

Praktikum 2



Tänase praktikumi teema on andmete importimine ja haldamine, sh andmetabelite ühendamine ja uute tunnuste arvutamine.

KUIDAS AVADA/KOPEERIDA EXCELISSE TEKSTI-FAILI VÕI INTERNETIS OLEVAT ANDMETABELIT?

--- Näide 1 ---

Samm 1.

Avage internetibrauseri aknas andmetabel

http://www.eau.ee/~ktanel/Exceli_koolitus_EMYS_2014/lehmad.txt

Inv_nr	Synnikp	Isa	lehmaSPAV	EHF_veresus
E1125	15.09.2002		65303 93	0,887
E1176	22.10.2002		65303 85	0,875
E3177	23.10.2004		56172 110	1
E3195	5.11.2004		56172 119	1
E3197	5.11.2004		56172 125	1
E3198	5.11.2004		56172 115	0,875
E3202	7.11.2004		56172 95	1
E3204	7.11.2004		56172 125	1
E3222	19.11.2004		56172 112	0,875

Samm 2.

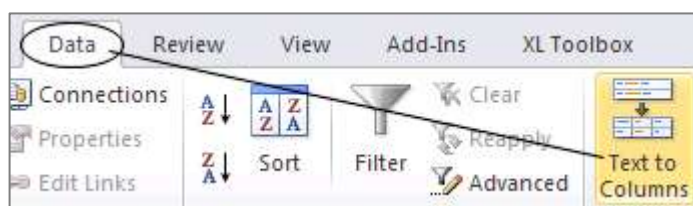
- Avage tühi Exceli fail,
- võtke blokki kogu internetibrauseri aknas avatud andmetabel (lihtsaim viis on kasutada klahvikombinatsiooni Ctrl+A),
- kopeerige see andmetabel arvuti vahemällu (näiteks klahvikombinatsiooni Ctrl+C abil),
- klõpsake Exceli töölehel lahtril A1 ning kleppige andmetabel sinna (näiteks klahvikombinatsiooni Ctrl+V abil).

Tulemusena oskab Excel jagada kopeeritud tabeli read küll eraldi ridadesse, aga jääb jänni veergude eraldamisega – kõik tabeli samas reas paiknevad väärtused on ühes lahtris:

	A	B	C	D	E	F	G
1	Inv_nr	Synnikip	Isa	lehmaSPAV	EHF_veresus		
2	E1125	15.09.2002	65303	93	0,887		
3	E1176	22.10.2002	65303	85	0,875		
4	E3177	23.10.2004	56172	110	1		
5	E3195	5.11.2004	56172	119	1		
6	E3197	5.11.2004	56172	125	1		
7	E3198	5.11.2004	56172	115	0,875		

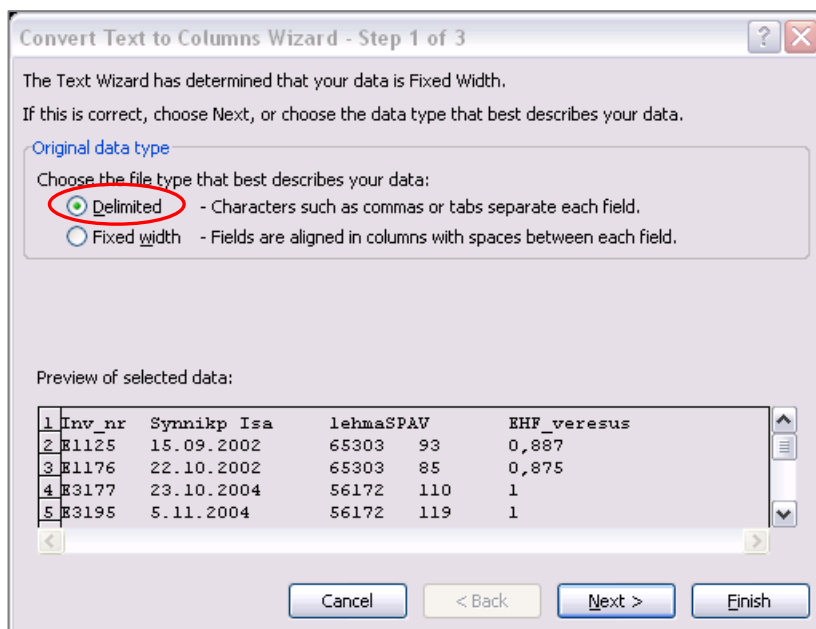
Samm 3.

