

Koostatud juhend on mõeldud lisamaterjalina kasutamiseks „Andmeanalüüsi“ kursuse kuulajatele. Näidiste ning õpetuste loomisel on kasutatud uuringu „Mina, Maailm ja Meedia 2008“ andmeid. Kõik näited põhinevad statistikapaketi SPSS versioonil 18.0. Juhendi koostas K.Osula.

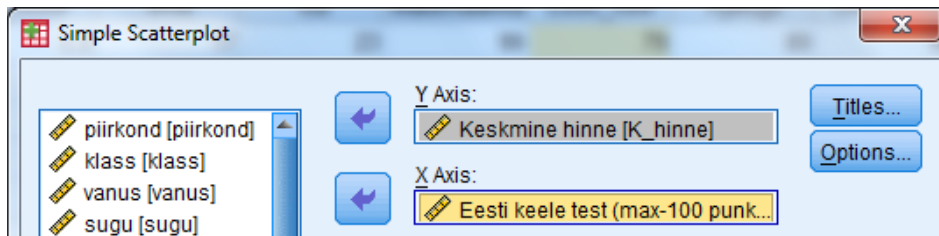
Juhendi sisukord:

- 1. Korrelatsiooniväli 1
- 2. Kahe intervalltunnuse vahelise seose arvutamine 2
- 3. Kahe intervalltunnuse vahelise seose arvutamine (erandlikud väärtused) 3
- 4. Seose uurimine kui vähemalt üks tunnustest on järjestustunnus 5
- 5. Seose uurimine nimetunnuse (ja/või binaarse tunnuse) korral 6

1. Korrelatsiooniväli

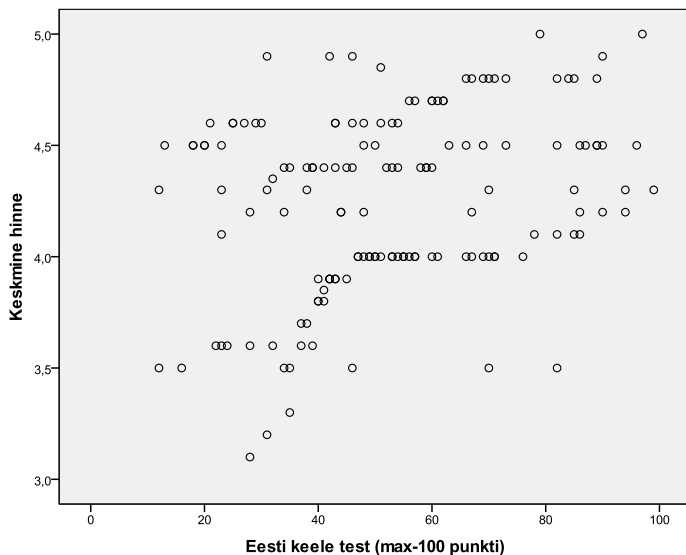
Koostame kahe tunnuse vahelise seose kirjeldamiseks korrelatsiooni välja.

Vali: *Graphs/Legacy Dialogs/Scatter/Dot.../Simple Scatter*



Määra x-teljel ja y-teljel kuvatavad tunnused.

