

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.

SAGEDUSTABELI KOOSTAMINE

Anname ülevaate ühe tunnuse väärtustest nende kokkuloendamise teel.

1. Koostame ülevaate tunnuse **haridustase** vastustest.

Valime *Analyze/Frequencies*



JÄRJESTUS-TUNNUS

Vali tunnus Haridustase ja vii see *Variable(s)* väljale.

Oluline on, et *Display frequency tables* ees oleks linnuke – siis koostatakse sagedustabel.

Tulemus kuvatakse tulemuste (*Output*) faili.

Statistics

N	Valid	1507
	Missing	0

Haridustase. Milline on Teie praegune haridustase?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Algharidus	47	3,1	3,1	3,1
Põhiharidus	228	15,1	15,1	18,2
Kutseharidus (ilma keskhariduseta)	83	5,5	5,5	23,8
Keskharidus	401	26,6	26,6	50,4
Kutseharidus + keskharidus	367	24,4	24,4	74,7
Rakenduslik kõrgharidus	170	11,3	11,3	86,0
Ülikooliharidus, kraadiharidus	211	14,0	14,0	100,0
Total	1507	100,0	100,0	

Esmalt antakse üldine ülevaade (*Statistics*) vastajate vastustest (kui paljud vastasid *Valid* ja kui palju ei vastanud sellele küsimusele *Missing*)

Sagedustabelis esitatakse tunnuse väärtused (*Valid*), nende esinemissagedused (*Frequency*) ning protsendid (*Percent*).

Tabeli kujundamiseks (sõnade muutmiseks, osa kustutamiseks) tee selle peal topeltklõps.

Peale tabeli valimist (topeltklõps tabeli peal) leiad lisavahendeid *Format* menüüst. Näiteks *Table Looks* võimaldab kujundada tabelit (värvilahendused).

Kujundatud lõpptulemus võiks välja näha näiteks selline:

Tabel 1. Vastajate kõrgem omandatud haridustase

	Vastajate arv	Osakaal
Algharidus	47	3,1%
Põhiharidus	228	15,1%
Kutseharidus (ilma keskhariduseta)	83	5,5%
Keskharidus	401	26,6%
Kutseharidus + keskharidus	367	24,4%
Rakenduslik kõrgharidus	170	11,3%
Ülikooliharidus, kraadiharidus	211	14,0%
KOKKU	1507	100,0

NIMITUNNUS

2. Koostame ülevaate tunnuse **perekonnaseis** vastustest.

Valime *Analyze/Frequencies*

Viime tunnuse Perekonnaseis *Variable(s)* väljale ning kinnitame oma valiku.

Statistics

Perekonnaseis		
N	Valid	1506
	Missing	1

Perekonnaseis

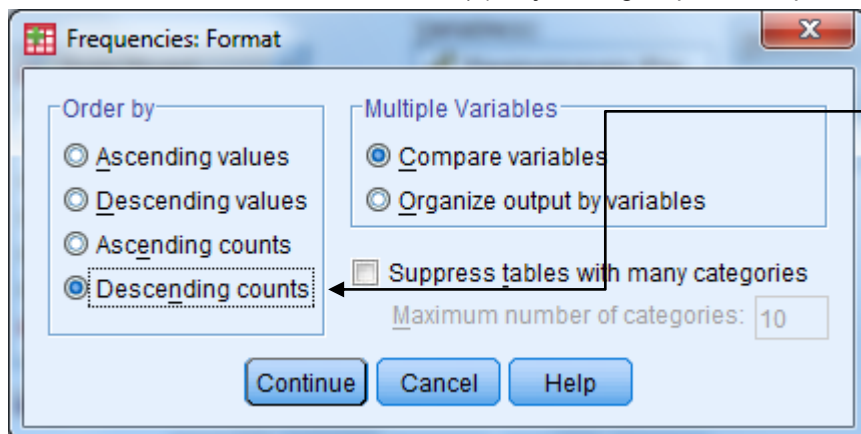
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid VASTAMATA	2	,1	,1	,1
Üksik (pole olnud abielus)	331	22,0	22,0	22,1
Ametlikus abielus	603	40,0	40,0	62,2
Vabaabielus (elan koos partneriga)	288	19,1	19,1	81,3
Lahutatud / Elan lahus	170	11,3	11,3	92,6
Lesk	112	7,4	7,4	100,0
Total	1506	99,9	100,0	
Missing System	1	,1		
Total	1507	100,0		

Esialgne tulemus on selline:

Juhul kui tabelis toodud kategooriad **ei ole sisuliselt tähenduslikus järjekorras**, siis järjestatakse tabeli read sageduste/osakaalude järgi.

