Sisukord

1.SISSEJUHATUS	2
2.EXCELI TÖÖLEHT NING ANDMETABELID	2
3.DIAGRAMMID	
3.1Diagrammide loomine	
3.2Diagrammide redigeerimine	4
3.3Diagrammide tööriistariba	5
3.4Histogrammid	6
4.STATISTIKA-PROTSEDUURID EXCELIS	10
4 1Kirifi davad apvnäitaiad	
4. 2V AHEKOKKI IVÕTETE TEGEMINE OTSE ANDMETABEI ISSE	10
4 3K OOND- JA RISTTAREI ID	12
4 3 1Sagedus- või koondtaheli koostamine	12
4 3 2Risttaheli redigeerimine	13
44Väärtuste suhteline asukoht variatsioonireas	14
4.5Korrelatsioon	
4.6Regressioon	
4.7t-test kahe valimi keskväärtuste võrdlemiseks	
4.8Põhilised statistika-funktsioonid	
5.ANDMEBAASI-PROTSEDUURID EXCELIS	
5.1Andmete sisestamine töölehel ja vastavas dialoogiaknas	
5.2Andmetabeli sorteerimine	
5.3Kirjete leidmine	
5.4Väljavõtete tegemine andmetabelist	
5.4.1Tingimus(t)ele vastavate kirjete kuvamine andmetabeli piirkonnas	
5.4.2Alamtabeli loomine uude töölehe piirkonda	
-	

1.Sissejuhatus

Excel on tabelarvutusprogramm, mis pakub mugavaid võimalusi andmete igakülgseks töötlemiseks ning analüüsimiseks. Käesolev õppematerjal ei sisalda Exceli kasutamise algkursust vaid püüab anda ülevaate põhilistest andmetöötlusvahenditest Excelis.

2. Exceli tööleht ning andmetabelid.

Erinevalt statistikapakettidest ning andmebaasisüsteemidest võib Exceli töölehele andmeid paigutada küllalt vabalt st. töölehe ridade ja veergude otstarve pole kindlalt määratud. Vaatamata sellele, tuleks enne andmete sisestamist läbi mõelda, milleks sisestatavaid andmeid hiljem kasutama hakatakse ning milliseid analüüsimeetodeid kasutada soovitakse. Kui tegemist on statistilise andmestikuga või lihtsa andmebaasiga, siis tuleks kindlasti järgida klassikalise andmetabeli ülesehitust: iga veerg andmetabelis vastaku ühele tunnusele ning iga rida vastavalt ühele objektile.

Andmetabel Exceli töölehel ei pea algama töölehe ülemisest vasakust nurgast st lahtrist A1, kuid tihti on mugavam töötada, kui töölehe esimesse ritta sisestatakse tunnusenimed ning iga järgmine rida on reserveeritud ühte objekti kirjeldavatele andmetele. Kui on vaja säilitada objektide järjekorranumbreid või objekte identifitseerivaid koode, siis tuleb need sisestada andmetabeli esimesse veergu, sest töölehe ridade numbrid järjekorranumbriteks ega identifikaatoriteks ei sobi.

Kuigi ühele töölehele (Sheet) võib koostada mitu erinevat andmetabelit on Exceli uuemates versioonides (5.0 ja uuemad) mugav võimalus paigutada erinevaid andmetabelid ühes failis e. töövihikus erinevatele töölehtedele. See lihtsustab märgatavalt andmete haldamist ning vähendab töölehe mõttetut edasi-tagasi kerimist. Soovitus: hoidke oma originaalandmestikku eraldi lehel (st paigutage kogu tulem uutele lehtedele)!

	X	Microsoft Exce	el - klass.xls									_ 8 ×
Valomiroal toimub	1	Eile Edit Vie	w <u>I</u> nsert F <u>o</u> rm	at <u>T</u> ools <u>D</u> ata <u>W</u>	/indow <u>H</u> elp							_ 8 ×
valenniear tonnuo	[) 🛩 🖪 🖉	j 🖪 💞 👗	🖻 🛍 🍼 🔊	🝷 C4 👻 🍓	ኛ Σ f*		🛍 🔮 🦧	100% 💌	(2)		
andmete ja valemite	Т	imes New Roman	• 8 •	B <i>I</i> <u>U</u> ≣		§%,	00. 0	律律目	- 🕭 - 🛔	4 -		
sisestamine ning		~~~	_			- 🗗 🔛	• 🗉 🕯		8 8			
redigeerimine		F5	-	-				and the second				
J	1		B	C	D	E	F	G	H	I	J	K 🔺
	1		SUGU	TESTI	TEST2	HINNE						
Veeru päises on	1	Alice	N	84	95	3						
tähtedega märgitud	/ 3	Becka	N	98	105	4						
untedega margitud	4	Gail	N	90	88	3		-				
veeru aadress	5	Karen	N	77	107	5						
	6	Kathy	N	84,5	90	4						
	7	Mary	N	112	106	5						
Rea päises on	8	Sandy	N	60,5	78,5	2						
numbritega märgitud		Sharon	N	112,5	115	5						
roo oodroog	10	Tammy	N	102,5	100,5	4						
iea aduless	11	Alfred	M	112,5	109	5						
	12	Duke	м	102,5	118,5	5						
	13	Guido	м	133	135	5						
Töölehe vahetamiseks	14	James	M	83	79,5	2						
klõpsi lehe allservas –	1.5	Jeffrey	м	84	90	3						
oleval nunul tekstiga			/ Sheet1 / She	eet2 /	576471			4			1.	
	Re	ady										
Sneet		Start 🛛 🛃	2 2	🔍 Exploring 🔋	Microsoft	Kicrosof	i 🗩 Co	rel CAP		En	\$ €])0£	11:24

3.Diagrammid

Vast kõige enam levinud meetod andmete esitamiseks on graafikud ja diagrammid. Selleks, et Excelis diagramme luua peab teie töölehel olema (koond)andmeid sisaldav tabel, mida te illustreerida tahate.

