

Data-analytiikan harjoitustyö

Lääketieteelliseen tiedekuntaan
pyrkiminen vuonna 2020

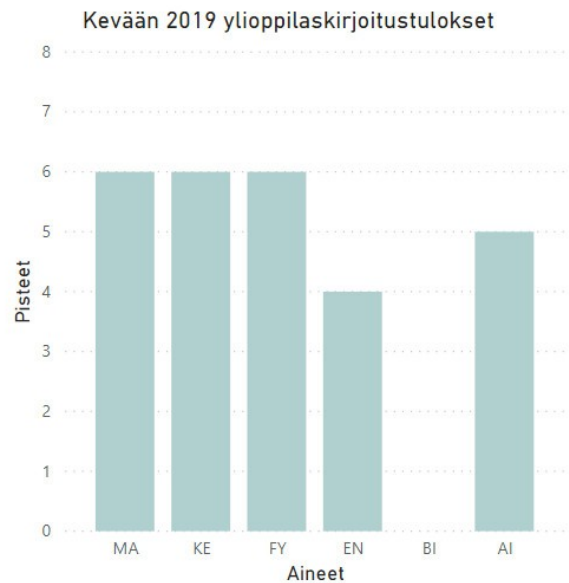
Virpi Karhula

Analytics Training Camp, Saranen Consulting Oy, 11.5.2020

Data-analytiikan harjoitustyö

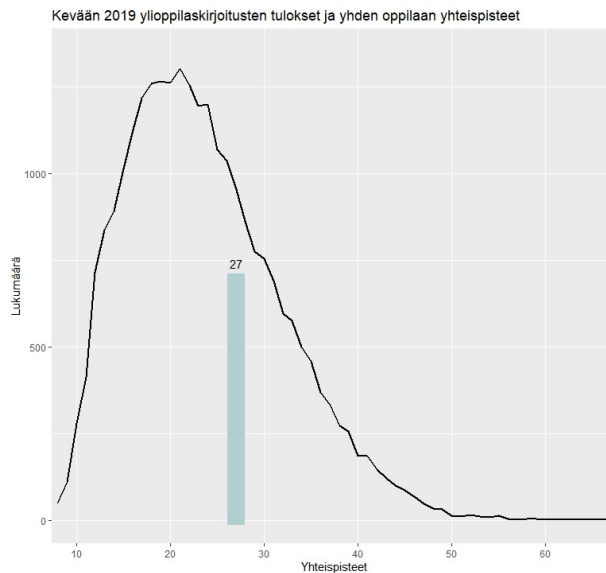
- Esityksessä kuvataan yhden ylioppilaan hakemispolkua lääketieteelliseen tiedekuntaan vuosina 2019 ja 2020.
- Harjoitustyössä visualisoidaan ja arvioidaan ylioppilasarvosanojen korotusten vaikutusta sisäänpääsymahdollisuuksiin.
- Tulokset lasketaan ja visualisoidaan käyttäen työkaluja
 - R, ggplot2 ja Power BI

Lukiolaisen ylioppilaskirjoitukset keväällä 2019



27

2019-K



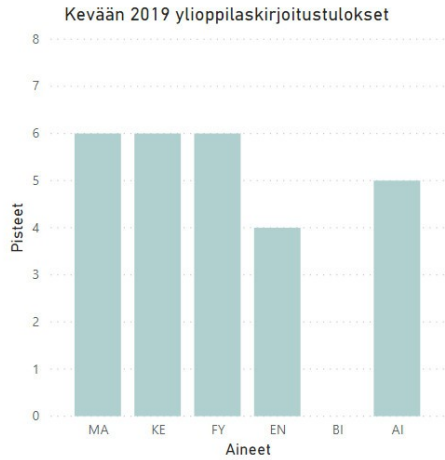
27

2019-K

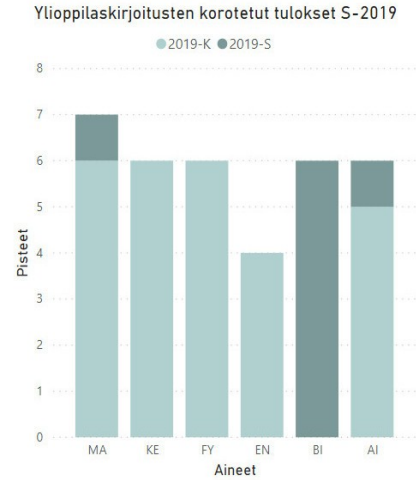
Opiskelijan painotettu
pistemäärä lääketieteellisen
tiedekunnan haussa oli 124.

