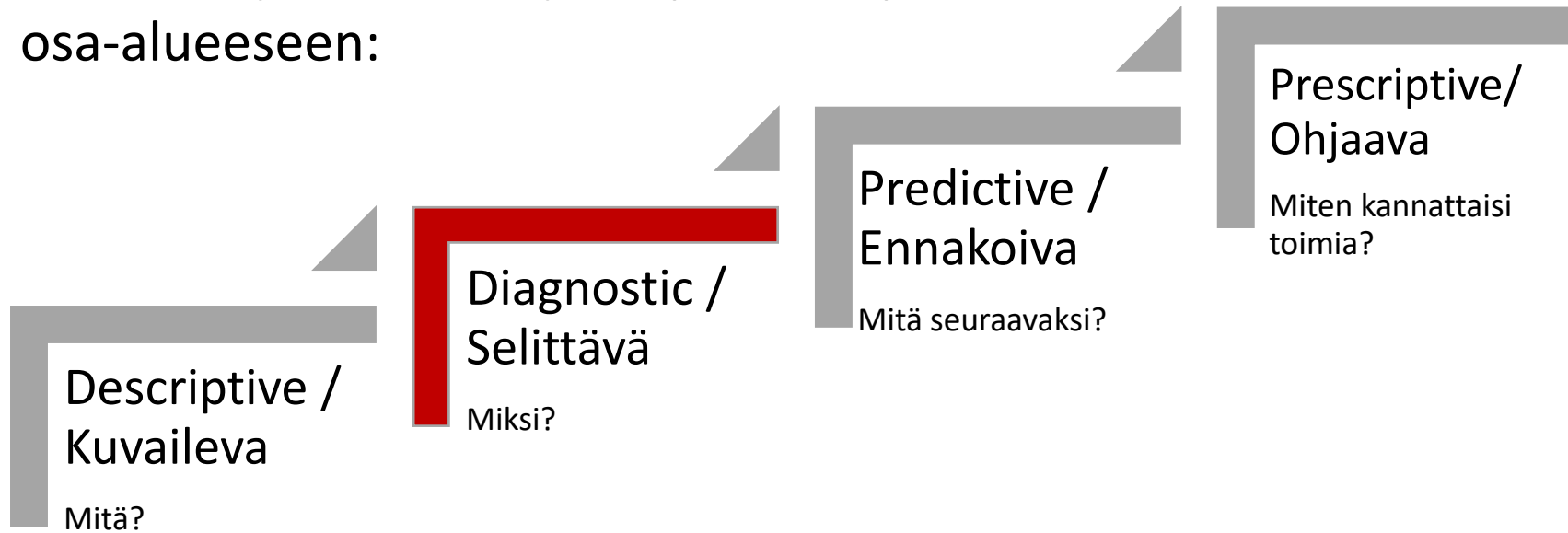


Selittävä analytiikka

Keskeiset menetelmät

Data-analytiikka

Data-analytiikka on tapana jakaa neljään osa-alueeseen:



Tämän esityksen aiheena on selittävä analytiikka

Muuttujat

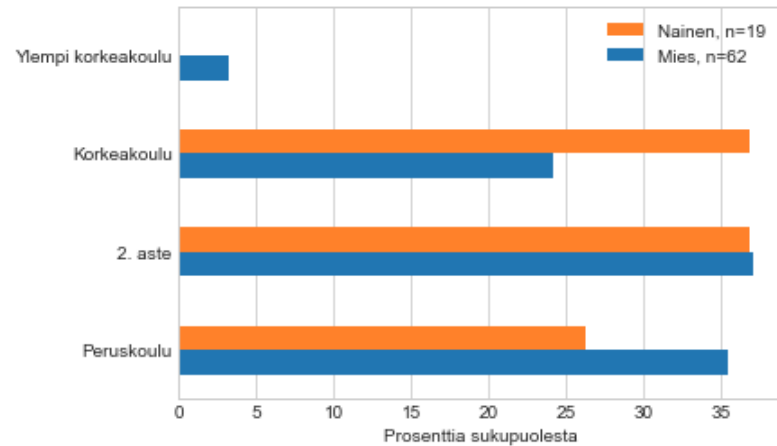
- Selittävässä analytiikassa on vähintään kaksi muuttujaa:
 1. Selitettävä muuttuja
 2. Selittävä muuttuja
- Käytettävät menetelmät riippuvat sekä selitettävän että selittävän muuttujan tyypistä
 - Sekä selitettävä että selittävä muuttuja voi olla joko kategorinen tai määrällinen
 - Näin ollen käyttötilanteet voidaan luokitella neljään erilaiseen tilanteeseen (katso seuraava dia)