Excel lubab teil valida neljateistkümne diagrammide põhitüübi vahel, millest igal on veel omakorda alamtüübid ning lisaks võib kogenum kasutaja ka oma äranägemise järgi erinevaid diagrammide tüüpe ühendada. Mõned neist tüüpidest erinevad üksteisest ainult väliste efektide poolest, teised on aga sobivad sisuliselt erinevate andmete jaoks. Tuleb meeles pidada, et mitte kõik pakutavad diagrammid ei sobi igasuguste andmete illustreerimiseks ning tulemuse sisulise korrektsuse eest vastutab <u>alati</u> diagrammi koostaja, mitte arvuti. Seepärast on mõistlik enne diagrammi koostamist põhjalikult läbi mõelda, millist joonist te saada tahate ning alles siis vastav diagramm arvutilt "tellida".

3.1Diagrammide loomine

- Selekteeri lahtrid, milles oleva info põhjal soovid diagrammi koostada (seda saab ka hiljem teha, kuid soovitavalt võiks siiski sellest alustada)
- Vajuta nuppu (ChartWizard) või vali menüü Insert/Chart...

• Vali sobiv diagrammi tüüp ja alamtüüp

• Vajuta nuppu *Next* >



• Vajadusel sisesta või paranda piirkonna aadress, milles olevaid andmeid soovid diagrammi loomisel kasutada.

NB! Kui märkisid vajaliku piirkonna enne nupule wajutamist korrektselt, siis pole vaja siin tegevust korrata.

- Määra andmeseeriate paiknemine (*Row* reas, *Column* veerus);
- Vajuta nuppu *Next* >

Lhart <u>ti</u> tle:		-				20
Category (X) axis:	50%				_	<u>-</u> 21
I	40 %	-			1	<u>9</u> 2
						- Series 1
Second category (X) axis:	20 %					- 13
J	10 %	-			-	-
Securid Value (1) axis.	0%	Tallinn	Hayumaa	Kirde-Eesti	Maad	•

• Vastavalt valitud diagrammi tüübile saad määrata:

Titles	digrammi ning telgede pealkirjad
Axes	telgede kuvamise/mittekuvamise ning tüübi
Gridlines	abijoonte kuvamise/mittekuvamise
Legend	legendi kuvamise/mittekuvamise ning asukoha
Data Labels	andmeseeriaid kirjeldava lisainfo kuvamise/mittekuvamise
Data Table	diagrammi loomisel aluseks olnud andmetabeli
	kuvamise/mittekuvamise

NB! Kõikide valikute tagajärjel toimuvad muudatused kajastuvad näiteaknas!

- Vajuta nuppu *Next* >
- Vali kas diagramm luuakse: *As a new sheet:* kui uus eraldi lehekülg *As object in:* kui objekt valitud leheküljel.
- Vajuta nuppu Finish

3.2Diagrammide redigeerimine

Olemasolevate jooniste **ümberpaigutamiseks** vajuta joonise piirkonnas hiire vasak klahv alla ning vea joonist klahvi all hoides sulle sobivas suunas. Vabasta hiire klahv, kui joonise koht on sobiv.

Joonise **suuruse muutmiseks** selekteeri joonis (üks klõps) ning vea joonise raamjoonele tekkinud täpikesi sobivas suunas. Nurgas olevad täpikesed võimaldavad muuta jooniste suurust proportsionaalselt kahel küljel korraga; serva keskel olevad täpikesed liigutavad vaid vastavat külge moonutades seega digrammi proportsioone.

Joonise ümberkujundamiseks:

- Selekteeri joonis (1 klõps)
- Kasuta tekkinud menüüd Chart... diagrammi põhimäärangute muutmiseks
- Klõpsi hiire parema klahviga joonise elemendil, mida soovid muuta ning vali tekkinud rippmenüüst sobiv tegevus

või

selekteeri joonise element, mida soovid muuta ning vali vastav tegevus (esimene alammenüü) menüüst *Format*



3.3Diagrammide tööriistariba

Diagrammide tööriistariba kuvamiseks vali View/Toolbars.../Chart



3.4Histogrammid

Kõik eelnevalt kirjeldatud viisil saadavad diagrammid sobivad juhul, kui töölehel on selline andmetabel, mille iga lahtri väärtus annab ühe diagrammi elemendi (tulba, sektori, punkti vms). Statistilise andmetabeli tunnuse kirjeldamiseks sellised diagrammid sageli ei sobi. Näiteks, olgu tabelis klassi õpilaste nimed ning nende veerandihinded matemaatikas. Kui eesmärgiks on võrrelda õpilasi nende hinnete põhjal, siis võiks küll teha tulpdiagrammi (mõistlik oleks andmetabel eelnevalt sorteerida), kuid sellisel kujul on ta üsna ebaülevaatlik ning soovitavalt tuleks selliseid diagramme vältida.