Pistemäärä ei riittänyt
sisäänpääsyyn keväällä 2019.

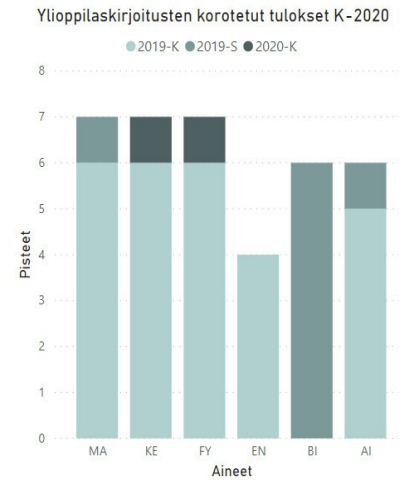
Arvosanojen korotus



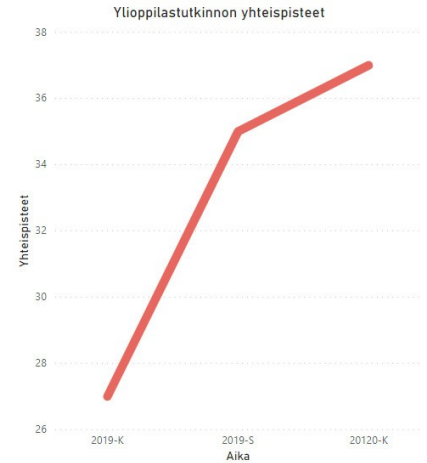
27
2019-K



35
2019-S



37
20120-K

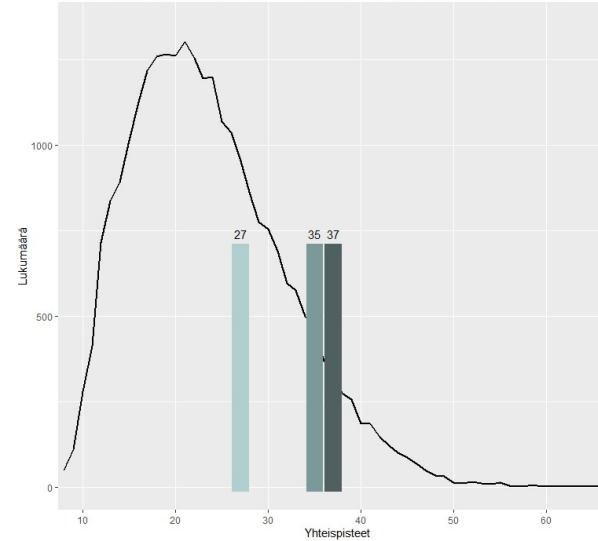


Arvosanojen korotus

Ylioppilaskirjoitusten korotetut tulokset K-2020



Kevään 2019 ylioppilaskirjoitusten tulokset ja yhden oppilaan korotetut yhteispisteet



```
ggplot(yo_values_and_one_student, aes(x=sum_points,y=amount)) +  
  geom_line(group = 1, size = 1,colour='black') +  
  geom_col(aes(x=sum_points,y=one_student_1st), size=4, colour='#b0d0d0') +  
  geom_text(aes(y=one_student_1st, label = "27"), vjust = -1, hjust = "center", show.legend = FALSE) +  
  geom_col(aes(x=sum_points,y=one_student_2nd), size=4, colour='#7c9999') +  
  geom_text(aes(y=one_student_2nd, label = "35"), vjust = -1, hjust = "center", show.legend = FALSE) +  
  geom_col(aes(x=sum_points,y=one_student_3rd), size=4, colour='#4f6062') +  
  geom_text(aes(y=one_student_3rd, label = "37"), vjust = -1, hjust = "center", show.legend = FALSE) +  
  ggtitle("Kevään 2019 ylioppilaskirjoitusten tulokset ja yhden oppilaan korotetut yhteispisteet") +  
  scale_x_discrete("Yhteispisteet", c(10,20,30,40,50,60),c("10","20","30","40","50","60")) +  
  ylim(0, 1350) +  
  ylab("Lukumäärä")
```