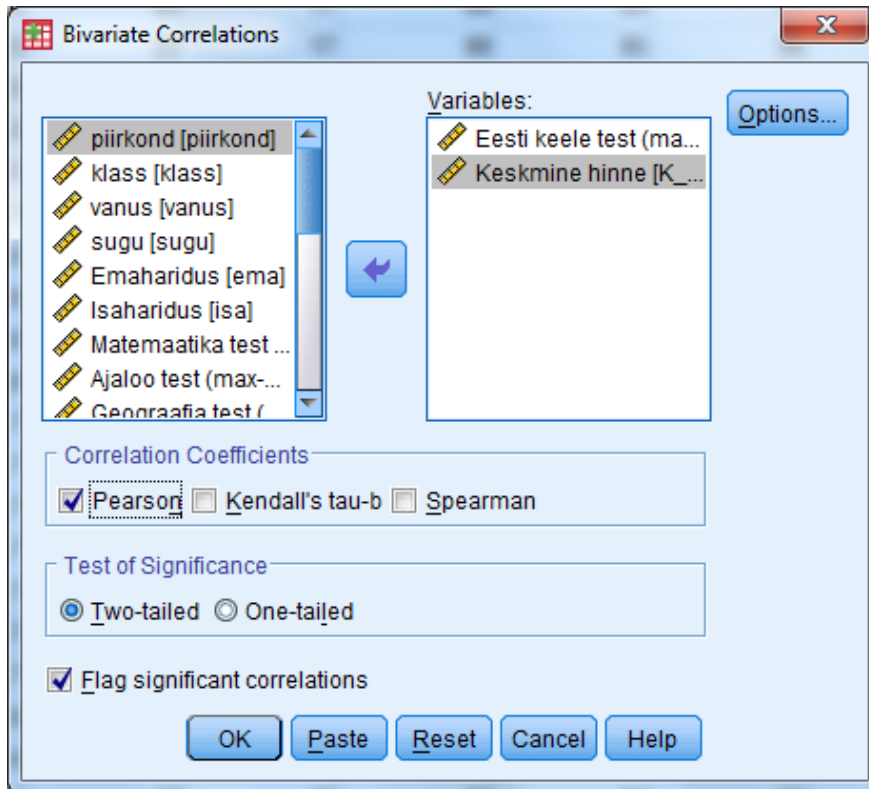
Tulemus kuvatakse tulemuste (*Output*) faili.



Koostatud diagrammi vaadeldes paistab silma kerge suundumus e punktiväli on koondunud tõusvas joones.

2. Kahe intervalltunnuse vahelise seose arvutamine

Vali: *Analyze/Correlate/Bivariate...*



Vii uuritavad tunnused *Variables* väljale.

Märgista sobivad korrelatsioonikordaja(d). Kuna on tegemist kahe intervalltunnusega, siis sobib *Pearsoni* kordaja.

Tulemus kuvatakse tulemuste (*Output*) faili.

Correlations			
		Eesti keele test (max-100 punkti)	Keskmine hinne
Eesti keele test (max-100 punkti)	Pearson Correlation	1	,299**
	Sig. (2-tailed)		,000
	N	150	150
Keskmine hinne	Pearson Correlation	,299**	1
	Sig. (2-tailed)	,000	
	N	150	150

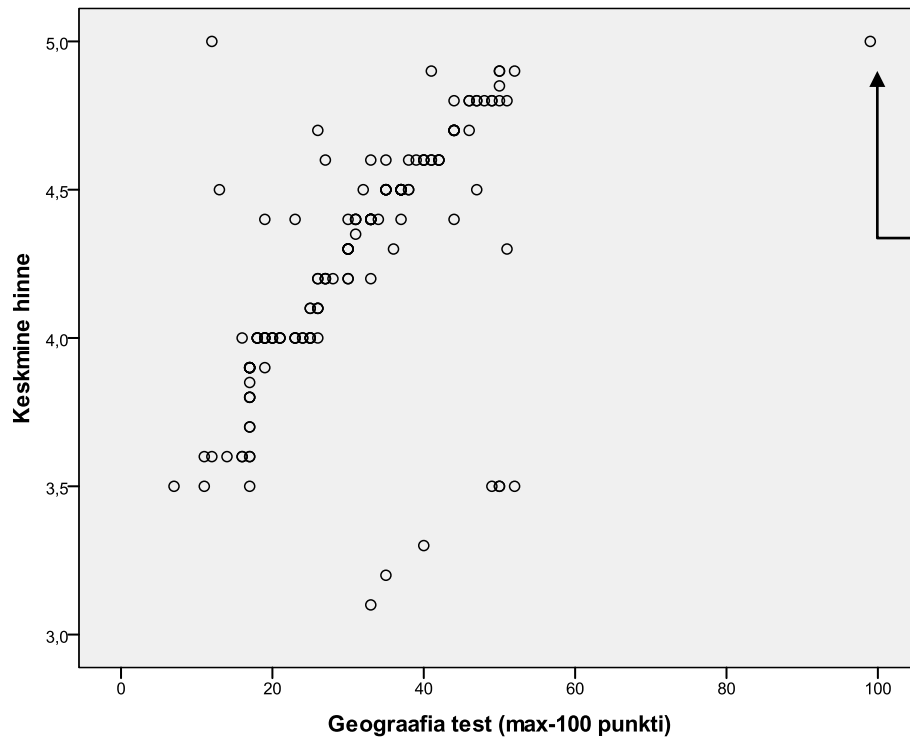
** . Correlation is significant at the 0.01 level (2-tailed).