Kui eesmärgiks on kirjeldada klassi kui tervikut st. näidata kui suur osa õpilastest said viie, kui suur osa nelja jne, siis on selleks kaks erinevat teed:

- 1. kõigepealt tuleks koostada koondtabel (nt kasutades funktsiooni *frequency*) ning alles selle põhjal vastav (tulp)diagramm,
- 2. tellida histogramm, mille puhul kasutaja annab ette vahemiku otspunktid ning Excel koondab ühe tulba alla kõik sellesse vahemikku kuuluvad väärtused.

Histogrammi koostamiseks:

• Sisesta töölehe vabasse veergu üksteise alla väärtused, mis määravad intervallide ülemised otspunktid (mõistlik on veerule anda ka pealkiri).

N1. Kui sisestad väärtused 2, 3, 4 ja 5 siis moodustub neli+üks intervalli e. vahemikku:

- 1) (kuni) hinne 2
- 2) (üle kahe kuni) hinne 3
- 3) (üle kolme kuni) hinne 4
- 4) (üle nelja kuni) hinne 5
- 5) (üle viie)
- N2. Tunnuse sissetulek puhul sisestades

väärtused 1000, 2000, 3000, 4000 tekivad vahemikud:

- 1) ... kuni 1000 kr
- 2) 1001 kr kuni 2000 kr
- 3) 2001 kr kuni 3000 kr
- 4) 3001 kr kuni 4000 kr
- 5) 4001 kr kuni ...

Nagu näed, saab histogrammi koostada ka pidevast arvtunnusest (n testitulemustest, pikkusest, kaalust, palgast jne). Siin tuleb aga jälgida seda, et esimese ja viimase vahemiku pikkused võivad jääda teistest erinevaks, mis on tihti eksitav tulemuste tõlgendamisel.

• Vali Tools/Data Analysis... ning nimekirjast Histogram. Vajuta OK.



- Sisesta väljale *Input Range* andmepiirkonna aadress (võid kasutada hiirega märkimist)
- Sisesta väljale *Bin Range* intervalli otspunktide aadress (võid kasutada hiirega märkimist)

NB! kui märgitud andmepiirkond sisaldab väljanime, siis peab pealkiri olema ka intervalli otspunktide veerul!

• Määra, kuhu paigutada tulem

Output Rangesamale lehele; sisesta piirkonna ülemise vasaku lahtri aadressNew Worksheetuuele lehele; võid sisestada töölehe aadressiNew Workbookuude töövihikusse ehk faili

• Määra, millist graafikut soovid

Pareto

sorteeritud diagramm (tulbad joonisel järjestatakse vastavalt sagedusele)

Cumulative Percentage sagedust

ge diagrammil väljendatakse kumulatiivset e kuhjuvat

Chart Output tavaline sagedusjaotust iseloomustav diagramm NB! Alati, ka siis kui jätad kõik graafikud märkimata, saad tulemiks sagedustabeli.



Ülaltoodud näite puhul võib tulemit pidada pigem tulpdiagrammiks kui histogrammiks (historgamm on



tulpdiagrammi spetsiifiline vorm, kus üks telg on alati (pidev) arvtelg ja teisel teljel on sagedus).

Kuna igasse vahemikku kuulub ainult üks väärtus, mis oli määratud vahemiku ülemiseks otspunktiks, siis on ka põhjendatud, et see väärtus asub diagrammil vahemiku keskpaigas.

Järgmise näite puhul, kus diagramm on mõeldud andma ülevaadet õpilaste pikkustest, on aga Exceli tulem ilma kasutajapoolse sekkumiseta täiesti eksitav:



Pideva tunnuse sagedusjaotust iseloomustaval histogrammil peaks tulbad asuma ilma vahedeta üksteise kõrval, et rõhutada seda, et kõik väärtused skaalal on võimalikud. Selle reegli eiramine ei ole õnneks väga eksitav kuigi võimendab ekslikku muljet, nagu oleks pikkused olnud mõõdetud 5cm täpsusega. Samas on väga eksitav, et vahemiku lõpp-punktid on paigutatud tulba keskele, sest tegelikult ei kuulu ju nt teise tulba alla mitte ainult need, kelle pikkus on 170cm vaid ka kõik need, kelle pikkus jääb 165cm ja 170cm vahele. (Teadmiseks, et tavaliselt statistikapaketid lähtuvad vahemike alguspunktidest ning see, et Excel eirab seda tava lähtudes vahemiku lõpp-punktist, suurendab segadust veelgi.)

Esimese vea parandamiseks tuleb joonise tulbal parema klahvi klõpsuga avada kiirmenüü ning valida sealt *Format Data Series*... ning edasi lehel *Options* väljal *Gap width:* muuta väärtus nulliks.

Teise vea parandamiseks tuleb vastav info muuta tekkinud sagedustabelis ning Excel redigeerib tabeli põhjal automaatselt ka diagrammi (vt. allpool toodud näidet).



Põhimõtteliselt saaks seega asendada vahemiku ka tema keskpunkti väärtusega (mida statistikapaketid tihti teevad). Antud näite puhul oleks see vähem eksitav kui esialgne

variant, kuid harjumatu kasutaja võib siiski arvata, et tegemist on üksikväärtuse ja mitte vahemikuga.

4. Statistika-protseduurid Excelis

4.1Kirjeldavad arvnäitajad

Exceli andmeanalüüsi töövahendite komplekt (*Tools/Add-Ins/ Analysis ToolPak*) pakub mitmeid statistika põhimeetodeid, sealhulgas kirjeldavaid arvnäitajaid.