Lääketieteellisen painotetut todistusvalintapisteet

yht	A	M	N	BI	FF	FY	HI	PS
20	4	NA	5	4	NA	NA	NA	NA
15	4	3	NA	NA	NA	2	NA	NA
17	3	4	NA	NA	NA	4	NA	NA
16	3	5	NA	NA	NA	NA	NA	NA
27	5	6	NA	NA	NA	6	NA	NA
21	4	NA	5	NA	NA	NA	NA	NA
20	6	5	NA	NA	NA	NA	NA	NA
13	2	NA	4	NA	NA	NA	NA	NA
17	5	NA	NA	NA	NA	NA	NA	NA

Ylioppilastutkintolautakunta jakaa datan csv-taulukkona
<https://www.ylioppilastutkinto.fi/ext/data/FT2019KD4001.csv>

```
# M[,16], pitka matikka (tai)
if (!is.na(yo2019k[index,16])) {
  sum_weighted <- sum_weighted
  + add_weighted_grade(yo2019k[index,16],
    0, 6.6, 13.2, 19.8, 26.4, 33.1, 39.7)
}
```

Ylioppilaista valittiin kaikki keväällä 2019
kemian ja biologian kirjoittaneet



LÄÄKETIETEELLISET ALAT	L	E	M	C	B	A
äidinkieli*	33,0	27,5	22,0	16,5	11,0	5,5
kemia	34,0	28,3	22,7	17,0	11,3	5,7
biologia	32,3	26,9	21,5	16,2	10,8	5,4
Matematiikka (pitkä tai lyhyt)						
matematiikka, pitkä	39,7	33,1	26,4	19,8	13,2	6,6
matematiikka, lyhyt	28,3	23,6	18,9	14,1	9,4	4,7

