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. Gruppide võrdlemine kui keskväärtust arvutada ei saa	1
1.1. Näide (andmeanalüüsi küsimus)	1
1.2. Risttabeli koostamine	2
1.3. Risttabeli koostamine kui tunnusel on palju erinevaid väärtuseid	3
1.4. Gruppide tulemuste graafiline võrdlemine – võrdlev tulpdiagramm	4
1.5. Kihtdiagramm	5
2. Gruppide võrdlemine keskväärtuse arvutamise kaudu	8

## 1. Gruppide võrdlemine kui keskväärtust arvutada ei saa

Lisaks tavapärasele ülevaate andmisele vastustest on sageli huvitav/vajalik analüüsi käigus võrrelda, milliseid vastuseid on erinevatesse gruppidesse/kategooriatesse kuuuvad isikud andnud.

#### 1.1. Näide (andmeanalüüsi küsimus)

Kui me vaatame järgnevat andmestikku, siis võime esitada näiteks niisuguse küsimuse:

Millist sissetulekut esineb meeste hulgas kõige rohkem ja kas see langeb kokku naiste hulgas kõige enamlevinud sissetulekuga?

	sugu	Vanus	Haridus	Sissetulek	töökat	v_palk_vastab_haridusele
1	naine	17	Põhiharidus	kuni 160€	kontoritöötaja	Üldiselt nõustun
2	mees	17	Põhiharidus	386-640€	kontoritöötaja	Nõustun täielikult
3	naine	19	Kutseharidus + keskharidus	Keeldus vastamast	kontoritöötaja	Üldiselt nõustun
4	mees	20	Kutseharidus + keskharidus	161-260€	juht	Nõustun täielikult
5	mees	20	Põhiharidus	261-385€	kontoritöötaja	Nõustun täielikult
6	mees	20	Põhiharidus	Keeldus vastamast	kontoritöötaja	Ei nõustu üldse
7	mees	20	Keskharidus	Keeldus vastamast	eraettevõtja	Üldiselt nõustun
8	mees	21	Ülikooliharidus. kraadiharidus	261-385€	kontoritöötaia	Nõustun täielikult

1 = "kuni 160€"	
2 = "161-260€"	
3 = "261-385€"	
4 = "386-640€"	

- 5 = "üle 640€"
- 9 = "Keeldus vastamast"

Meie andmestikus on sissetulek kogutud kategoriseeritult e järjestustunnusena.

Seega ei saa me arvutada meeste-naiste sissetulekute keskväärtust vaid peame valima gruppide (meeste ja naiste) vastuste võrdlemiseks teise meetodi.

Arvutame sissetulekute sagedused ja nendest lähtuvad protsendid võrreldavate gruppide lõikes e koostame võrdleva sagedustabeli e risttabeli.

Kuna meie andmestikus ei ole mehi ja naisi ühepalju, ei saa me võrdlemisel lähtuda gruppide sagedustest vaid peame võrdlemisel kasutama gruppide osakaalusid (protsente), kusjuures protsendid tuleb arvutada nii, et iga võrreldav grupp moodustaks kokku 100%.

Sugu						
		Frequency	Percent			
Valid	Mees	694	46,2			
	Naine	808	53,8			
	Total	1502	100,0			

### 1.2. Risttabeli koostamine

Risttabeli koostamiseks vali Analyze/Descriptive Statistics/Crosstabs...



Viime selle tunnuse, mille gruppe me omavahel võrdlema hakkame Row(s): väljale. Sellisel juhul saame tulemuseks tabeli, milles meeste ja naiste vastused asuvad ridades (üksteise all).

Column(s): väljale viime uuritava tunnuse.

🔢 Crosstabs: Cell Di	splay		Lisame loodavasse ris
Counts	]		protsendid.
✓ Observed			Ave Celle pupp pipg v
Expected			
Percentages	Residuals		
Row	Unstandardized	+	
Column	Standardized		
Total	Adjusted standardized		

ttabelisse grupeeriva tunnuse

ali reaprotsendid.

Tulemuseks saime risttabeli:

				Sissetulek ühe pereliikme kohta kuus							
			kuni 160€	161-260€	261-385€	386-640€	üle 640€	Total			
Sugu	Mees	Count	87	144	210	154	73	668			
		% within Sugu	13,0%	21,6%	31,4%	23,1%	10,9%	100,0%			
	Naine	Count	103	186	243	202	43	777			
		% within Sugu	13,3%	23,9%	31,3%	26,0%	5,5%	100,0%			
Total		Count	190	330	453	356	116	1445			
		% within Sugu	13,1%	22,8%	31,3%	24,6%	8,0%	100,0%			

#### Sugu \* Sissetulek ühe pereliikme kohta kuus Crosstabulation

Tabelist on näha, et 13% meestest ning 13,3% naistest elab peres, kus sissetulek ühe pereliikme kohta on väiksem kui 160€ kuus jne. e tabelist tulemuste lugemiseks vaadata võrreldavate gruppide protsente.