- Vali Tools/Data Analysis...
- Vali Descriptive Statistics



	TESTI		
Keskvaartus	Mean	100,6	
Standardviga	Standard Error	5,0	
Mediaan	Median	99,5	
Mood	Mode	84,0	
Standardhalve	Standard Deviation	21,7	
Valimi dispersioon	Sample Variance	468,9	
Ekstsess	Kurtosis	0,3	
Asummeetria	Skewness	0,5	
Ulatus	Range	89,5	
Mın. väärtus	Minimum	60,5	
Max vaartus	Maxımum	150,0	
Summa	Sum	1910,5	
Kirjete arv	Count	19,0	
USAIQUSINCEIVAII:			
Mean \pm Conf.Level	Confidence Level(95,0%	10,4	

4.2Vahekokkuvõtete tegemine otse andmetabelisse

Sul on võimalus teha vahekokkuvõtteid otse andmetabelisse. Selleks peaks andmetabel olema eelnevalt sorteeritud selle tunnuse järgi, mille põhjal moodustuvate gruppide jaoks sa kokkuvõtteid teha soovid.

- Sorteeri andmetabel ja paiguta aktiivne lahter andmetabeli piirkonda
- Vali Data/Subtotals...

Vali grupeeriy tunnus	Subtotal	? ×
van grapeenv tainas	At Each Change in:	
Vali kokkuvõttev	SUGU	
funktsioon	Use Function:	
	Average	🔄 👔 📙 <u>R</u> emove All
Märgi tunnus(ed), mille kohta kokkuvõtteid soovid	Add Subtotal to:	······
	HINNE	
Asendada olemasolevad		
коккиνоней	Replace Current Subtotals	
Lehevahetus gruppide	I. Page Break Between Groups	
vahel	1 <u>S</u> ummaly Below Data	
Paigutada kokkuvõtted gr (märkimata, siis grupp	ruppide alla, bide ette)	Kõigi vahekokkuvõtete eemaldamiseks

NIMI	SUGU	KAAL	TESTI	TEST2	HINNE
Alice	Ν	56,5	84	95	3
Becka	Ν	65,3	98	105	4
Gail	Ν	64,3	90	88	3
Karen	Ν	56,3	77	107	5
Kathy	Ν	59,8	84,5	90	4
Mary	Ν	66,5	112	106	5
Sandy	Ν	51,3	60,5	78,5	2
Sharon	Ν	62,5	112,5	115	5
Tammy	Ν	62,8	102,5	100,5	4
-	N Average		91,2	98,3	3,9
Alfred	М	69	112,5	109	5
Duke	М	63,5	102,5	118,5	5
Guido	М	67	133	135	5
James	М	57,3	83	79,5	2
Jeffrey	М	62,5	84	90	3
John	М	59	99,5	119	5
Philip	М	72	150	136	5
Robert	М	64,8	128	133	5
Thomas	М	57,5	85	102	4
William	М	66,5	112	107,5	4
	M Average		109,0	113,0	4,3
	Grand Average		100,6	106,0	4,1

4.3Koond- ja risttabelid

Exceli protseduur *PivotTable* võimaldab luua statistilise andmestiku põhjal nii sagedustabeleid kui ka muid kokkuvõtvaid tabeleid. Järgnevas on toodud mõningaid näiteid lihtsamatest sagedus- ja koondtabelitest:

Sagedustabel	
HINNE	Total
2	2
3	3
4	5
5	9
Grand Total	19

Sagedustabel	SIICII	r	
Sageuuscaser	2000		
HINNE	М	N	Grand Total
2	10,0%	11,1%	10,5%
3	10,0%	22,2%	15,8%
4	20,0%	33,3%	26,3%
5	60,0%	33,3%	47,4%
Grand Total	100,0%	100,0%	100,0%

Keskmine HINNE		
SUGU	Total	
M		4,30
N		3,89
Grand Total		4,11

		SUGU	
HINNE	Keskmised	М	Ν
2	Test l	83,0	60,5
	Test 2	79,5	78,5
3	Test 1	84,0	87,0
	Test 2	90,0	91,5
4	Test 1	98,5	95,0
	Test 2	104,8	98,5
5	Test 1	120,9	100,5
	Test 2	125,1	109,3

4.3.1Sagedus- või koondtabeli koostamine

- (Jäta aktiivne lahter andmestiku piirkonda ja) Vali menüü Data/PivotTable ...
- Kui lähteandmed asuvad Exceli töölehel aktsepteeri valik *Microsoft Excel List or Database* ja vajuta nuppu *Next*
- Sisesta lähteandmete piirkonna aadress väljale Range
 * Kui aktiivne lahter oli jäetud andmetabeli piirkonda, siis sisestatakse kogu tabeli aadress automaatselt
- Vajuta nuppu Layout...