Väärtuste veergudesse jagamiseks tuleb võtta blokki kõik andmetabeli read (vaikimisi nad seda ongi) ning valida *Data*-sakilts käsk *Text to Columns*:



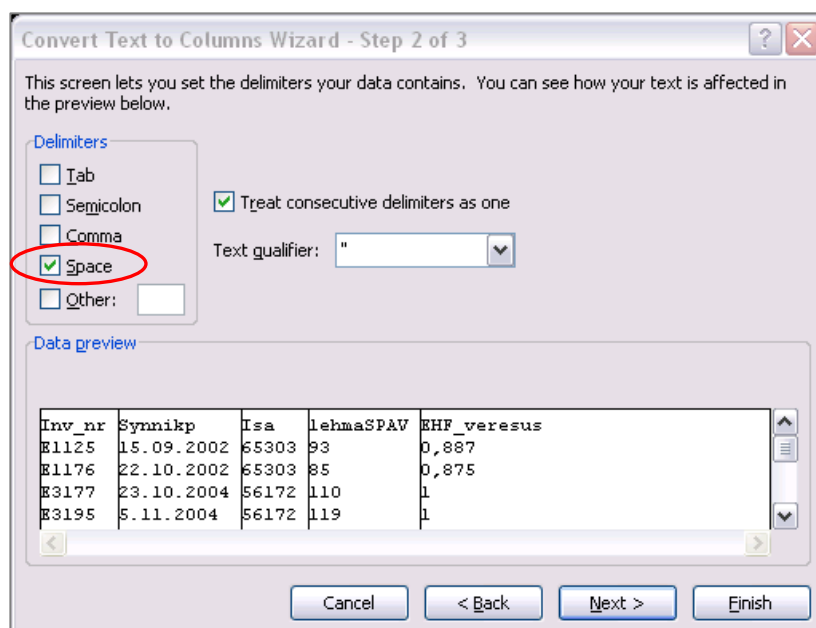
Avanenud aknas tuleb Excelile öelda, kas

- veerud imporditavas andmetabelis on kindla fikseeritud laiussega (vaikimisi variant) või
- on veerud eraldatud mingi kindla sümboliga – see variant on õige antud juhul.



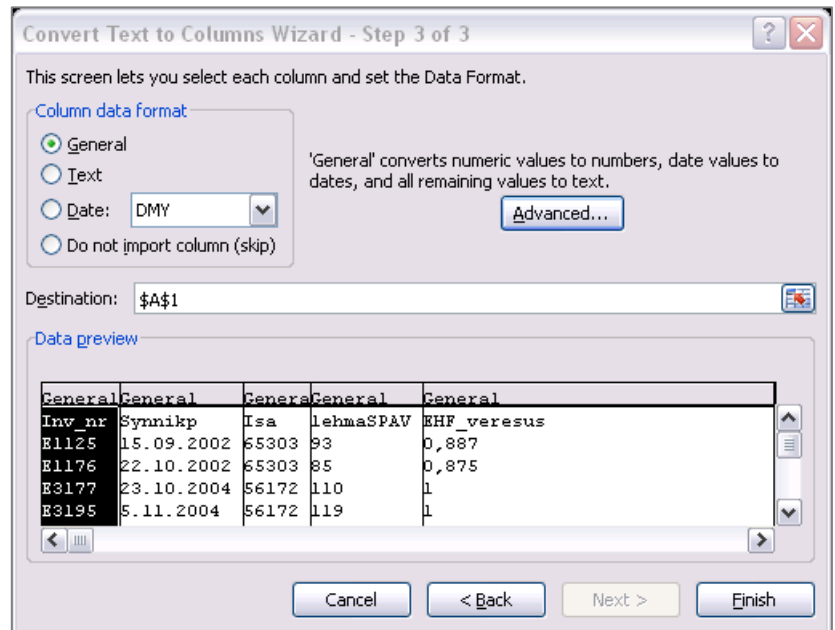
Järgmisel sammul tahab Excel teada, mis sümboliga on veerud eraldatud – antud juhul on selleks tühik (*Space*).

NB! Excel näitab vertikaaljoontega, millistelt kohtadelt ta antud veergude eraldaja puhul tabeli veergudeks jagab (**proovige näiteks**, kuidas näeks imporditav tabel välja siis, kui määrata veergude eraldajaks punkt – valige *Other* ja trükkige selle järel lahtrisse punkt – või siis, kui eraldada veerud nii punktide kui ka tühikute kohalt).



Tabeli veergudesse jagamise viimasel sammul saate soovi korral

- määrata valitud veergudele teksti või kuupäeva formaadi (see viimane enamasti kuupäevade õigesti importimist lihtsamaks ei tee – kui Excel saab aru, et mingis veerus on tegu kuupäevadega, saab ta seda nagunii, kui aga ei saa, ei aita ka mingi valik antud aknas),
- jätta mõne veeru importimata (selekteeritud veerule tuleb omistada formaat *Do not import column (skip)*),
- täpsustada, milline on imporditavas tabelis arvude kümnendkohtade ja tuhandete eraldaja (nupu *Advanced...* all avanenavas aknas – see koht on **ülioluline, kui imporditavas tabelis on kümnendkohtade eraldajaks punkt** eesti keelele omase koma asemel!),
- lasta imporditav tabel kirjutada kuhugi mujale – lahtrisse *Destination* tuleb ette anda tabeli soovitava asukoha ülemise vasaku nurga aadress (näiteks H1; selle valiku tulemusena tekib töölehele kaks tabelit – üks, mis on ette antud reeglite kohaselt veergudeks jagatud, ja algne, veergudeks jagamata tabel).