ind	ordered_ke_bi
1	20.5
2	22.1
3	25.2
4	25.2
5	26.7

Lääketieteellisen valintaperusteet v. 2020

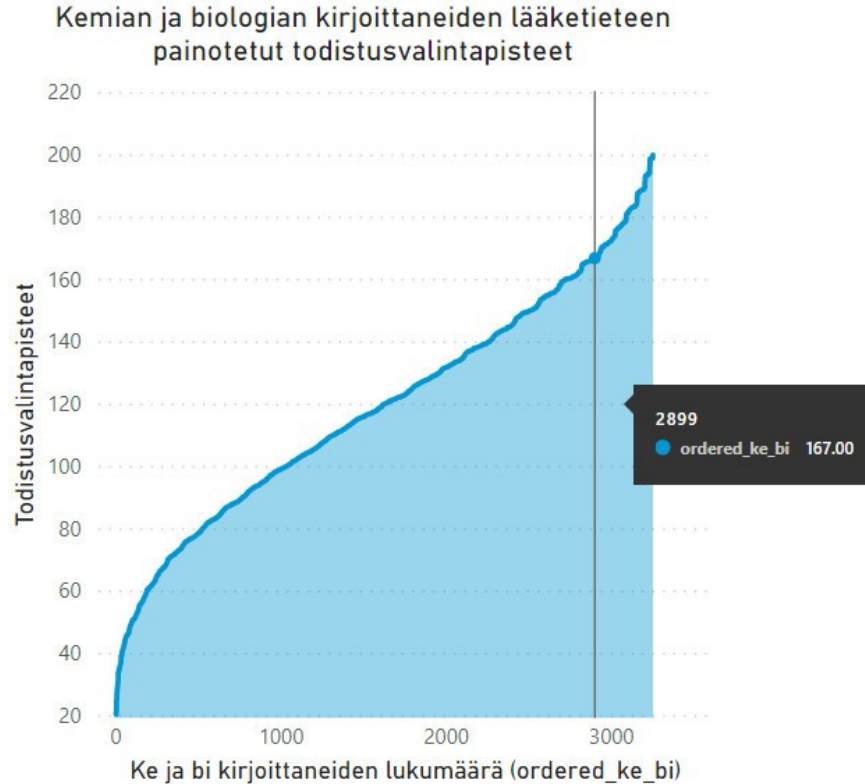
- sisäänotto yhteensä 725 hlöä

1) Todistusvalinta, ensikertalaiset 48,75 %

2) Todistusvalinta, ei-ensikertalaiset 26,25 %

3) Pääsykoe 25 %, 1) etänä 2) paikan päällä

Lääketieteellisen painotetut todistusvalintapisteet



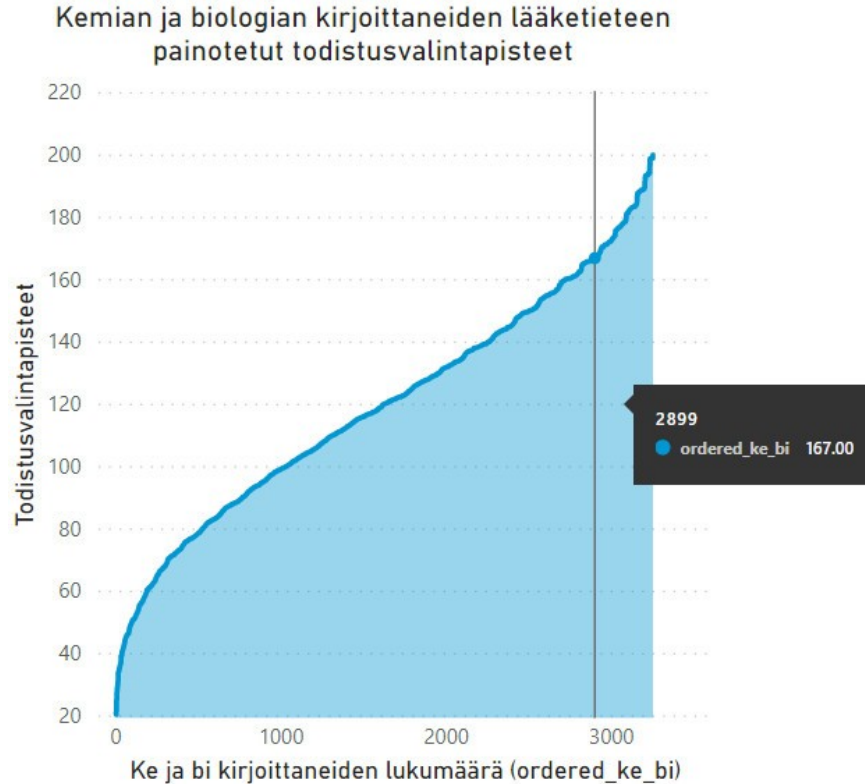
Kuvaajassa keväällä 2019 kaikki kemian ja biologian kirjoittaneiden arvosanat on kerrottu lääketieteen painoituksin todistusvalintapisteiksi.

Ensikertalaisista todistusvalinnalla valitaan 48,75 % (353 hlöä)

Aika	Painotetut pisteet
K-2019	124
S-2019	163
K-2020	173.9

Jos kuvaaja kuvaisi kaikkia hakijoita, pistemäärä 173.9 riittäisi sisäänpääsyyn.

Datan virhelähteet



- Ei ole saatavilla tietoa:
 - Moniko kemian ja biologian kirjoittanut on hakemassa lääketieteelliseen?
 - Ylioppilasarvosanojen **korotukset** (tod.näk. E→L 8 %)
 - Moniko on ilman opiskelupaikkaa (ensikertalainen) ja moniko on jo vastaanottanut opiskelupaikan?
- Lääketieteelliseen pääsee sisään **keskimäärin kolmantena hakuvuonna**
 - Pitäisi ottaa huomioon usean vuoden kirjoittaneet
- Todella suuri virhemarginaali
 - Ensikertalaisten kiintiössä valitaan 48,75 % (353 hakijaa)
 - Mahdoton tietää, minkä vuoden ylioppilaita ensikertalaiset ovat ja paljonko he ovat korottaneet ylioppilasarvosanojaan

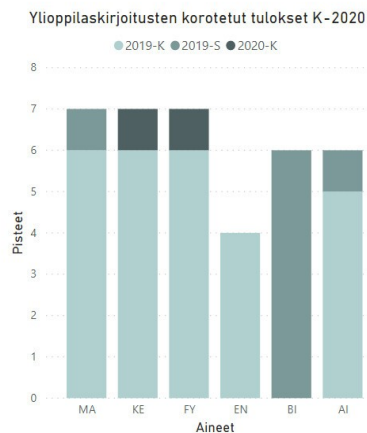
MAFY-valmennuksen arvion mukaan ensikertalaisten todistuspisteiden alaraja Helsingin lääketieteelliseen tiedekuntaan on tänä vuonna 171.

