

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 tulpdiaagramm .....	4
1.5. Kihtdiaagramm .....	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 kuuavad 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	töökata	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€	kontoritöötaja	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.

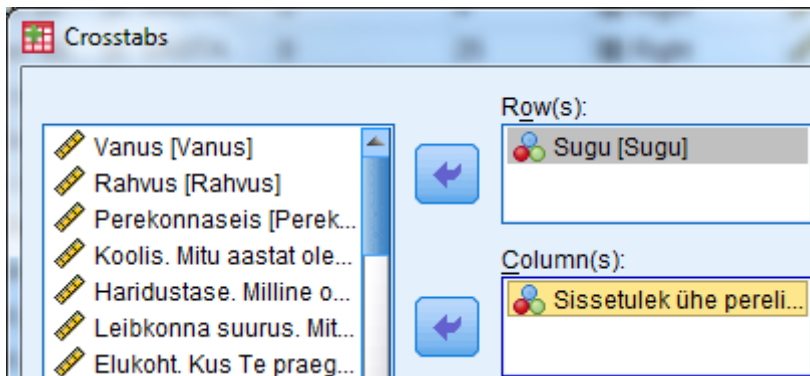
Arvutame sissetulekute sagedused ja nendest lähtuvad protsendid võrreldavate gruppide lõikes e koostame võrdleva sagedustabeli e risttabeli.

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

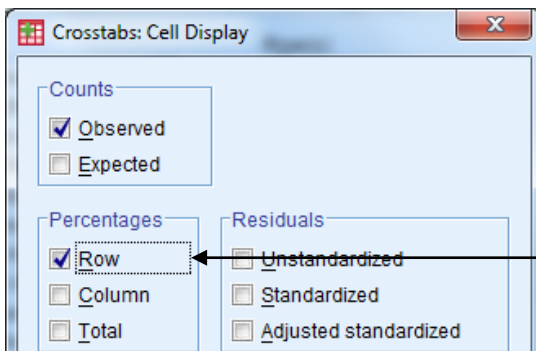
### 1.2. Risttabeli koostamine

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



Viime selle tunnuse, mille grupe 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.



Lisame loodavasse risttabelisse grupeeriva tunnuse protsendid.

Ava *Cells* nupp ning vali reaprotsendid.

Tulemuseks saime risttabeli:

Sugu * Sissetulek ühe pereliikme kohta koos Crosstabulation								
		Sissetulek ühe pereliikme kohta koos					Total	
		kuni 160€	161-260€	261-385€	386-640€	üle 640€		
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%

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					Total
		kuni 160€	161-260€	261-385€	386-640€	üle 640€	
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.

Sissetulek ühe pereliikme kohta kuus \* Sugu Crosstabulation

			Sugu		Total
			Mees	Naine	
Sissetulek ühe pereliikme kohta kuus	kuni 160€	Count	87	103	190
		% 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%

Row(s):

Column(s):

Exact...  
 Statistics...  
 Cells...  
 Format...

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.

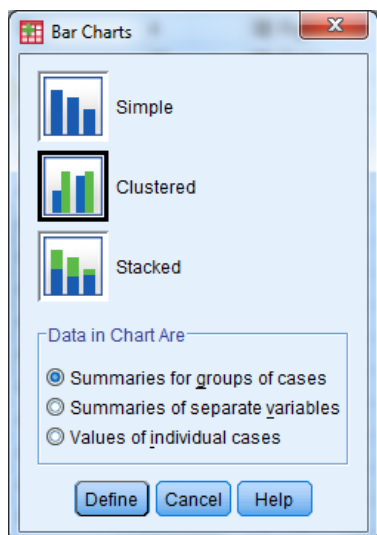
Järgneva tabeli puhul on ühte tabelisse koondatud kaks tunnust, millel mõlemal on suhteliselt palju erinevaid väärtuseid. Tabelist tendentside ja suundumuste väljalugemine ei ole sellises olukorras alati lihtne (mõnikord on see lausa võimatu).