Antud juhul ei ole neist võimalustest vaja kasutada ühtegi, võib koheselt klõpsata nupul 'Finish' ja tulemuseks ongi korrektselt veergudesse jagatud andmetabel:

	A	B	C	D	E
1	Inv_nr	Synnikp	Isa	lehmaSPA	EHF_veresus
2	E1125	15.09.2002	65303	93	0,887
3	E1176	22.10.2002	65303	85	0,875
4	E3177	23.10.2004	56172	110	1
5	E3195	5.11.2004	56172	119	1
6	E3197	5.11.2004	56172	125	1
7	E3198	5.11.2004	56172	115	0,875
8	E3202	7.11.2004	56172	95	1
9	E3204	7.11.2004	56172	125	1
10	E3222	19.11.2004	56172	112	0,875

Või kas ikka on kõik korrektne? Kiirem võimalus kontrollida arvutunnuste korrektsust on klikkida vastaval veerul ja vaadata järgi maksimaalsed ja minimaalsed väärtused ekraani all paiknevalt olekuribalt:

	A	B	C	D	E	F	G	H	I
1	Inv_nr	Synnikp	Isa	lehmaSPA	EHF_veresus				
2	E1125	15.09.2002	65303	93	0,887				
3	E1176	22.10.2002	65303	85	0,875				
4	E3177	23.10.2004	56172	110	1				
5	E3195	5.11.2004	56172	119	1				
6	E3197	5.11.2004	56172	125	1				

Average: 87,14041849 Count: 412 Numerical Count: 411 **Min: 0,86** Max: 135 Sum: 35814,712

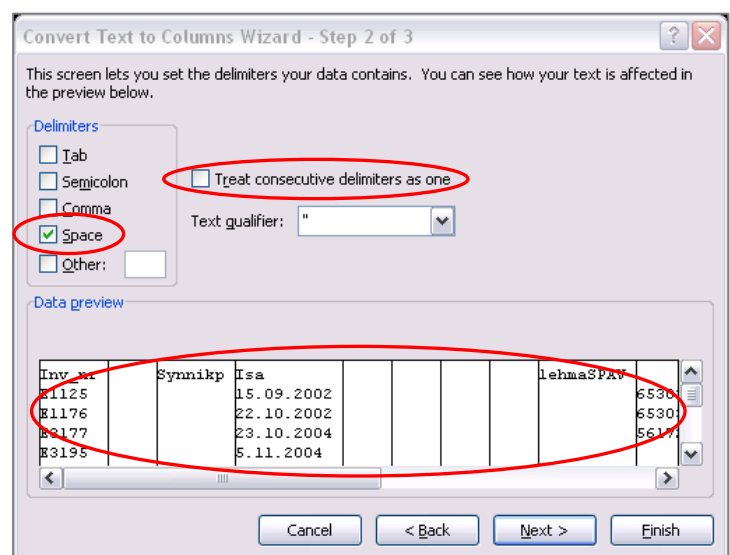
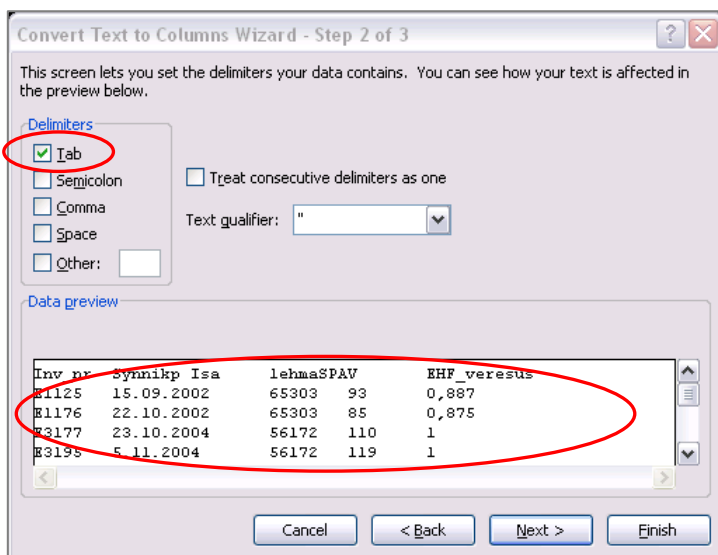
SPAV on lehma suhteline piimajõudluse aretusväärtus, mille keskmine väärtus on 100 punkti ning mille miinimum ja maksimum jäävad tavaliselt 70 ja 130 kanti. SPAV-i väärtus 0,86, nagu on kirjas tabeli all minimaalse väärtusena, on ilmselgelt ebareaalne. Milles probleem? Kerides andmetabelit allapoole selgub, et osadel lehmadel on justkui puudu nende holsteini veresuse andmed, samas on neil SPAV-i väärtus ebareaalselt väike:

Ilmselt on neil lehmadel tegelikult puudu hoopis SPAV, aga andmete importimisel on Excel jätnud tühja veeru vahelt ära.

Mida taolisel juhul teha?