Kindlasti ei saa võrdlemisel lähtuda gruppide sagedustest (saab ainult siis kui võrreldavad grupid on täpselt ühesuurused) ning järgmine risttabel gruppide vastuste võrdlemiseks ei sobi.

Sugu \* Sissetulek ühe pereliikme kohta kuus Crosstabulation

Count							
Sissetulek ühe pereliikme kohta kuus							
		kuni 160€	161-260€	261-385€	386-640€	üle 640€	Total
Sugu	Mees	87	144	210	154	73	668
	Naine	103	186	243	202	43	777
Total		190	330	453	356	116	1445

Risttabelit võib koostada ka nii, et

võrreldavate gruppide tulemused on paigutatud veergudesse.

Oldstetuiek	gu orossius	alation			
			Su		
			Mees	Naine	Total
Sissetulek ühe pereliikme	kuni 160€	Count	87	103	190
kohta kuus		% within Sugu	13,0%	13,3%	13,1%
	161-260€	Count	144	186	330
		% within Sugu	21,6%	23,9%	22,8%
	261-385€	Count	210	243	453
		% within Sugu	31,4%	31,3%	31,3%
	386-640€	Count	154	202	356
		% within Sugu	23,1%	26,0%	24,6%
	üle 640€	Count	73	43	116
		% within Sugu	10,9%	5,5%	8,0%
Total		Count	668	777	1445
		% within Sugu	100,0%	100,0%	100,0%

#### Sissetulek ühe pereliikme kohta kuus \* Sugu Crosstabulation



Meeles tuleb pidada, et võrdlemiseks on vaja tabelisse lisada protsendid võrdleva tunnuse järgi. Kui võrreldav tunnus (sugu) on paigutatud veergu (*Column*(s)), siis tuleb *Cells* nupu all kaasa võtta veeruprotsendid (*Percentages/Column*).

# 1.3. Risttabeli koostamine kui tunnusel on palju erinevaid väärtuseid

Mahukate risttabelite koostamisel tuleks kaaluda tunnus(t)e väärtuste grupeerimist, et saavutada ülevaatlikum ja selgem pilt andmetest.

Järgneva tabeli puhul on ühte tabelisse koondatud kaks tunnust, millel mõlemal on suhteliselt palju erinevaid väärtuseid. Tabelist tendentside ja suundumuste väljalugemine ei ole sellises olukorras alati lihtne (mõnikord on see lausa võimatu).

					Haridustase. M	illine on Teie prae	gune haridustase	19		
			Alqharidus	Pőhiharidus	Kutseharidus (ilma keskhariduset a)	Keskharidus	Kutseharidus + keskharidus	Rakenduslik kõrgharidus	Ülikooliharidu s, kraadiharidus	Total
Sissetulek ühe	kuni 160€	Count	10	39	12	61	44	19	5	190
pereliikme kohla kuus		% within Haridustase. Milline on Teie praegune haridustase?	21,3%	18,1%	14,8%	15,8%	12,6%	11,7%	2,5%	13,1%
	161-260€	Count	18	70	21	93	84	21	23	330
		% within Haridustase. Milline on Teie praegune haridustase?	38,3%	32,4%	25,9%	24,0%	24,1%	12,9%	11,4%	22,8%
	261-385€	Count	9	64	30	114	115	52	69	453
		% within Haridustase. Milline on Teie praegune haridustase?	19,1%	29,6%	37,0%	29,5%	33,0%	31,9%	34,2%	31,3%
	386-640€	Count	7	34	16	95	80	57	67	356
		% within Haridustase. Milline on Teie praegune haridustase?	14,9%	15,7%	19,8%	24,5%	22,9%	35,0%	33,2%	24,6%
	üle 640€	Count	3	9	2	24	26	14	38	116
		% within Haridustase. Milline on Teie praegune haridustase?	6,4%	4,2%	2,5%	6,2%	7,4%	8,6%	18,8%	8,0%
Total		Count	47	216	81	387	349	163	202	1445
		% within Haridustase. Milline on Teie praegune haridustase?	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Sissetulek ühe pereliikme kohta kuus \* Haridustase. Milline on Teie praegune haridustase? Crosstabulation

# 1.4. Gruppide tulemuste graafiline võrdlemine – võrdlev tulpdiagramm

Jätkame pt.1.1. püstitatud küsimusega: Millist sissetulekut esineb meeste hulgas kõige rohkem ja kas see langeb kokku naiste hulgas kõige enamlevinud sissetulekuga?