Tunnuste vahelist seost iseloomustab korrelatsioonikordaja väärtus (0,299) järgmiselt:

- Kuna kordaja on positiivne (pluss märgiga), siis on tegemist positiivse (samasuunalise) seosega. Mida kõrgem oli õpilase keskmine hinne, seda kõrgem oli ka eesti keele testi tulemus ning vastupidi, mida madalam oli keskmine hinne, seda vähem punkte sai õpilane eesti keele testi eest.
- Korrelatsioonikordaja absoluutväärtus näitab seose tugevust. Meie näites on tegemist nõrga seosega.

3. Kahe intervalltunnuse vahelise seose arvutamine (erandlikud väärtused)

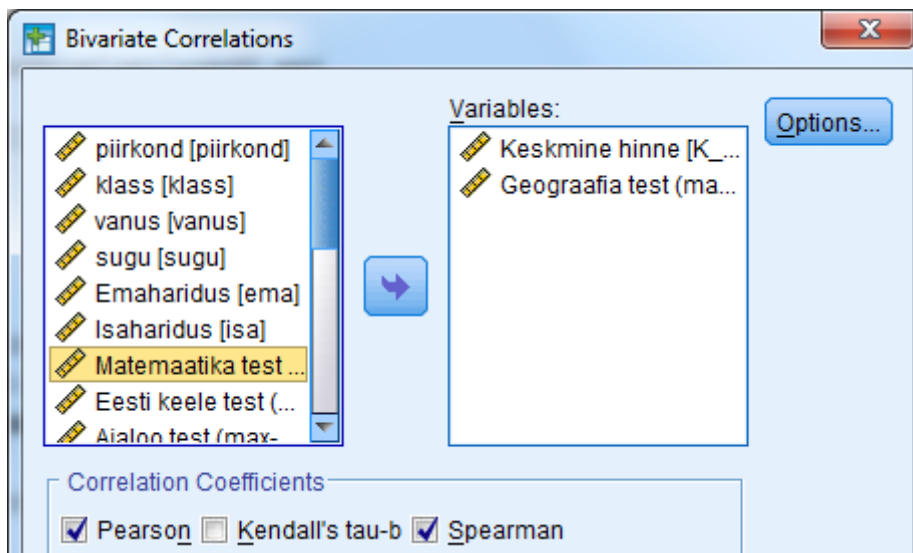
Koostame korrelatsioonivälja.



Näeme, et punktipilv on koondunud tõusvas joones, kuid jaotusel esineb teistest eraldiseisvaid väärtuseid, mis võivad mõjutada *Pearsoni* kordaja suurust.

INTERVALL-TUNNUSED

Arvutame mõlemad kordajad (Pearson, Spearman).