Üks variant on püüda andmeid uuesti importida, püüdes kasutada mõnda muud veergude eraldajat (näiteks tühiku asemel TAB-i) ja vaadates, kuidas muutub imporditav tabel, kui keelata järjestikuste eraldajate ühena käsitlemine. Võite avada uue Exceli dokumendi (lihtsaim viis selleks on klahvikombinatsioon Ctrl+N), kopeerida sinna internetist uuesti lehmade andmetabeli ja püüda seda jagada veergudeks. Tulemus ei ole positiivne – valik TAB ei suuda veerge eraldada ja kasutades veergude eraldajana tühikut, annab mitme järjestikuse tühiku ühe sümbolina käsitlemise keelamine veel segasema tulemuse:

	A	B	C	D	E
1	Inv_nr	Synnikip	Isa	lehmaSPAV	EHF_veresu
328	E5109	19.06.2008	56274	108	1
329	E5110	20.06.2008	56274	0,907	
330	E5120	25.06.2008	65210	106	1
331	E5126	28.06.2008	56274	115	1
332	E5134	4.07.2008	65699	1	
333	E5142	6.07.2008	65699	0,938	



Mis edasi?

Üks variant oleks uurida, kas ehk on mingi võimalus saada Excelisse imporditav tabel kätte mingis muus formaadis, näiteks mingi muu veergude eraldajaga (levinuil juhul kas semikooloni või komaga – nn csv- ehk *Comma delimited*-failina), sest siis on puuduva väärtuse kohal kaks veergude eraldajat ning Excel suudab selle alusel veerud lahku lüüa.

Teine variant oleks püüda olla nutikas ja moodustada mingi algoritmi alusel juba õigeid väärtuseid sisaldavad veerud. Ühte ja alati kehtivat lahendust siinkohal aga pole. Antud juhul saab näiteks kasutada asjaolu, et holsteini tõu veresuse väärtused peavad olema 0 ja 1 vahel, samas kui SPAV ei saa kunagi omandada nii väikeseid väärtuseid.

See tähendab, et kasutades mõningaid IF-funktsioone, on võimalik tekitada valesti veergudesse jagatud andmetabeli kõrvale juba õigeid väärtuseid sisaldavad veerud ning seejärel kirjutada nendega üle algused veerud (vt järgmine lk).

	A	B	C	D	E	F	G	H	I
1	Inv_nr	Synnikip	Isa	lehmaSPAV	EHF_veresus		Õige SPAV	Õige EHF-veresus	
2	E1125	15.09.2002	65303	93	0,887		=IF(D2>1;D2;"")	=IF(E2>0;E2;IF(D2<=1;D2;""))	

Nõ õige SPAV-i tarvis vaatab ülalolev funktsioon, kas SPAV-i veerus olev arv on suurem kui üks, kui on, väljastatakse algse SPAV-i veerus olev väärtus, vastasel juhul jäetakse aga lahter tühjaks (väärtuseks on "").

Nipid, märkused, soovitused.

NB! Jättes IF-funktsioonis viimase argumendi lihtalt ette andmata (mõeldes, et kui SPAV on väiksem kui üks, siis ei kirjuta sinna lahtrisse midagi), kirjutab Excel lahtrisse tegelikult nulli, mitte ei jäta lahtrit tühjaks!

	D	E	F	G
1	lehmaSPAV	EHF_veresus		Õige SPAV
353	114	1		114
354	0,907			0

Õige EHF-veresuse lahteris olev valem

=IF(E2>0;E2;IF(D2<=1;D2;""))

- o vaatab esmalt, kas tegelik EHF-veresus on suurem kui null (st kas lahter ei ole tühi),
- o kui vastus on jah, võetakse väärtuseks algse veresuse lahtris olev arv,
- o kui aga vastus on ei, kontrollitakse edasi, kas ehk algse SPAV-i veerus olev väärtus on väiksem-võrdne ühest,
- o kui on, on veresus ilmselt kirjas algse SPAV-i veerus ja tulebki võtta sealt,
- o vastasel juhul loetakse veresus mitte teadaolevaks ja vastav lahter jäetakse tühjaks.

Nüüd on nii SPAV-i kui ka veresuse andmed juba õigetes veergudes

	D	E	F	G	H
1	lehmaSPAV	EHF_veresus		Õige SPAV	Õige EHF-veresus
327	1				1
328	108	1		108	1
329	0,907				0,907
330	106	1		106	1

ja nende väärtustega tuleb üle kirjutada algsed imporditud väärtused:

Inv_nr	Synnikip	Isa	lehmaSPAV	EHF_veresus	Õige SPAV	Õige EHF-veresus
E1125	15.09.2002	65303	93	0,887	93	0,887
E1176	22.10.2002	65303	85	0,875	85	0,875
E3177	23.10.2004	56172	110	1	110	1
E3195	5.11.2004	56172	119	1	119	1

Kleebi teisiti / Paste special -> Väärtused / Values

Kopeeri (Ctrl+C)

Ja viimase sammuna tuleks need abiread, kuhu sai arvutatud õiged SPAV-d ja veresused, ära kustutada.

--- Näide 2 ---

Teine variant teksti-faili Excelisse importida on kasutada lihtsalt *Open*-käsku.