Bar Charts	Võrdlemise graafiliseks esitamisek	s kasutatakse praktikas kõige rohkem
	tulpdiagrammi võimalusi.	
Simple	Vali: Graphs/Legacy Dialogs/Bar	
Clustered Clustered Stacked Data in Chart Are Summaries for groups of cases Summaries of separate variables Values of individual cases	Dialoogiaknas vali: ☑ Clustered ☑ Summaries for groups of cases	
Define Cancel Help		
Define Clustered Par Summarie	s for Groups of Cases	Edasi vii Define Clusters by: väljale
	s for Groups of Cases	tunnus <b>sugu</b> (arupeeriv tunnus) ning
Ва	rs Represent	kategooriate esitamiseks tunnus
🖋 Vanus [Vanus] 🛛 🖆 👩	Nofcases @%ofcases	
Rahvus [Rahvus]	Cum N O Cum %	SISSETUIER Category Axis: Valjale.
Verekonnaseis [	) Other statistic (e.g. mean)	
Haridustase Milli	<u></u>	Kuna võrreldavate gruppide
🖉 🥔 Leibkonna suuru	Variable:	suurused ei ole samad, tuleb
🛷 Elukoht. Kus Te p		võrdlemisel välia tuua grupeeriva
🖉 Töökoht Mis tüüp		voraionnoon vaija tada grupoonva

	Kuna võrreldavate gruppide
	suurused ei ole samad, tuleb
	võrdlemisel välja tuua grupeeriva
stic	tunnuse väärtused protsentides.
	Selleks vali Bars Represent valikus
eliikme koht	⊙ % of cases

Change Statis Þ Majanduslik oluk... k1. Kuivõrd jätkub... Þ 🔗 k2. Kuivõrd jätkub... Category Axis: 🔗 k3. Kuivõrd jätkub... 🗞 Sissetulek ühe per k4. Kuivõrd jätkub... Þ Define Clusters by: k5. Kuivõrd jätkub...

\lambda Sugu (Sugu)

D

P

k6. Kuivõrd jätkub..



Tulemuseks saime võrdleva tulpdiagrammi:



Antud graafikul on esitatud meeste-naiste vastused lähtudes vastajate arvust (*Count*). See ei ole sisuliselt korrektne kuna võrreldavate gruppide suurused ei ole samad. Gruppide vastuste võrdlemisel tuleb kasutada protsentuaalset jaotust.

## 1.5. Kihtdiagramm

Kihtdiagrammis esitatakse võrreldavate kategooriate vastused tulpades, mis moodustavad (iga tulp eraldi) kokku 100%.



Näites võrreldakse erinevaid hinnanguid oma majanduslikule olukorrale andnud isikuid ja nende sissetulekute suurust. Loodud kihtdiagrammi asemel saaks koostada ka risttabeli:

			Sissetulek ühe pereliikme kohta kuus					
			kuni 160€	161- 260€	261- 385€	386- 640€	üle 640€	Total
Majanduslik olukord. Milline järgnevatest	Ei oska öelda	Count %	5 26,3%	5 26,3%	7 36,8%	2 10,5%	0 ,0%	19 100,0%
kirjeldustest vastab Teie arvates kõige paremini Teie	Praeguse sissetuleku juures on väga raske hakkama saada	Count %	40 38,1%	41 39,0%	16 15,2%	7 6,7%	1 1,0%	105 100,0%
leibkonna praegusele sissetulekute tasemele?	Praeguse sissetuleku juures on üsna raske hakkama saada	Count %	77 21,5%	111 31,0%	129 36,0%	36 10,1%	5 1,4%	358 100,0%
	Saan/ saame hakkama praeguse sissetuleku juures	Count %	65 8,1%	161 20,0%	276 34,3%	236 29,3%	67 8,3%	805 100,0%
	Elan/ elame hästi praeguse sissetuleku juures	Count %	3 1,9%	12 7,6%	25 15,8%	75 47,5%	43 27,2%	158 100,0%
Total		Count %	190 13,1%	330 22,8%	453 31,3%	356 24,6%	116 8,0%	1445 100,0%

# Majanduslik olukord. Milline järgnevatest kirjeldustest vastab Teie arvates kõige paremini Teie leibkonna praegusele sissetulekute tasemele? \* Sissetulek ühe pereliikme kohta kuus Crosstabulation

Kihtdiagrammi loomiseks vali: Graphs/Legacy Dialogs/Bar...