• Risttabeli koostamiseks vea grupeeriv(ad) tunnus(ed) väljale *ROW* siin määrad tunnuse, mille väärtused pannakse reapäisesse ja/või *COLUMN* siin määrad tunnuse, mille väärtused pannakse veerupäisesse

- Kui soovid mingi tunnuse väärtuste põhjal koostada erinevad risttabelid, siis vea vastav tunnus väljale *PAGE* (N naistest ja meestest eraldi tabelite loomiseks)
- Lahtrite sisu määramiseks vea väljale *DATA* tunnus, mille kohta soovid teha kokkuvõtte (sagedustabeli puhul on selleks üks rea- või veerupäiseid määravatest tunnustest)
- Kokkuvõtva funktsiooni muutmiseks tee topeltklõps väljale *DATA* veetud tunnusel ning vali nimekirjast sobiv funktsioon:

0	5
Sum	väärtuste summa
Count	sagedus e. kirjete arv
Average	aritmeetiline keskmine
Max	maksimaalne väärtus
Min	minimaalne väärtus
Product	korrutis
CountNums	arvuliste väärtuste arv
StdDev	hinnang üldkogumi standardhälbele
StdDevp	valimi standardhälve
Var	hinnang üldkogumi dispersioonile
Varp	valimi dispersioon

PivotTable Field		?
Source Field: HII	NNE	OK
Name: Count of HIN	INE	Cancel
<u>S</u> ümmarize by: Sum	.	<u>D</u> elete
Count Average		<u>N</u> umber
Max Min		Options >>
Product Count Nums	-	

- Kokkuvõtva tabeli lahtrite numbriformaadi määramiseks vajuta nuppu Number...
- Kui soovid lisavõimalusi lahtri sisu määrava funktsiooni koostamiseks siis vajuta nuppu *Options>>*

% of row	lahtri osakaal reas
% of column	lahtri osakaal veerus
% of total	lahtri osakaal kogu tabelis

4.3.2Risttabeli redigeerimine

Sul on võimalus muuta olemasoleva risttabeli kokkuvõtvaid funktsioone ning rea ja/või veeru päiseid määravate tunnuste kuvamist.

- Selekteeri risttabeli väli, mida soovid muuta
- Vali Data/PivotTable Field...
- * Kokkuvõtva funktsiooni muutmisel toimi samuti nagu funktsiooni esialgsel valimisel;

* Rea ja/või veeru päiste muutmisel toimi vastavalt järgnevale joonisele:

Andmetöötlus Excel'is

Vali, kas selle tunnuse	PivotTable Field	? ×
väärtuste põhjal moodus-	N <u>a</u> me: HINNE	OK
veerud või eraldi tabelid	Orientation	Cancel
	Subtotals	<u>D</u> elete
Kui soovid muuta vahe-	C Automatic Sum	
kokkuvõtete funktsiooni,	C Custom Average	
siis vali Custom ning	Max Min	
sobiv funktsioon	Product	
	Hide <u>I</u> tems:	
Selekteeri need	2	
väärtused,	4	
mida sa risttabelis	5	
kuvada ei soovi		

- Andmeväljal oleva üksteise alla paigutatud arvnäitajate ümberpaigutamiseks eraldi veergudesse vea halli taustaga lahter *Data* tabeli veerupäiste kohale (enamasti sinna, kus on veerupealkiri *Total*)
- Et tabelit oleks võimalik paindlikult kujundada, tuleb selle sisust teha koopia jättes kujunduselemendid kleepimata. Selleks:
 - * märgista e selekteeri kogu tabel
 - * vali *Edit/Copy*
 - * vii aktiivne lahter uue tabeli ülemise vasaku nurga kohale ja
 - * vali *Edit/Paste Special/Values*
 - * kujunda tabel vastavaid tööriistu kasutades

4.4Väärtuste suhteline asukoht variatsioonireas

Tihti on vaja teada väärtuse asukohta variatsioonireas e teisisõnu: kui suur osa väärtusi on ühest konkreetsest väärtusest väiksemad või suuremad. Selleks:

	Järjek.nr esialgses reas		Järjek.nr. sorteeritud re	as Pi	otsentiil
• Vali Tools/DataAnalysis				<u> </u>	
Vali nimekirjast	POI	.nt	TESTI	Rank	Percent
Rank and Percentile		16	150	1	100,00%
		12	133	2	94,40%
 Dialoogiakna määrangute 		17	128	3	88,80%
valimiseks vaata selgitustega jooni	st				
leheküliel nr 0	50	14	84	15	16,60%
ichekuijei ili 9.		13	83	17	11,10%
		4	77	18	5,50%
		7	60,5	19	,00%

4.5Korrelatsioon

Korrelatsioonimaatriksi saamiseks (seosekordaja Pearsoni r):

 Paiguta andmestikus kõrvuti need tunnused, mille vahelisi korrelatsioone soovid leida, sest Excel ei aktsepteeri antud juhul eraldiasuvaid andmepiirkondi! Selleks:
 * paiguta aktiivne lahter veergu, millest vasakule soovid tühja tunnust või märgi nii mitu veergu, kui palju tühje tunnuseid andmete ümberpaigutamiseks vajad

* vali Insert/Columns

* selekteeri veeru päisel klõpsates tunnused, mida soovid ümber paigutada

* vali Edit/Cut

* paiguta aktiivne lahter eelnevalt tekitatud tühja piirkonna ülemisse vasakusse nurka ja vali *Edit/Paste*

* kustuta tühjad veerud *Edit/Delete*

- Vali Tools/DataAnalysis...
- Vali nimekirjast Correlation ning vajuta OK

	Correlation
Sisesta andmete piirkond	n input
	Input Range: \$D\$1:\$E\$20
Andmeseeriad on veergudes	Gravinad Pro
või ridades	
	<u>Help</u>
Esimeses reas/veerus on/ei ole	
pealkirjad	Output options
	C Output Range:
Sisesta väljundi niirkond	• New Worksheet <u>Ply</u>
Sisesta valjanar prirkona	C New Workbook

Korrelatsioonimaatriks

	motivatsioon	sisseastumishinne	keskmine hinne	eksam
motivatsioon	1			
sisseastumishinne	0,440	1		
keskmine hinne	0,256	0,505	1	
eksam	0,287	0,306	0,370	1