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

			Haridustase. Milline on Teie praegune haridustase?							Total
			Algharidus	Põhiharidus	Kutseharidus (ilma keskharidusest)	Keskharidus	Kutseharidus + keskharidus	Rakenduslik kõrgharidus	Ülikooliharidus, kraadiharidus	
Sissetulek ühe pere liikme kohta kuus	kuni 160€	Count	10	39	12	61	44	19	5	190
		% 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%

### 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?

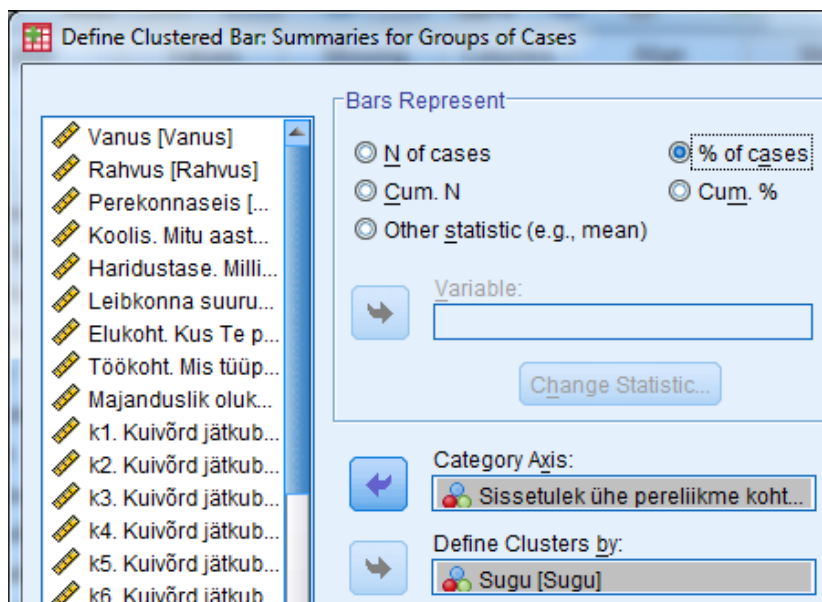


Võrdlemise graafiliseks esitamiseks kasutatakse praktikas kõige rohkem tulpdiagrammi võimalusi.

Vali: *Graphs/Legacy Dialogs/Bar...*

Dialogiaknas vali:

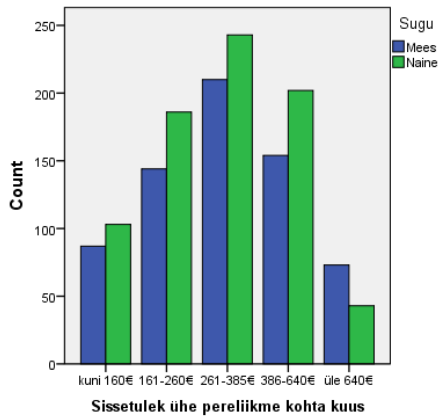
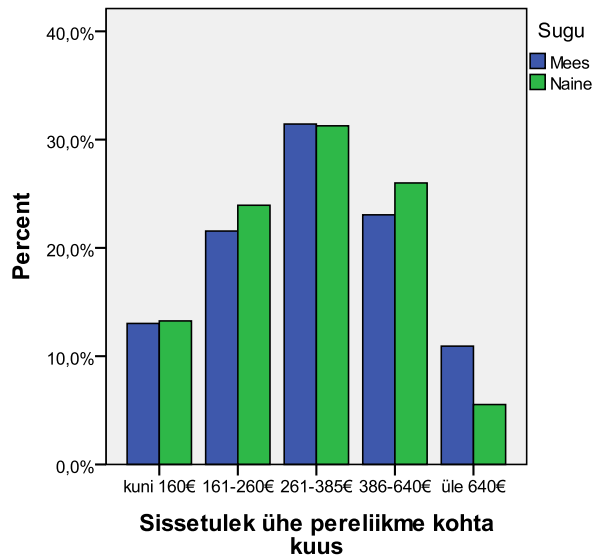
- Clustered
- Summaries for groups of cases



Edasi vii *Define Clusters by*: väljale tunnus **sugu** (grupeeriv tunnus) ning kategooriate esitamiseks tunnus **sissetulek** *Category Axis*: väljale.