Tulemuseks saame kaks korrelatsioonitabelit:

Correlations

		Keskmine hinne	Geograafia test (max-100 punkti)
Keskmine hinne	Pearson Correlation	1	,605**
	Sig. (2-tailed)		,000
	N	150	150
Geograafia test (max-100 punkti)	Pearson Correlation	,605**	1
	Sig. (2-tailed)	,000	
	N	150	150

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

Correlations

			Keskmine hinne	Geograafia test (max-100 punkti)
Spearman's rho	Keskmine hinne	Correlation Coefficient	1,000	,705**
		Sig. (2-tailed)	.	,000
		N	150	150
	Geograafia test (max-100 punkti)	Correlation Coefficient	,705**	1,000
		Sig. (2-tailed)	,000	.
		N	150	150

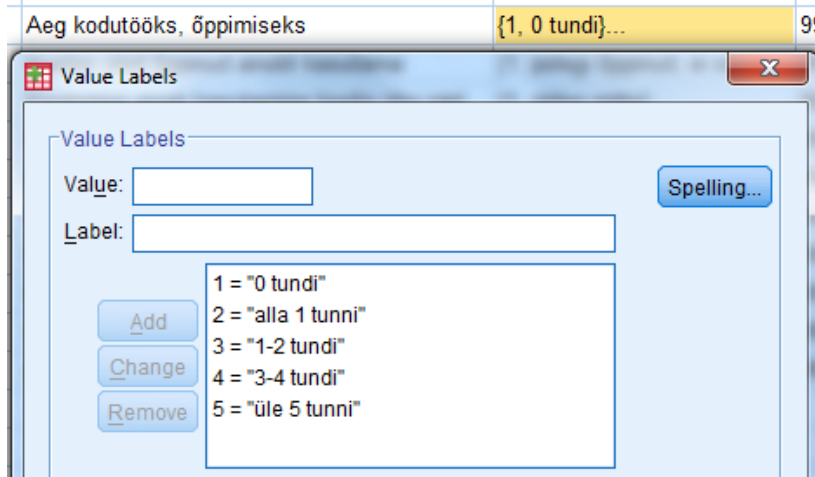
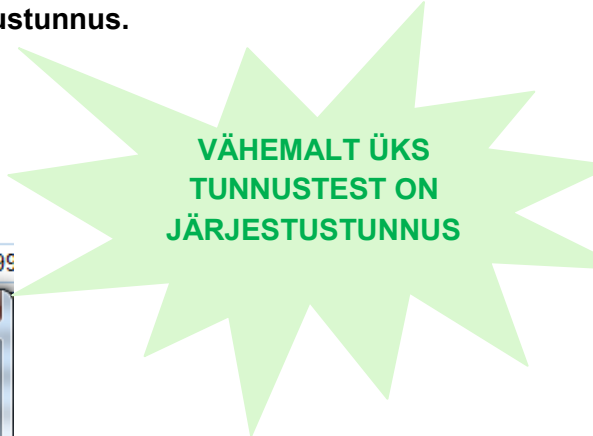
** . Correlation is significant at the 0.01 level (2-tailed).

Järelduse tegemiseks kasutame Spearmani kordajat (0,705), sest teistest eraldi paiknev väärtus muutis Pearsoni kordaja väärtuse tegelikkusest madalamaks.

- Kuna kordaja on positiivne (pluss märgiga), siis on tegemist positiivse (samasuunalise) seosega. Mida kõrgem oli õpilase keskmine hinne, seda kõrgem oli ka geograafia testi tulemus ning vastupidi, mida madalam oli keskmine hinne, seda vähem punkte sai õpilane geograafia testi eest.
- Korrelatsioonikordaja absoluutväärtus näitab seose tugevust. Meie näites on tegemist tugeva seosega.