4.6Regressioon

Regressioonanalüüsi tellimiseks (meetod: linaarne regressioon):

- Paiguta andmestikus kõrvuti sõltumatud tunnused, sest Excel ei aktsepteeri antud juhul eraldiasuvaid andmepiirkondi! (vt juhiseid eelmisest punktist!)
- Vali Tools/DataAnalysis...
- Vali nimekirjast Regression ning vajuta OK



	Regression	Statistics
Mitmene seosekordaja	Multiple R	0,425
Seosekordaja ruut	R Square	0,181
Parandatud R ²	Adjusted R Square	0,176
Mudeli standardviga	Standard Error	123,294
Objectide arv	Observations	465

			ANOVA		
	Vabadusast- mete arv	Ruuthälvete summa	Keskruut e dispersioon	F	Olulisuse tõenäosus p
	df	SS	MS	F	Significance F
Regression	3	1548102,109	516034,036	33,946	0,000
Residual	461	7007900,953	15201,521		
Total	464	8556003,062			

	Kordajad	Standardviga	t statistik	Olulisuse tõenäosus p	Alumine usalduspiir	Ulemine usalduspiir
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept = konstant	106,875	13,559	7,882	0,000	80,229	133,520
motivatsioon	2,196	0,583	3,769	0,000	1,051	3,341
sisseastumishinne	4,874	2,991	1,629	0,104	-1,005	10,752
keskmine hinne	9,108	1,583	5,754	0,000	5,997	12,218

4.7t-test kahe valimi keskväärtuste võrdlemiseks.

t-testi tellimiseks kahe valimi võrdlemiseks:

- Sorteeri andmetabel grupeeriva tunnuse järgi (nii, et iga grupi tulemused oleks määratavad tervikliku piirkonnana) või paiguta gruppide tulemused eraldi veergudesse
- Vali Tools/DataAnalysis...
- Sõltumatute valimite puhul vali nimekirjast: *t-Test: Two-Sample Assuming Equal Variances* kui valimite hajuvused on sarnased või

t-Test: Two-Sample Assuming Unequal Variances kui valimite hajuvused on erinevad ning vajuta *OK*

• Sõltuvate valimite puhul vali nimekirjast: *t-Test: Paired Two Sample for Means*



t-Test: Two-Sample Assuming Equal Variances

		1 grupp	2 grupp
		Variable 1	Variable 2
Keskväärtus	Mean	3,685	4,393
Dispersioon	Variance	0,139	0,183
Objektide arv	Observations	30	19
Kogudispersioon	Pooled Variance	0,156	
H₀:alamkogumite keskväärtuste erinevus	Hypothesized Mean Difference	0	
Vabadusastmete arv	df	47	
t-väärtus	t Stat	-6,122	
Ühepoolse testi olulisuse tõenäosus p	P(T<=t) one-tail	0,000	
Ühepoolse testi kriitiline t-väärtus	t Critical one-tail	1,678	
Kahepoolse testi olulisuse tõenäosus p	P(T<=t) two-tail	0,000	
Kahepoolse testi kriitiline t-väärtus	t Critical two-tail	2,012	

PS! z-Test ja F-test For Two-Sample Variances on tellitavad analoogselt t-testiga!

4.8Põhilised statistika-funktsioonid

AVEDEV(number1;number2;...) argumendiks: arvuliste väärtuste hulk tulemuseks: hälvete absoluutväärtuste keskväärtus AVERAGE(number1;number2;...) a: arvuliste väärtuste hulk t: keskväärtus (aritmeetiline keskmine) CHITEST(actual range; expected range) a: tegelike väärtuste piirkond; oodatavate väärtuste piirkond t: ²/test CONFIDENCE(alpha;stdev,size) a: usaldusnivoo;standardhälve;valimi suurus t: usaldusintervall üldkogumi keskväärtusele CORREL(array1;array2) a: massiiv1;massiiv2 (tunnus1;tunnus2) t: korrelatsioonikoefitsent COUNT(value1;value2;...) a: väärtuste hulk t: loendab väärtuste hulgas olevad numbrid COUNTA(value1;value2;...) a: väärtuste hulk t: loendab mittetühjad lahtrid märgitud piirkonnas COUNTBLANK(range) a: piirkond t: loendab tühjad lahtrid märgitud piirkonnas COUNTIF(range; criteria) a: piirkond;tingimus t: loendab tingimusele vastavad väärtused märgitud piirkonnas DEVSQ(number1;number2;...) a: arvuliste väärtuste hulk t: ruuthälvete summa FREQUENCY(data array; bins array) a: andmete massiiv; intervallide ülemiste otspunktide massiiv t: sagedused NB! Kasutamiseks: • Selekteeri lahtrid, millesse soovid valemi (sagedused) sisestada • Sisesta valem käsitsi trükkides, mitte menüüde abil • Vajuta CTRL+SHIFT+ENTER FTEST(array1;array2)

a: massiiv1;massiiv2 (tunnus1;tunnus2)

t: f-test ühepoolse olulisustõenäosusega dispersioonide võrdlemiseks

KURT(number1;number2;...) a: arvuliste väärtuste hulk t: ekstsess

LARGE(array;k) a: massiiv1;konstant t: k-s väärtus alates suurimast

MAX(number1;number2;...) a: arvuliste väärtuste hulk t: maksimaalne väärtus

MEDIAN(number1;number2;...) t: mediaan

MIN(number1;number2;...) t: minimaalne väärtus

MODE(number1;number2;...) t: mood

PEARSON(array1;array2) a: massiiv1;massiiv2 (sõltumatu tunnus; sõltuv tunnus) t: Pearsoni korrelatsioonikordaja