Kuna võrreldavate gruppide suurus ei ole samad, tuleb võrdlemisel välja tuua grupeeriva tunnuse väärtused protsentides. Selleks vali *Bars Represent* valikus  % of cases

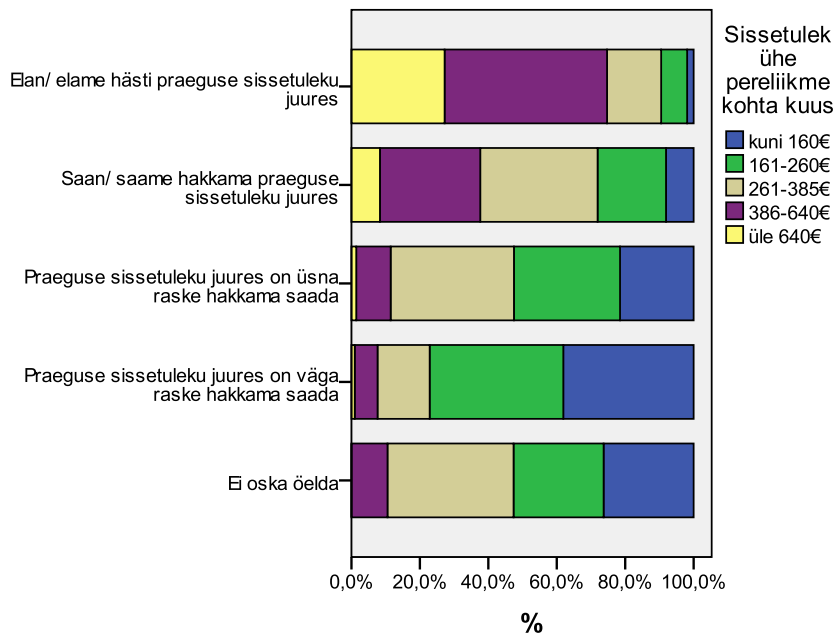
Tulemuseks saime võrdleva tulppiagrammi:



Antud graafikul on esitatud meeste-naiste vastused lähtudes vastajate arvust (*Count*). See ei ole sisuliselt korrektne kuna võrreldavate gruppide suurus 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:

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

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

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

Dialogiaknas vali:

- Stacked
- Summaries for groups of cases

**Bars Represent**

N of cases       % of cases  
 Cum. N       Cum. %  
 Other statistic (e.g., mean)

Variable:

[Change Statistic...](#)

---

Category Axis:

---

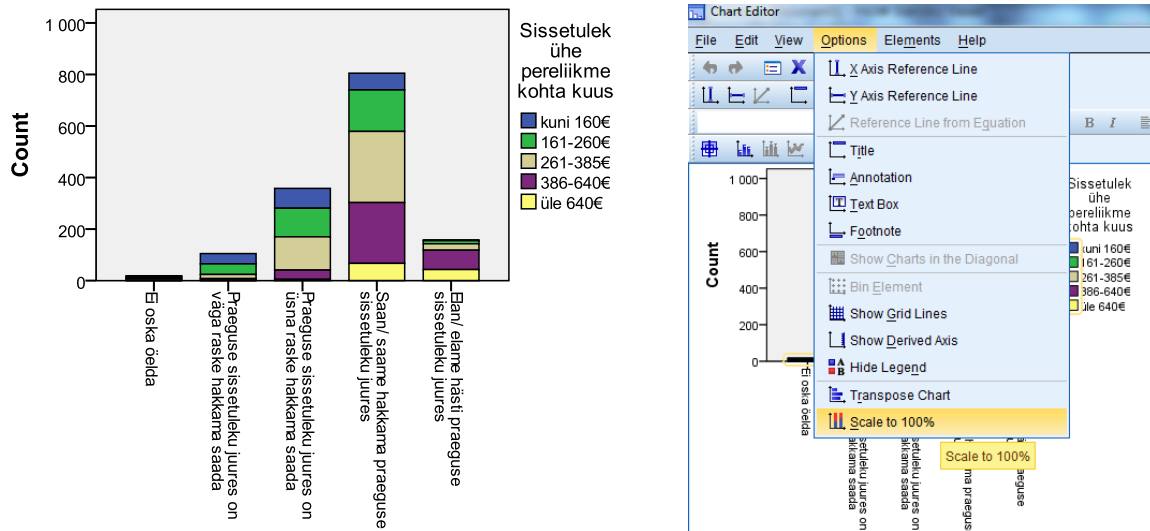
Define Stacks by:

Tulba kõrguseks määra vastajate arv  N of cases

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

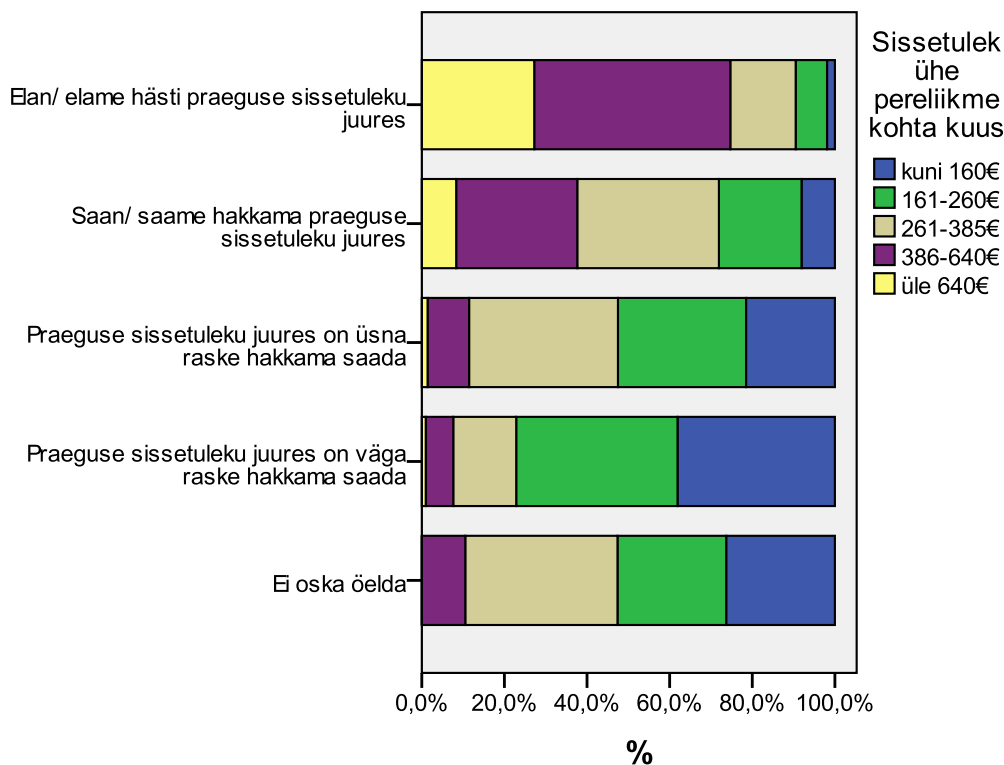
*Define Stacks by* väljale vii uuritav tunnus.

