

8. Kirjeldavad arvnäitajad

Variatsioonirida - rida, kus me oleme numbrilise tunnuse väärtused järjestanud suuruse järgi.

Ulatus - maksimaalse ja min väärtuse vahe.

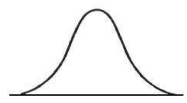
Mediaan – variatsioonireaa keskel paiknev väärtus, mis jagab vaatlustulemused kahte ossa, pooled on mediaanist suuremad ja pooled väiksemad.

Mood – jaotuse väärtus, mis esineb kõige sagedamini.

Kvartiilid - jagavad vaatlustulemused nelja võrdsesse ossa.

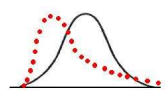
Standardhälve – hajuvuse näitaja, mis arvestab kõiki vaatlustulemusi ning näitab kui palju üksikud tulemused erinevad keskmisest. Mida suurem on hajuvus, seda rohkem nad erinevad ning seda suurem on standardhälve. Kui kõik vaatlustulemused on ühesugused (Nt. kõik tudengid said kontrolltööl 15 palli), siis hajuvust ei ole ja standardhälve on 0. (standardhälve ei ületa tavaliselt poolt jaotuse ulatusest)

Jaotuse kuju.

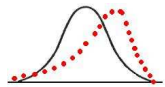


Lisaks eelnevale on ka graafikutel oluline osa jaotuse kuju kirjeldamisel. Kõige enam esinev jaotuse kuju on *sümmeetriline*. Kuid see pole ainuvõimalik. Jaotust, mis ei ole sümmeetriline (on "väljavenitatud") nimetatakse *asümmeetriliseks*.

Jaotuse asümmeetriat iseloomustab asümmeetria koefitsent (kui $a=0$ on jaotus sümmeetriline)



* Kui jaotus on väljavenitatud paremalt (saba jääb paremale poole) on **positiivne asümmeetria**: $a>0$



* Kui jaotus on väljavenitatud vasakult (saba jääb vasakule poole) on **negatiivne asümmeetria**: $a<0$

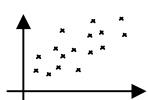
Mean		Aritmeetiline keskmine
Median		Mediaan
Mode		Mood (Multiple modes exist. The smallest value is shown- multimodaalne jaotus)
Std. Deviation		Standardhälve
Variance		Dispersioon (standardhälve ²)
Skewness		Asümmeetria
Kurtosis		Ekstsess
Range		Ulatus (max-min)
Minimum		Minimaalne väärtus
Maximum		Maksimaalne väärtus
Percentiles	25	25-s protsentiil (alumine kvartiil) – väiksemaid väärtuseid 25%
	40	40-s protsentiil - väiksemaid väärtuseid 40%
	50	50-s protsentiil – mediaan
	75	75-s protsentiil – ülemine kvartiil (väiksemaid väärtuseid 75%)

9. Korrelatsioonanalüüs

Siiani uurisime, kas ühe tunnuse väärtused on paremad kui teised või kas tunnuse jaotus on ühtlane. Edasi vaatame, kas kahe tunnuse vahel on seos ning milline see on: keskmine hinne lõputunnistusel on kindlasti seotud õpilase üldise intelligentsusega ja võimekusega, aga ta on ka seotud tervisliku seisundi jpm.

Esmalt on vaja vaadata, et tunnused ei oleks sõltumatud – omandavad väärtusi üksteisest sõltumata. Selliste tunnuste ühisel uurimisel ei ole mõtet, see ei anna meile mingit lisainformatsiooni võrreldes nende kahe tunnuse eraldi uurimisega.

Graafilise ülevaate kahe tunnuse vahelisest seosest annab **korrelatsiooniväli** e. **hajuvusdiagramm**

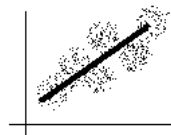


(kahemõõtmeline diagramm, millel kaks uuritavat tunnust määravad teljed ning igat katseisikut tähistatakse punktiga). Jooniselt näeme, et punktid paiknevad vasakult paremale tõusva pilvena – see viitab ühisele tendentsile tunnuste käitumises. Ühe tunnuse suurem väärtus omandab suurema väärtuse ka teises tunnuses.