Dialoogiaknas vali:

# ☑ Stacked

 $\ensuremath{\boxtimes}$  Summaries for groups of cases

-Bars Represent					
N of cases	© % of c <u>a</u> ses				
© <u>C</u> um. N	© Cu <u>m</u> . %				
◎ Other <u>s</u> tatistic (e.g., mean)					
Variable:					
•					
C <u>h</u> ange Stat	iistic				
Category A <u>x</u> is:					
🛛 🔌 Majanduslik oluko	ord. Milline jär				
Define Stacks by:					
📩 🔒 Sissetulek ühe pe	ereliikme koht				

Tulba kõrguseks määra vastajate arv ⊙N of cases

*Category Axis* väljale vii tunnus, mille erinevaid gruppe sa omavahel võrdlema hakkad.

Define Stacks by väljale vii uuritav tunnus.

🔣 Chart Editor 1 000-Sissetulek <u>File Edit View</u> Options Elements Help ühe pereliikme 🕈 🖻 🔚 🗙 L X Axis Reference Line 800 kohta kuus Y Axis Reference Line LEZ C Count kuni 160€ Keference Line from Equation B I 600 161-260€ 🖶 🗽 🛍 🗠 Title 261-385€ Annotation 400- 386-640€
 üle 640€ Sissetulek ühe pereliikme ohta kuus 1 000 Text Box 800 200 L→ Footnote cuni 160€ 161-260€ 261-385€ 386-640€ Count 600 Book Charts in the Diagonal 0 Element 400 <u>Saan/ saame hakkama praeguse</u> sissetuleku juures E oska öelda Praeguse sissetuleku juures väga raske hakkama saada Praeguse sissetuleku juures üsna raske hakkama saada <u>.</u>⊟an/ elame hästi praeguse sissetuleku juures How Grid Lines 200 Show Derived Axis B Hide Legend oska 🛓 T<u>r</u>anspose Chart öelda Scale to 100% setuleku akkama s setuleku juures o akkama saada Scale to 100% aeguse ma praegus i juures saada g 9 9 9

Võrdlemiseks sobiva kihtdiagrammi loomiseks, ava *Chart Editori* aken ning teisenda skaala 100%-le (vali *Options* menüüst käsk *Scale to 100%*).

Kirjeldava teksti paremaks lugemiseks on antud diagrammi teljed ka ära vahetatud.

Andmesiltide lisamine ei ole hetkel ilmselt parim mõte, kuna väga väikese suurusega kihte on mitu ning nende sildid hakkaksid üksteist katma.



# 2. Gruppide võrdlemine keskväärtuse arvutamise kaudu

Uurime, kuivõrd sõltuvad hinnangud riiete ostmiseks raha jätkumisele vastajate vanusest e kas näiteks vanemad inimesed pigem ei vaja või ei osta riideid?

Kirjeldavate arvnäitajate arvutamiseks erinevate gruppide kohta vali: Analyze/Means/Compare Means...

Means	A LOUGH & A MARKET	×
Means  Sugu [Sugu]  Rahvus [Rahvus]  Koolis. Mitu aastat olete kooliharidust sa Haridustase. Milline on Teie praegune h Leibkonna suurus. Mitu inimest elab Tei Elukoht. Kus Te praegu elate? [Elukoht]	Dependent List Vanus [Vanus] Layer 1 of 1 Previous Independent List	Options
<ul> <li>Sissetulek ühe pereliikme kohta kuus [S</li> <li>Töökoht. Mis tüüpi organisatsioonis Te t</li> <li>Majanduslik olukord. Milline järgnevatest</li> <li>k1. Kuivõrd jätkub raha: korralikuks toitu</li> </ul>	Meteorie List.         Image: Reset         Cancel         Help	

Options nupu avamisel saad määrata, milliseid arvnäitajaid arvutatakse.

Means: Options			×	
Statistics:		_	<u>C</u> ell Statistics:	
Median	4		Mean	
Grouped Median			Number of Cases	
Std. Error of Mean			Standard Deviation	
Sum			Minimum	
Range			Maximum	
First				
Last				
Variance		-		
Kurtonin			l	

Lõplik tulemus kinnitab meie püstitatud hüpoteesi.

Vanus						
k2. Kuivõrd jätkub raha: Riiete ostmiseks	Mean	Ν	Std. Deviation	Minimum	Maximum	
Ei vaja seda, ei kuluta sellele	60,57	35	13,774	21	74	
Ei jätku, oleme pidanud	49,39	115	17,726	15	74	
Mõnikord jätkub, mõnikord	42,76	658	16,294	15	74	
mitte Üldiselt jätkub	42,34	693	16,334	15	74	
Total	43,49	1501	16,669	15	74	

#### Report