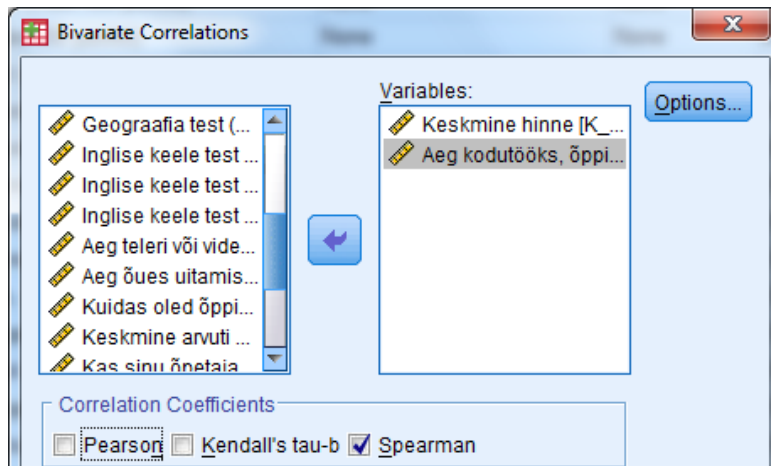
4. Seose uurimine kui vähemalt üks tunnustest on järjestustunnus.

Uurime keskmise hinde ja õppimiseks ning kodutöödeks kuluva aja vahelist seost. Kodutöödeks ja õppimisele kuluvat aega on uuritud järgmiselt:



Kui vähemalt üks uuritavatest tunnustest on järjestustunnus, siis tuleb seose uurimiseks kasutada Spearmani kordajat.

Vali: *Analyze/Correlate/Bivariate...*



Correlations

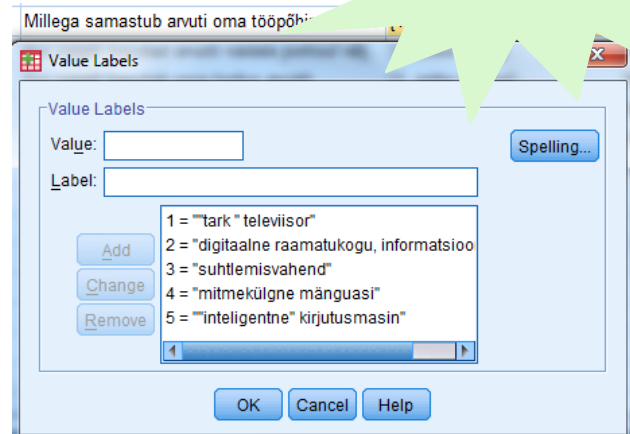
			Keskmine hinne	Aeg kodutööks, õppimiseks
Spearman's rho	Keskmine hinne	Correlation Coefficient	1,000	-,066
		Sig. (2-tailed)	.	,425
		N	150	150
	Aeg kodutööks, õppimiseks	Correlation Coefficient	-,066	1,000
		Sig. (2-tailed)	,425	.
		N	150	150

Kuna korrelatsioonikordaja absoluutväärtus on 0,07, siis saame järeldada, et keskmine hinne ning aeg kodutöödeks ja õppimiseks ei ole omavahel seotud.

5. Seose uurimine nimitunnuse (ja/või binaarse tunnuse) korral

Klassikaliste korrelatsioonikordajate kasutamise asemel saab anda ülevaate kahe tunnuse ühiskäitumisest (proportsioonide jagunemisest) kahemõõtmelise sagedustabeli e risttabeli abil.

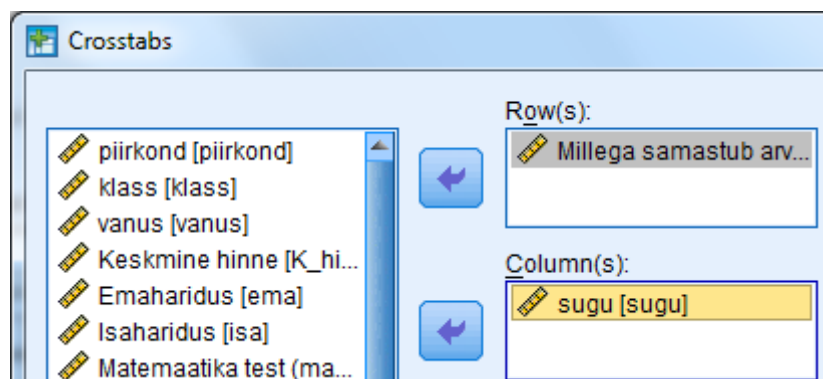
Uurime poiste tüdrukute vastuseid küsimusele „Millega samastub arvuti oma tööpõhimõttelt kõige enam?“