- Salvestage fail

http://www.eau.ee/~ktanel/Exceli_koolitus_EMYs_2014/klypsid.txt

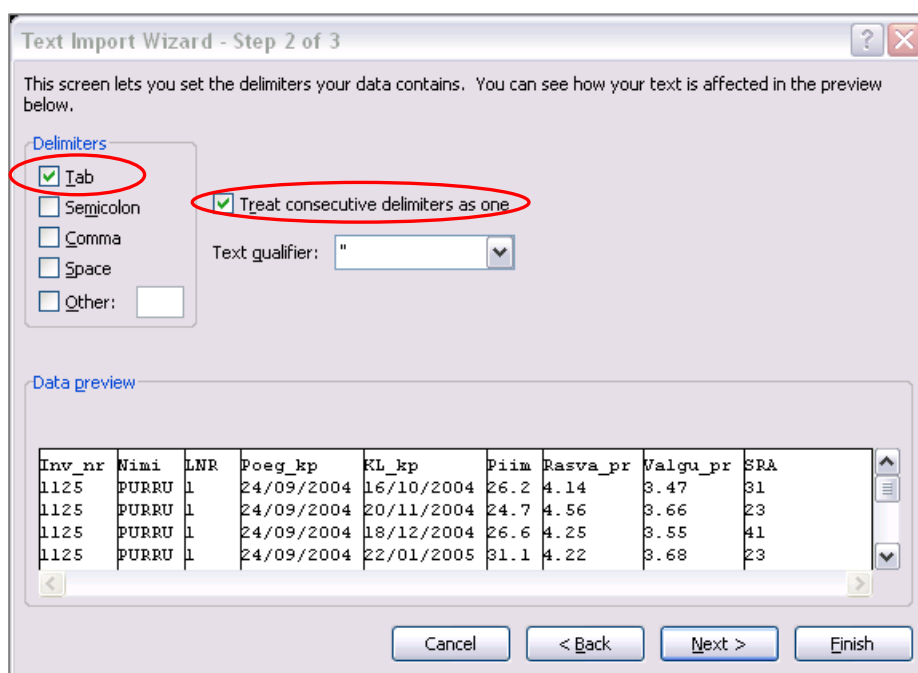
(klikkides näiteks faili nimel kursuse kodulehel hiire parempoolse nupuga ning valides avanenud rippmenüüst käsu 'Save File As...' või 'Salvesta link kui...' või ...

- Valige Excelis käsk *Open*, määrake avatava faili tüübiks teksti-fail

Files of type:

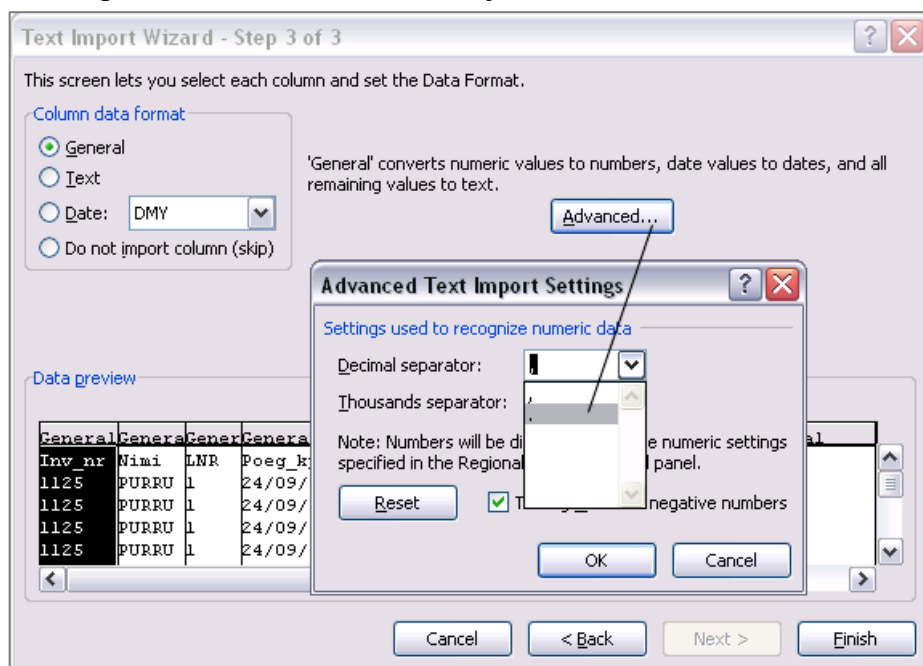
ning avage salvestatud kontrollüpside andmestik 'klypsid.txt'.