Kahe tunnuse vahelise seose iseloomustamiseks vaadatakse:

* **seose kuju** – määrab geomeetiline joon, millele punktide parv kõige lähedasem on.

Kõige sagedamini on selleks sirgjoon – lineaarne seos. N: õppeedukus ja õppimisele kulunud aeg



* **seose suunda**

positiivne – kui ühe tunnuse väärtuste kasvades, kasvavad ka teise tunnuse väärtused ja vastupidi: ühe tunnuse väärtuste kahanedes, kahanevad ka teise tunnuse väärtused.

negatiivne – kui ühe tunnuse väärtuste kasvades, teise tunnuse väärtused kahanevad ja vastupidi: ühe tunnuse väärtuste kahanedes, teise tunnuse väärtused kasvavad.

* **seose tugevust**

Mida tugevam on seos kahe tunnuse vahel, seda tihedam (joone lähedale koonduv) on punktivarv.

Seose tugevust näitab korrelatsioonikordaja, mille absoluutväärtus muutub 0...1

Kui kahe tunnuse vahel seost ei ole, on kordaja absoluutväärtus 0

Kui kahe tunnuse vahel on funktsionaalne seos, on kordaja abs.väärtus 1

Kordaja	Seose tugevus
0,0 – 0,2	Olematu, väga nõrk
0,2 – 0,4	Nõrk
0,4 – 0,7	Keskmine
0,7 – 0,9	Tugev
0,9 – 1	Väga tugev

Kordaja	Seose tugevus
0,0 – 0,3	Olematu, väga nõrk seos
0,3 – 0,7	Keskmise tugevusega
0,7 - 1	Tugev seos

* **Korrelatsioonikordajad.**

Pearson r – mõlemad tunnused numbrilised, on tundlik erandite suhtes.

Spearman ρ – üks või mõlemad tunnused järjestustunnus(ed), vähendab erandite mõju, kuna kordaja leidmisel kasutatakse astakuid (järjenumbreid), mitte konkreetseid väärtusi.

Kendall τ - üks või mõlemad tunnused järjestustunnus(ed), ei ole tundlik erandite suhtes, kuna vaadeldakse objektide paare.

* **Näited**

Correlations		vanus	Tunde majapidamistöõdeks (tööpäeviti)
vanus	Pearson Correlation	1	,198**
	Sig. (2-tailed)		,000
	N	882	873
Tunde majapidamistöõdeks (tööpäeviti)	Pearson Correlation	,198**	1
	Sig. (2-tailed)	,000	
	N	873	873

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

Pearson Correlation – Pearsoni korrelatsioonikordaja (r)

Sig. (2-tailed) – olulisustõenäosus

N – objektide arv

1) seos kahe tunnuse vahel on positiivne ($r=0,198$) *vanemad inimesed koristavad rohkem, nooremad vähem*

2) tunnuste vahel on nõrk seos.

Kuna üks tunnustest on järjestustunnus (tervislik seisund), seega Pearsoni kordajat leida ei saa, arvutame Spearmani kordaja

Correlations			sissetulek viimasel kuul	Tervislik seisund
Spearman's rho	sissetulek viimasel kuul	Correlation Coefficient	1,000	-,254**
		Sig. (2-tailed)		,000
		N	829	821
Tervislik seisund	Tervislik seisund	Correlation Coefficient	-,254**	1,000
		Sig. (2-tailed)	,000	
		N	821	873

** . Correlation is significant at the .01 level (2-tailed).

1) seos kahe tunnuse vahel on negatiivne ($r= -0,254$) *mida suurem on sissetulek, seda parem on hinnang oma tervislikule seisundile ja vastupidi*

2) tunnuste vahel on nõrk seos