Menetelmät muuttujien tyypin mukaan

		<u>Selittävä muuttuja</u>	
		Kategorinen	Määrällinen
<u>Selitettävä muuttuja</u>	Kategorinen	<ul style="list-style-type: none">• Ristiintaulukointi• Pylväskaavio• Khiin neliö -testi	<ul style="list-style-type: none">• Ristiintaulukointi, jossa selittävän muuttujan arvot luokiteltu• Khiin neliö –testi
	Määrällinen	<ul style="list-style-type: none">• Tunnuslukujen vertailu selittävän muuttujan määräämissä ryhmissä• Ruutu- ja janakaavio eli boxplot• T-testi tai Mann-Whitney U -testi• F-testi tai Kruskal-Wallis -testi	<ul style="list-style-type: none">• Hajontakaavio• Korrelaatiokerroin• Korrelaatiokertoimen testaus

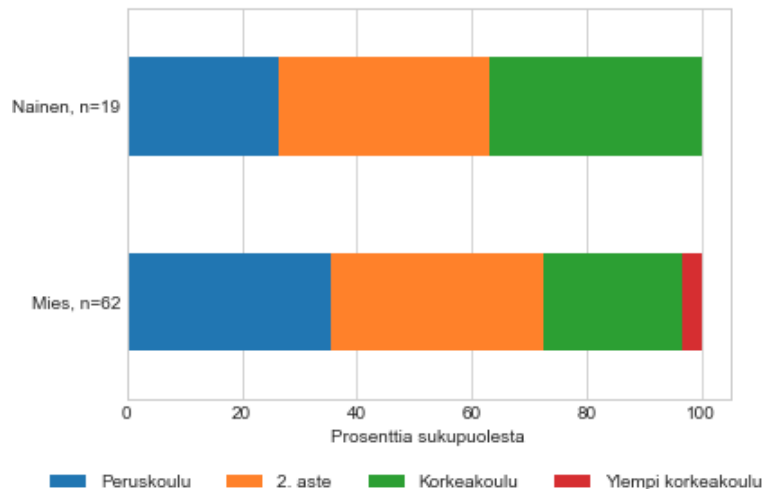
Ristiintaulukointi

Jos käytät prosentteja, niin muista n-arvot!



	Mies, n=62	Nainen, n=19
Peruskoulu	35.5 %	26.3 %
2. aste	37.1 %	36.8 %
Korkeakoulu	24.2 %	36.8 %
Ylempi korkeakoulu	3.2 %	0.0 %

Khiin neliö -testin $p = 0.607 > 0.050$.
Ei merkitsevää eroa miesten ja naisten välillä.



Tunnuslukujen vertailu

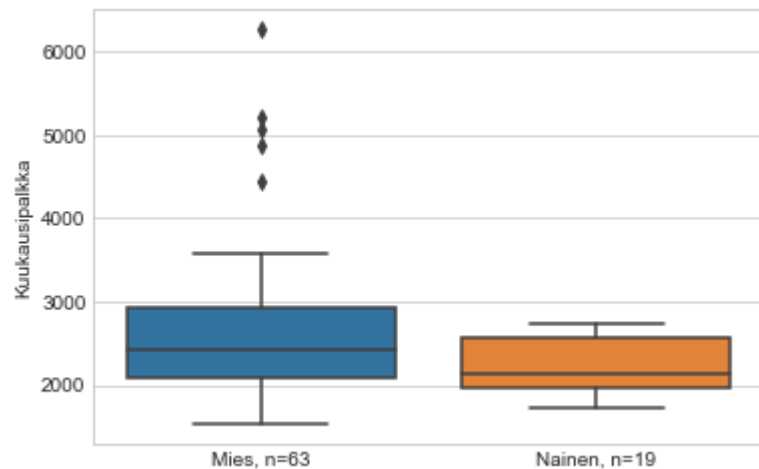
Muista n-arvot kaavioon!

	Mies	Nainen
Lukumäärä	63	19
Keskiarvo	2664	2233
Keskihajonta	930	341
Pienin	1521	1715
Alaneljännes	2086	1969
Mediaani	2417	2144
Yläneljännes	2925	2574
Suurin	6278	2729