NIMI-TUNNUSED

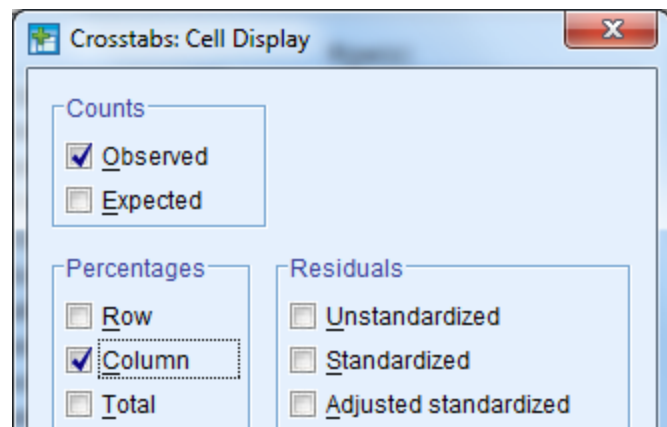
Vali: *Analyze/Descriptive Statistics/Crosstabs...*

Vii üks tunnustest *Row(s)* ning teine *Column(s)* väljale.

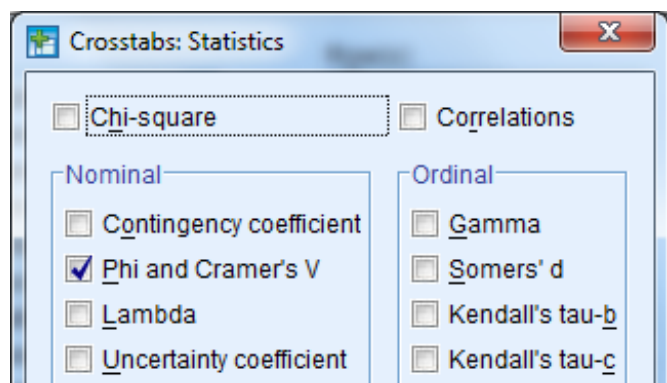


Ava *Cells* nupp ning võta kaasa grupeeriva tunnuse protsendid.

Kuna meie näites on võrreldavaks tunnuseks sugu, mis asub veerus (*Columns* väljal), siis võtame kaasa veeruprotsendid.



Statistics nupu alt saame kaasa võtta *Cramer* V kordaja, mis sarnaselt teistele korrelatsioonikordajatele näitab seose tugevust kuid mitte suunda (*Cramer* V on alati positiivne ning muutub vahemikus 0...+1)



Tulemuseks saame risttabeli:

Millega samastub arvuti oma tööpõhimõttelt kõige enam * sugu Crosstabulation

			sugu		Total
			naissoost	meessoost	
Millega samastub arvuti oma tööpõhimõttelt kõige enam	"tark " televiisor	Count	2	5	7
		% within sugu	3,7%	5,2%	4,7%
	digitaalne raamatukogu, informatsiooni ladu	Count	30	42	72
		% within sugu	55,6%	43,8%	48,0%
	suhtlemisvahend	Count	14	16	30
		% within sugu	25,9%	16,7%	20,0%
	mitmekülgne mänguasi	Count	5	25	30
		% within sugu	9,3%	26,0%	20,0%
	"intelligentne" kirjutusmasin	Count	3	8	11
		% within sugu	5,6%	8,3%	7,3%
Total		Count	54	96	150
		% within sugu	100,0%	100,0%	100,0%

Lisaks kuvatakse Crameri V väärtus (0,23), millest saame järeldada, et seos on nõrk.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,229	,096
	Cramer's V	,229	,096
N of Valid Cases		150	