- Tällä arviolla 173.9 pistettä riittäisi sisäänpääsyyn

Harjoitustyön työkalut

- Pylväsdiagrammit
 - Data: LibreOfficen Calc-ohjelmalla tallennettu csv-tiedosto
 - Visualisointi: Power BI
- Datan sijoittaminen taulukkoon (rivit ja sarakkeet)

Subjects	2019-K	2019-S	2020-K
AI	5	1	0
MA	6	1	0
KE	6	0	1
FY	6	0	1
BI	0	6	0
EN	4	0	0



time	subject	grade	improvement
K-2019	AI	5	0
K-2019	MA	6	0
K-2019	KE	6	0
K-2019	FY	6	0
K-2019	BI	0	0
K-2019	EN	4	0
S-2019	AI	6	1
S-2019	MA	7	1
S-2019	KE	0	0
S-2019	FY	0	0

Harjoitustyön työkalut

- Lääketieteelliseen painotetut valintapistekuvaajat
 - R-koodilla lääketieteellisen painokerrointen laskeminen ylioppilastuloksissa kemian ja biologian kirjoittajien tuloksiin
 - Tulosdatan siirto data.frame-muodossa csv-tiedostoon
- Visualisointi
 - Ylioppilastutkintodata: R, ggplot
 - Todistuvaintadata: Power BI

```
ggplot(yo_values_and_one_student, aes(x=sum_points,y=amount)) +  
  geom_line(group = 1, size = 1,colour='black') +  
  geom_col(aes(x=sum_points,y=one_student_1st), size=4, colour='#b0d0d0') +  
  geom_text(aes(y=one_student_1st, label = "27"), vjust = -1,  
    hjust = "center", show.legend = FALSE) +  
  ggtitle("Kevään 2019 ylioppilaskirjoitusten tulokset ja yhden oppilaan  
    yhteispisteet") +  
  scale_x_discrete("Yhteispisteet", c(10,20,30,40,50,60),  
    c("10","20","30","40","50","60")) +  
  ylim(0, 1350) +  
  ylab("Lukumäärä")
```



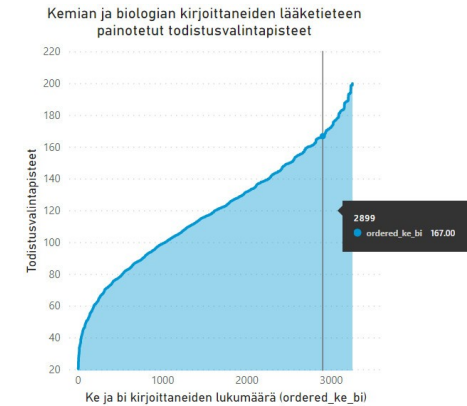
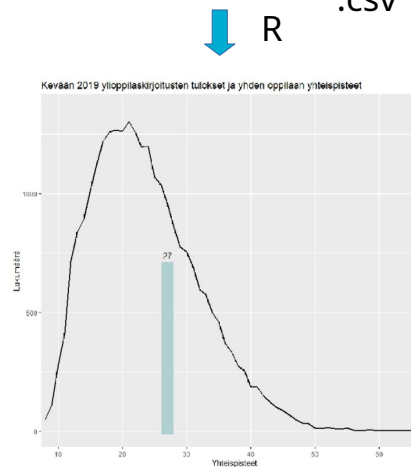
yht	A	M	N	BI	FF	FY	HI	PS
20	4	NA	5	4	NA	NA	NA	NA
15	4	3	NA	NA	NA	2	NA	NA
17	3	4	NA	NA	NA	4	NA	NA
16	3	5	NA	NA	NA	NA	NA	NA
27	5	6	NA	NA	NA	6	NA	NA
21	4	NA	5	NA	NA	NA	NA	4
20	6	5	NA	NA	NA	NA	NA	4
13	2	NA	4	NA	NA	NA	2	NA
17	5	NA	NA	NA	NA	NA	3	NA



ind	ordered_ke_bi
1	20.5
2	22.1
3	25.2
4	25.2
5	26.7



.CSV



R-koodi GitHubissa

https://github.com/vkarhula/Data_Analysis

Kiitos!

Virpi Karhula
vkarhula@gmail.com
+358 40 5634653