- Avaneb juba nähtud teksti-faili importimise / veergudesse jagamise aken, kus
 - esimesel sammul tuleb öelda, et veerud on kindla eraldajaga (mitte fikseeritud laiusega);
 - teisel sammul tuleb määrata veergude eraldaja – NB! erinevalt internetilehelt kopeeritud tabelist on antud juhul veergude eraldajaks tabulaator (*Tab*) ning vältimaks asjata tühjade veergude tegemist tuleks anda Excelile käsk käsitleda mitut järjestikust veergude eraldajat ühena (*Treat consecutive delimiters as one*);



- kolmandal sammul tuleb nupul *Advanced...* klikkimise järel avanenud aknas öelda

Excelile, et imporditavas teksti-failis on kümnendkohtade eraldajana kasutusel punkt – sellisel juhul tõlgendabki Excel punkti arvus komana ning eesti keele seadistuse puhul konverteeritakse arvud automaatselt komadega arvudeks (ja mitte kuupäevadeks nagu vastasel juhul võib juhtuda).



Tulemus:

	A	B	C	D	E	F	G	H	I
1	Inv_nr	Nimi	LNR	Poeg_kp	KL_kp	Piim	Rasva_pr	Valgu_pr	SRA
2	1125	PURRU	1	24.09.2004	16.10.2004	26,2	4,14	3,47	31
3	1125	PURRU	1	24.09.2004	20.11.2004	24,7	4,56	3,66	23
4	1125	PURRU	1	24.09.2004	18.12.2004	26,6	4,25	3,55	41
5	1125	PURRU	1	24.09.2004	22.01.2005	31,1	4,22	3,68	23
6	1125	PURRU	1	24.09.2004	21.02.2005	30,6	4,23	3,75	28
7	1125	PURRU	1	24.09.2004	21.03.2005	26,7	4,54	3,77	39
8	1125	PURRU	1	24.09.2004	22.04.2005	25,6	4,57	3,6	46

Nipid, märkused, soovitused.

Peamine probleem seoses erinevatest andmebaasidest või internetist pärit andmetabelite Excelis avamisega on see, et erinevat tüüpi tunnused on vormindatud erinevalt.

Eestis on arvutid enamasti eesti keele seadistuses, mis tähendab arvudes kümnendkoha eraldajana koma kasutamist ning vaikimisi kuupäevaformaati Excelis kujul 'päev.kuu.aasta' (näiteks 3.05.2013). Andmebaasides on aga reeglina inglise keele seadistused, mille puhul kümnendkoha eraldajana arvudes kasutatakse punkti ja kuupäeva esituses on kuu number enne päeva numbrit või siis on päeva, kuu ja aasta eraldajana kasutatud hoopis kaldkriipsu, sageli on kuude nimed esitatud ka üldse sõnaliselt (näiteks May 3, 2013, seda eriti Ameerika inglise keele puhul).

Variandid, kuidas andmeid Excelisse importida (mis alati ei pruugi töötada – seepärast ongi neid mitu).

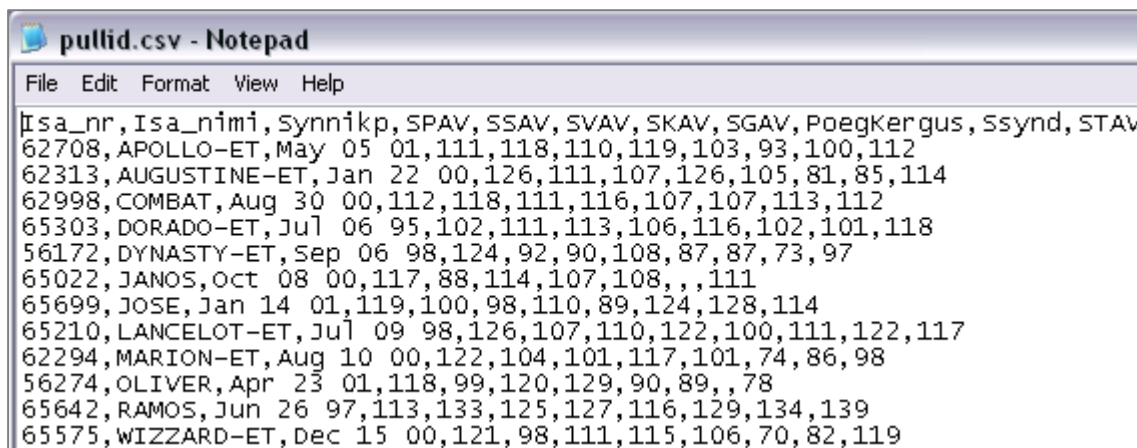
- Kasutage eelnevalt kirjeldatud tekstifaili importimise või lahtrite sisu veergudesse jagamise võimalust.
- Muudke enne inglise keeleruumi reeglite järgi vormindatud andmete Excelisse kopeerimist või tekstifaili avamist arvuti seadistust kontrollpaneelilt: *Control Panel* → *Regional and Language Options* → *English (United Kingdom)*;
juhul, kui imporditavas failis on kuupäevades kuude nimed antud sõnaliselt, võib proovida ka USA inglise keele (*English (United States)*) seadistust.
- Kopeerige imporditav tabel Notepadi, asendage seal kõik punktid komadega ja kuupäevade puhul kõik kaldkriipsud punktidega ning kopeerige sedasi muudetud tabel Excelisse.