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.

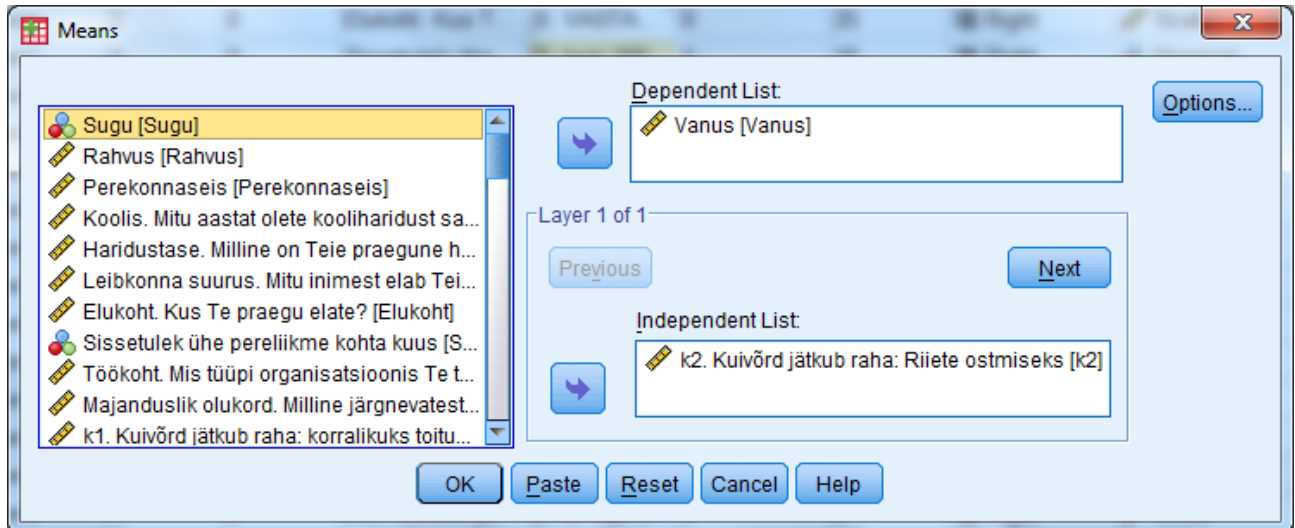
Andmesiltide lisamine ei ole hetkel ilmselt parim mõte, kuna väga väikese suurusega kihte on mitu ning nende sildid hakkaksid üksteist katma.



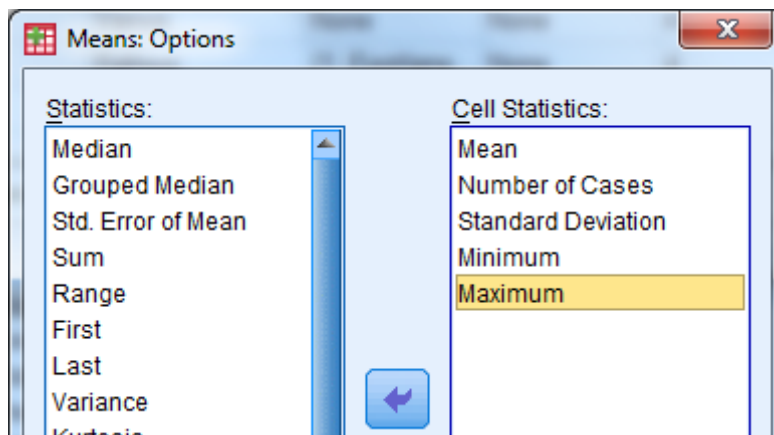
## 2. Gruppide võrdlemine keskväärtuse arvutamise kaudu

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...*



*Options* nupu avamisel saad määrata, milliseid arvnäitajaid arvutatakse.



Lõplik tulemus kinnitab meie püstitatud hüpoteesi.

### Report

Vanus

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