PERCENTILE(array;k) a: massiiv;konstant t: k-s protsentiil

PERCENTRANK(array;x;significance)

a: massiiv;väärtus;komakohtade arv t: x-ist väiksemate arvude osakaal määratud piirkonnas kümnendmurruna (protsentide saamiseks korruta 100-ga)

PERMUT(number;number_chosen) a: objektide arv; valitavate obj. arv t: permutatsioonide arv

QUARTILE(array;quart)

a: massiiv; mitmes kvartiil (0=min, 1, 2=mediaan, 3, 4=max) t: kvartiil

RANK(number;ref;order)

a: arv;väärtuste piirkond;järjekord (0=kahanev, suv.teine=kasvav) t: antud arvu järjenumber

(võrdsetele väärtustele omistatakse järjekorras esimese positsiooni järjenumber ning järgmine järjenumber jääb kasutamata)

ROUND(number;num_digits)

a: arv; komakohtade arv

t: ümardab arvu etteantud komakohtade arvuni

ROUNDDOWN(number;num_digits)

a: arv; komakohtade arv

t: ümardab arvu esimese väiksema 0-ga lõppeva arvuni

ROUNDUP(number;num digits) a: arv; komakohtade arv t: ümardab arvu esimese suurema 0-ga lõppeva arvuni MROUND(number; multiple) a: arv; jagaja t: ümardab arvu 'jagajaga' jaguva arvuni RSQ(known y's;known x's) a: massiiv1;massiiv2 (sõltuv tunnus; sõltumatu tunnus) t: Pearsoni korrelatsioonikordaja ruut SKEW(number1;number2;...) t: asümmeetriakordaja SMALL(array;k) a: massiiv;konstant t: k-s väärtus alates vähimast SUM(number1;number2) a: arvuliste väärtuste hulk t: summa STANDARDIZE(x;mean;standard dev) a: arv;keskväärtus;standardhälve t: standardväärtus STDEV(number1;number2;...) a: arvuliste väärtuste hulk t: hinnang üldkogumi standardhälbele (N-1) STDEVP(number1;number2;...) t: valimi standardhälve TRIMMEAN(array;percent) a: massiiv; osakaal kümnendmurruna t: korrigeeritud keskväärtus (ilma ekstreemsete väärtusteta) TTEST(array1;array2;tails;type) a: massiiv1; massiiv2; mitmepoolne test; testi tüüp 1= sõltuvad tunnused, 2= sõltumatud tunnused võrdsete hälvetega, 3= sõltumatud tunnused erinevate hälvetega t: t-test (t-väärtuse olulisustõenäosus) VAR(number1;number2;...) a: arvuliste väärtuste hulk t: hinnang üldkogumi dispersioonile VARP(number1;number2;...) t: valimi dispersioon ZTEST(array;x;sigma) a: andmemassiiv; arv; üldkogumi standardhälve (kui ära jätta, siis kasutatakse valimi standardhälvet) t: X-i standardiseeritud väärtus andmetega määratud jaotuse puhul

5. Andmebaasi-protseduurid Excelis

Andmebaasiprotseduuridena mõistetakse järgnevas andmetabeli sorteerimist, andmetabelist vajalike kirjete leidmist, väljavõtete tegemist andmetabelist vastavalt etteantud tingimustele jms.

5.1Andmete sisestamine töölehel ja vastavas dialoogiaknas

Lihtsa andmebaasi (e andmetabeli) koostamisel tuleb järgida põhimõtet, et iga rida andmetabelis vastab ühele objektile ning seda rida nimetatakse **kirjeks**. Iga veerg aga vastab ühele objekte kirjeldavale **tunnusele**.

Andmetabeli esimene rida Exceli töölehel võiks sisaldada (rangelt soovitav) tunnuse- e. veerunimesid; need moodustavad tabeli päise. Enne andmete sisestamist tuleb läbi mõelda, milliseid tunnuseid uus andmetabel sisaldama hakkab ning väga soovitav on kõigepealt sisestada veerunimed ning alles seejärel hakata sisestama andmeid.

Andmete sisestamisel otse töölehele tuleb toimida vastavalt juba tuttavatele Exceli võimalustele ning siin pole veerunimede eelneval defineerimisel määravat tähtsust. Kui aga soovid andmeid sisestada vormindatud dialoogiaknas nagu seda tehakse enamuses andmebaasisüsteemides, siis peavad kõik tabeli veerud eelnevalt defineeritud olema.

Andmete sisestamiseks vormindatud dialoogiaknas:

- Vii aktiivne lahter andmetabeli piirkonda
- Vali Data/Form ...
- Sisesta andmed ning vajuta uue kirje saamiseks nuppu *New* (klahvi *Enter* vajutus viimasel väljal annab automaatselt uue kirje)



- * Kirjete lehitsemiseks dialoogiaknas kasuta nooleklahve
- * Kuvatud kirje kustutamiseks vajuta nuppu Delete (kirje kustutatakse jäädavalt!)
- * Kuvatud kirjes tehtud muudatuste tühistamiseks vajuta nuppu Restore

5.2Andmetabeli sorteerimine

- Vii aktiivne lahter andmetabeli piirkonda
- Vali Data/Sort...