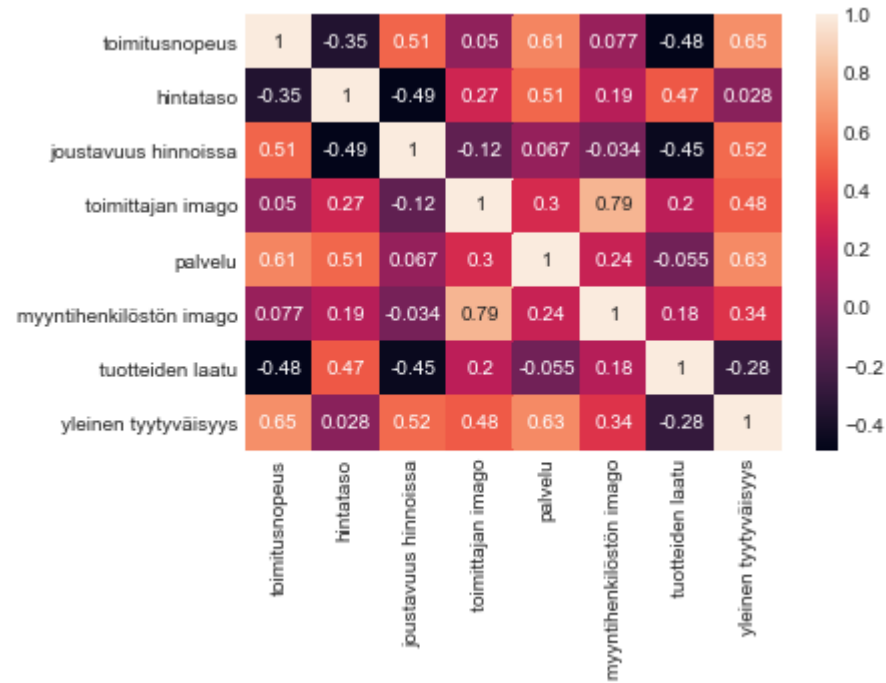
Riippumattomien otosten t-testin 2-suuntainen

$p = 0.003 < 0.050$.

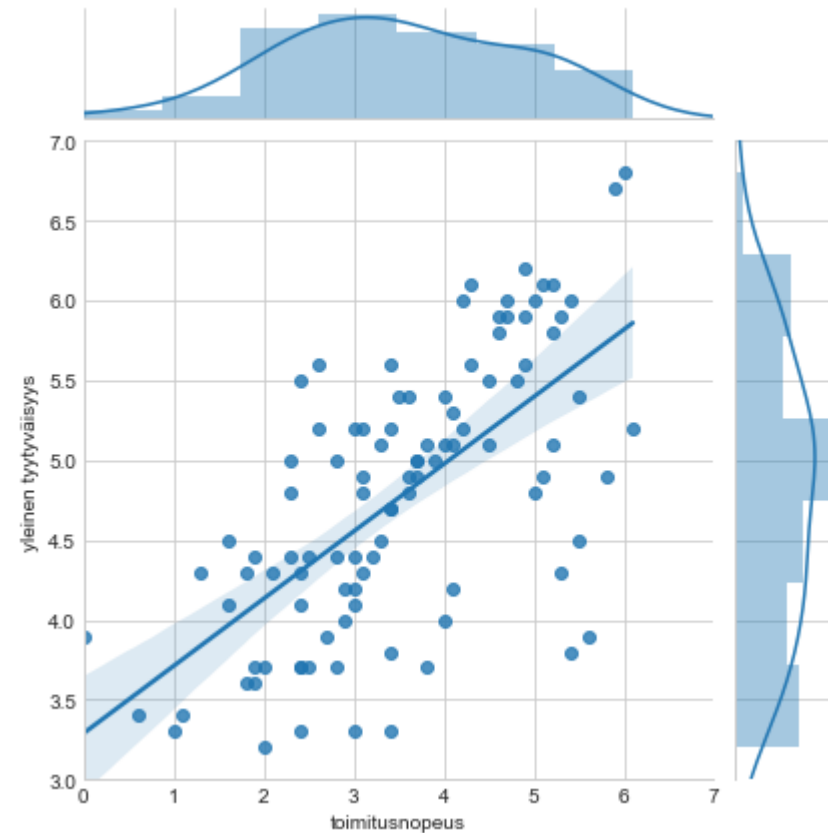
Merkitsevä ero miesten ja naisten palkkakeskiarvon välillä.



Korrelaatiokerroin ja hajontakaavio



Korrelaatiokertoimia (n = 100). Tällä otoskoolla itseisarvoltaan 0.2 ja sitä suuremmat korrelaatiokertoimet ovat merkitseviä ($p < 0.05$).



Välineet selittävään analytiikkaan

Selittävää analytiikkaa voi harjoittaa monenlaisilla välineillä:

- **Taulukkolaskennalla** (esimerkiksi Excel)
 - Merkitsevyyden testaukseen puutteelliset työkalut
- **Tilasto-ohjelmalla** (esimerkiksi SPSS)
 - Helppo suorittaa merkitsevyyden testaus
- **Koodaamalla** (Python tai R)
 - Paras vaihtoehto, jos analysoidaan toistuvasti rakenteeltaan samanlaisia dataa tai jos analysoitava data päivittyy aika ajoin
 - Tämän esityksen kaaviot ja taulukot laadittu Pythonilla