Sagedustabeli **järjestamiseks sageduste järgi** koostame uue sagedustabeli. Valime *Analyze/Frequencies*

Viime tunnuse Perekonnaseis *Variable(s)* väljale ning klõpsame nupul *Format*.



Määrame loodava tabeli järjestatavuse (*Order by*) nii, et tulemused esitataks sageduste kahanevas järjekorras (kõige suuremad väärtused ülal, väiksemad allpool)

Kujundatud lõpptulemus võiks välja näha näiteks selline:

Tabel 2. Perekonnaseis

	Vastajate arv	Osakaal
Ametlikus abielus	603	40,0%
Üksik (pole olnud abielus)	331	22,0%
Vabaabielus (elan koos partneriga)	288	19,1%
Lahutatud / Elan lahus	170	11,3%
Lesk	112	7,4%
Vastamata	2	,1%
Kokku	1506	99,9%
Vastus puudub	1	,1%
KOKKU	1507	100,0%

3. Intervalltunnuse puhul tuleb vahet teha, kas tunnusel on vähe või palju erinevaid väärtuseid.

Väheste erinevate väärtustega tunnuse korral järgi näidet nr 1 (haridustase).

**INTERVALL-
TUNNUS**

Vanus

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 21	39	2,6	2,6	2,6
49	37	2,5	2,5	5,0
18	35	2,3	2,3	7,4
39	35	2,3	2,3	9,7
48	35	2,3	2,3	12,0
53	33	2,2	2,2	14,2
25	31	2,1	2,1	16,3
52	31	2,1	2,1	18,3
58	31	2,1	2,1	20,4
74	31	2,1	2,1	22,4
37	30	2,0	2,0	24,4
20	29	1,9	1,9	26,3
28	29	1,9	1,9	28,3
42	29	1,9	1,9	30,2
59	29	1,9	1,9	32,1
62	29	1,9	1,9	34,0
64	29	1,9	1,9	36,0
17	28	1,9	1,9	37,8
27	28	1,9	1,9	39,7
31	28	1,9	1,9	41,5
36	28	1,9	1,9	43,4
40	28	1,9	1,9	45,3
67	28	1,9	1,9	47,1
29	27	1,8	1,8	48,9
35	27	1,8	1,8	50,7
43	27	1,8	1,8	52,5
45	27	1,8	1,8	54,3
46	27	1,8	1,8	56,1



Kui tunnusel on aga **palju erinevaid väärtuseid**, näiteks sissetulekud on kõikidel vastajatel tõenäoliselt erinevad, siis sagedustabel andmete kokkuvõtmiseks **ei sobi**.

Andmestikus kultuur.sav on selliseks tunnuseks vanus. Koostades vanuse väärtustest sagedustabeli, on see liiga mahukas, et seda andmete esitamiseks kasutada.

Kui on aga tingimata vaja esitada paljude erinevate väärtustega intervalltunnuse tulemusi tabeli kujul, siis tuleks kaaluda andmete koondamist väiksematesse gruppidesse (*Transform/Visual Binning*) ning koostada sagedustabel grupeeritud väärtustest.

Grupeerime tunnuse vanus väärtused (Transform/Visual Binning) ning moodustame järgmised vahemikud:

Koostades grupeeritud tunnuse väärtuste kohta sagedustabeli, võiks **kujundatud lõpptulemus** välja näha näiteks selline:

Tabel 3. Vanus

	Vastajate arv	Osakaal
<= 19	115	7,6%
20 - 29	278	18,4%
30 - 39	254	16,9%
40 - 49	284	18,8%
50 - 59	260	17,3%
60 - 69	216	14,3%
70+	100	6,6%
KOKKU	1507	100,0%

**BINAARNE-
TUNNUS**

4. Binaarse tunnuse (sugu) väärtuste **kokkuloendamiseks** sagedustabel sobib.

		Sugu			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Mees	698	46,3	46,3	46,3
	Naine	809	53,7	53,7	100,0
Total		1507	100,0	100,0	

Aga kas sellise tabeli töösse lisamine on põhjendatud ja vajalik?

Ilmselgelt mitte, kuna iga tulem (tabel, diagramm vms) on vaja pealkirjastada ning sellele on vaja lisada järelalus. Antud juhul dubleerikisme me sama infot lausa kaks korda.

Soolise jaotuse kohta piisab kui kirjutada:

Kokku osales uuringus 1507 vastajat, kellest 698 (46,3%) olid mehed ja 809 (53,7%) naised.