	Sort ? ×	
	Stinnipäev	
	nimi eesnimi sünnipäev	
	telefon aadress1 C Descending	
Vali kas sinu	Ascending Descending	
andmetabelil on päiserida	My List Has	
1		_

- Vali väljalt *Sort By* tunnus, mille alusel tabel sorteerida
- Määra, kas sorteerida kasvavasse (Ascending) või kahanevasse (Decending) järjekorda
- Kui soovid need kirjed, mis on esimese tunnuse osas sarnased, sorteerida omavahel mingi teise tunnuse järgi, siis vali väljalt *Then By* teine tunnus jne.

5.3Kirjete leidmine.

Mugavaim viis kirjete leidmiseks mingi teadaoleva tingimuse alusel on vormindatud dialoogiakna kasutamine.

- Vii aktiivne lahter andmetabeli piirkonda
- Vali Data/Form...
- Vajuta nuppu Criteria
- Sisesta vastavale väljale tingimus, mille alusel kirjet otsida (kasuta loogikaoperaatoreid)



5.4Väljavõtete tegemine andmetabelist.

Väljavõtete tegemisel andmetabelist on sul võimalus määrata, kas soovid tingimus(t)ele vastavaid kirjeid näha andmetabeli piirkonnas (st. tingimus(t)ele mittevastavad kirjed peidetakse ajutiselt) või paigutada tingimus(t)ele vastavad kirjed mingisse uude töölehe piirkonda (esialgne tabel jääb muutmata).

5.4.1Tingimus(t)ele vastavate kirjete kuvamine andmetabeli piirkonnas

- Vii aktiivne lahter andmetabeli piirkonda
- Vali Data/Filter/AutoFilter
- Klõpsi noolega nupul vastava veeru päises ning vali nimekirjast üks olemasolevatest tunnuse väärtustest (sel juhul on loogikaoperaatoriks =) või *Custom...* keerulisema tingimuse koostamiseks.
- Koosta tingimus

	H101	•	Tallinn						
	Α	В	С	D	E	F	G	Н	I
1	nimi 🕞	eesnimi 👻	sünnipä 👻	telefon 🕞	aadress 🗸	aadress 🗸	aadress 🗸	aadress 🗸	indeks 🖵
2	Aleksandr	Aleksandra	18.01.94		Kosmonau	ıdi 7-21	Raj	olamk 🔺	
3	Alle	Maarika	29.12.84		Pärsama k	Leisi v.	Raj	olamk.	
4	Antipina	Kerli	01.03.85		Vilde 121b	-30	Ta	inn	
5	Arak	Sigrid	05.07.93		Jalaka 12-(5	Tar	tu	EE0012
6	Arhipova	Veera	17.12.88	63355-19	Punane 33	-36	Val V65	gamk. 🔄	EE0001
7	Baskakova	Zinaida	10.11.88		Narva mnt	40-24	Jőhvi Vilja	andimk. 🔽	а
8	Bokman	Helen	15.06.87		Liivalaia 42	2-51 👘		Tallinn	EE
9	Burk	Alice	28.04.90		Filtri tee 8-	6 .		Tallinn	EE0001
10	Formão	Acco	27 NG Q1		Nurmo 42.5)7	Tooblo v	Lääna mk	EE3160



Filtri mõju tühistamiseks

klõpsi noolega nupul defineeritud tingimusega veeru päises ning vali nimekirjast All või vali Data/Filter/Show All

5.4.2Alamtabeli loomine uude töölehe piirkonda

Selleks, et eraldada andmetabelist teatud tingimustele vastavaid kirjeid tuleb kõigepealt töölehele sisestada nn tingimus(t)e piirkond. Tingimuste piirkond peab sisaldama esimeses reas nende veergude nimesid, mille kohta tingimus soovitakse kehtestada. Sellele järgnevas reas (või ridades) tuleb aga loogikaoperaatorite abil koostada tingimus e. filter.

- NB! Ühes reas olevad tingimused ühendatakse operaatoriga JA ning erinevates ridades olevad tingimused ühendatakse operaatoriga VÕI
- N. Järgmine tingimus eraldab kõik need lapsed kellel on telefon <u>või</u> kes elavad Tallinnas ja on sündinud ennem 1992. aastat.

telefon	sünnipäev	aadress4
>0		
	<01.01.92	Tallinn

• Vali Data/Filter/Advanced Filter...

Advanced Filter	? ×
Action C Filter the List, in-place	OK
• Copy to Another Location	Cancel
List Range:	
<u>C</u> riteria Range:	
Сору <u>t</u> о:	
Unique <u>R</u> ecords Only	

- Vali Copy To Another Location
- Sisesta andmete piirkonna aadress väljale List Range
 - * Kui aktiivne lahter oli jäetud andmetabeli piirkonda, siis sisestatakse aadress automaatselt

* Piirkond võib olla väiksem kui andmetabel (n. kui sa ei soovi väljavõttesse kõiki tunnuseid, mis on andmetabelis), kuid ta peab sisaldama kindlasti neid veerge, mille kohta seatakse tingimus ning ta peab olema pidev (st eraldi asuvaid veerge ei tohi selekteerida)

- Sisesta tingimuse piirkonna aadress väljale *Criteria Range* * Tingimuste piirkond peab olema andmetabeliga ühel töölehel
- Sisesta uue tabeli ülemise vasaku nurga aadress väljale *Copy To:* * Uus tabel peab olema andmetabeliga ühel töölehel
- Märgi *Unique Records Only*, kui soovid, et väljavõttes kuvataks vaid üksteisest erinevad kirjed.