--- Näide 3 ---

Importige Excelisse pullide andmestik:

http://www.eau.ee/~ktanel/Exceli_koolitus_EMYs_2014/pullid.csv

Salvestage see andmestik arvutisse ja avage enne Excelis avamist igaks juhuks Notepadis (see on tekstifailide puhul soovitatav, et saada esmane ülevaade andmete struktuurist ja vormingust):



```
pullid.csv - Notepad
File Edit Format View Help
Isa_nr,Isa_nimi,Synnikp,SPAV,SSAV,SVAV,SKAV,SGAV,PoegKergus,Ssynd,STAV
62708,APOLLO-ET,May 05 01,111,118,110,119,103,93,100,112
62313,AUGUSTINE-ET,Jan 22 00,126,111,107,126,105,81,85,114
62998,COMBAT,Aug 30 00,112,118,111,116,107,107,113,112
65303,DORADO-ET,Jul 06 95,102,111,113,106,116,102,101,118
56172,DYNASTY-ET,Sep 06 98,124,92,90,108,87,87,73,97
65022,JANOS,Oct 08 00,117,88,114,107,108,,111
65699,JOSE,Jan 14 01,119,100,98,110,89,124,128,114
65210,LANCELOT-ET,Jul 09 98,126,107,110,122,100,111,122,117
62294,MARION-ET,Aug 10 00,122,104,101,117,101,74,86,98
56274,OLIVER,Apr 23 01,118,99,120,129,90,89,,78
65642,RAMOS,Jun 26 97,113,133,125,127,116,129,134,139
65575,WIZZARD-ET,Dec 15 00,121,98,111,115,106,70,82,119
```

Nagu näha, on selles failis veergude eraldajaks hoopis koma, kuupäev on formaadis 'päev kuu aasta', kusjuures inglise keeles ja aastaarv on vaid kahenumbriline (näiteks 'May 05 01'), ning kahel pullil on ilmselt mõni aretusväärtus puudu (selle tõenduseks on järjestikuste komade esinemine).

Proovides seda tabelit Excelis avada ning seejärel teksti veergudesse jagada, määrates veergude eraldajaks koma, on tulemus järgmine:

	A	B	C	D	E	F	G	H	I	J	K
1	Isa_nr	Isa_nimi	Synnikp	SPAV	SSAV	SVAV	SKAV	SGAV	PoegK	Ssynd	STAV
2	62708	APOLLO-ET	May 05 01	111	118	110	119	103	93	100	112
3	62313	AUGUSTINE-ET	Jan 22 00	126	111	107	126	105	81	85	114
4	62998	COMBAT	Aug 30 00	112	118	111	116	107	107	113	112
5	65303	DORADO-ET	Jul 06 95	102	111	113	106	116	102	101	118
6	56172	DYNASTY-ET	Sep 06 98	124	92	90	108	87	87	73	97
7	65022	JANOS	Oct 08 00	117	88	114	107	108			111
8	65699	JOSE	Jan 14 01	119	100	98	110	89	124	128	114
9	65210	LANCELOT-ET	Jul 09 98	126	107	110	122	100	111	122	117
10	62294	MARION-ET	Aug 10 00	122	104	101	117	101	74	86	98
11	56274	OLIVER	Apr 23 01	118	99	120	129	90	89		78
12	65642	RAMOS	Jun 26 97	113	133	125	127	116	129	134	139
13	65575	WIZZARD-ET	Dec 15 00	121	98	111	115	106	70	82	119

Excel saab teksti veergudesse jagamisega kenasti hakkama, ka puuduvatele väärtustele vastavad tühjad lahtrid on õiges kohas, aga pullide sünnikuupäevade kuupäevana käsitlemises jääb Excel jänni (tõenduseks, et Excel ei suuda antud formaadis olevaid sünnikuupäevi kuupäevadeks konverteerida, on nende joondumine lahtri vasaku, mitte parema serva järgi – vaikumisi joondab Excel teksti alati vasaku serva ja arvud parema serva suhtes ning mistahes ajaväärtused on oma sisult arvud).

Paremaks ei tee pilti ka arvuti kontrollpaneelilt seadistuste inglise keeleruumile vastavateks muutmine – kasutatud kuupäeva formaat on lihtsalt sedavõrd ebatraditsiooniline.

Lahenduseks on

- kuupäeva jagamine eraldi kuuks, päevaks ja aastaks,
- kuu inglisekeelsete lühendite asendamine kuu numbritega,
- kahekohalise aastaarvu asendamine neljakohalise aastaarvuga ning
- seejärel arvulistest komponentidest jällegi kuupäeva kokkupanek.

Kuupäeva veeru kolmeks eri veeruks jagamiseks saab jällegi kasutada juba tuttavat käsku *Text to Columns*, **seejuures tuleb kuupäeva sisaldava veeru järele eelnevalt lisada kaks tühja veergu**, sest vastasel juhul kirjutatakse päeva ja aasta numbritega üle järgnevate veergude sisu.

Kuu inglisekeelsete lühendite numbritega asendamiseks võib kasutada käsu *Replace* abi – see ei ole küll kõige elegantsem tegevus, aga et kuid on üksnes 12 tk, on kaksteistkordne *Replace*-käsu rakendamine ehk kõige kiirem (kui andmeid on nii vähe, kui antud näites, on kiirem viis muidugi lihtsalt kuu nimede asemele numbrite trükkimine ☺).

Tulemus:

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	Isa_nr	Isa_nimi	Synnikip			SPAV	SSAV	SVAV	SKAV	SGAV	PoegK	Ssynd	STAV	
2	62708	APOLLO-ET		5	5	1	111	118	110	119	103	93	100	112
3	62313	AUGUSTINE-ET		1	22	0	126	111	107	126	105	81	85	114
4	62998	COMBAT		8	30	0	112	118	111	116	107	107	113	112
5	65303	DORADO-ET		7	6	95	102	111	113	106	116	102	101	118
6	56172	DYNASTY-ET		9	6	98	124	92	90	108	87	87	73	97
7	65022	JANOS		10	8	0	117	88	114	107	108			111
8	65699	JOSE		1	14	1	119	100	98	110	89	124	128	114
9	65210	LANCELOT-ET		7	9	98	126	107	110	122	100	111	122	117
10	62294	MARION-ET		8	10	0	122	104	101	117	101	74	86	98
11	56274	OLIVER		4	23	1	118	99	120	129	90	89		78
12	65642	RAMOS		6	26	97	113	133	125	127	116	129	134	139
13	65575	WIZZARD-ET		12	15	0	121	98	111	115	106	70	82	119

Järgmise sammuna on vaja neljakohalisi aastarve, sest vastasel juhul tõlgendab Excel kõiki aastanumbreid 19-ga algavatena. Kuna antud juhul on selge, et

- kõik arvud, mis on väiksemad kui käesoleva aasta kaks viimast numbrit (13), peavad vastama 21. sajandile (ning on esitatavad kujul 2000 + 'kaks viimast numbrit'), ja
- ülejäänud 20. sajandile (meil lihtsalt ei saa olla praegu kasutusel pulle sünniaastaga 1912 ja enne seda; aastarv on siis esitatav kujul 1900 + 'kaks viimast numbrit'),

saab sünniaasta teha neljakohaliseks lihtsalt IF-funktsiooni ja sobiva liitmistehte abil:

Tulemusena on nüüd andmetabelis üheselt mõistetavad arvilised veerud nii kuupäeva, kuu kui ka aasta tarvis:

C	D	E	F
Kuu	Päev		Aasta
5	5	1	2001
1	22	0	2000
8	30	0	2000
7	6	95	1995
9	6	98	1998
10	8	0	2000
1	14	1	2001
7	9	98	1998
8	10	0	2000
4	23	1	2001
6	26	97	1997
12	15	0	2000

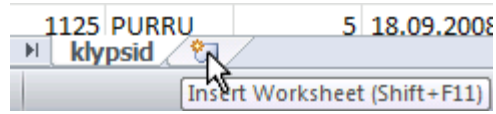
C	D	E	F
Sünnikp			
5	5	1	=IF(E2<14;2000+E2;1900+E2)
1	22	0	2000
8	30	0	2000
7	6	95	1995
9	6	98	1998
10	8	0	2000
1	14	1	2001
7	9	98	1998
8	10	0	2000
4	23	1	2001
6	26	97	1997
12	15	0	2000

Funktsioon DATE() võimaldab ühendada aasta, kuu ja päeva numbrit kuupäevaks:

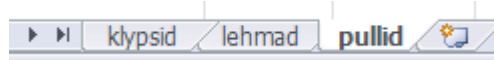
A	B	C	D	E	F	G
Isa_nr	Isa_nimi	Kuu	Päev		Aasta	Sünnikp
62708	APOLLO-ET	5	5	1	2001	=DATE(F2;C2;D2)
62313	AUGUSTINE-ET	1	22	0	2000	DATE(year; month; day)
62998	COMBAT	8	30	0	2000	30.08.2000
65303	DORADO-ET	7	6	95	1995	6.07.1995
56172	DYNASTY-ET	9	6	98	1998	6.09.1998
65022	JANOS	10	8	0	2000	8.10.2000
65699	JOSE	1	14	1	2001	14.01.2001
65210	LANCELOT-ET	7	9	98	1998	9.07.1998
62294	MARION-ET	8	10	0	2000	10.08.2000
56274	OLIVER	4	23	1	2001	23.04.2001
65642	RAMOS	6	26	97	1997	26.06.1997
65575	WIZZARD-ET	12	15	0	2000	15.12.2000

--- Teksti-failide importimise peatüki lõpetuseks ---

Lisage kontrollüpside faili kaks uut töölehte



ning kopeerige neile vastavalt lehmade ja pullide andmetabelid (nimetage vastavad töölehed ka seal sisalduvate andmete järgi):



NB! Pullide sünniaja kohta jätke alles vaid sünniaasta ja kuupäev.

Salvestage loodud kolme töölehega tööraamat Exceli failina (näiteks nimega 'klypsid.xlsx').

ANDMETABELITE ÜHENDAMINE

Lisage lehmade tabelisse iga lehma järele tema isa nimi, isa sünniaasta ning kõik pullide tabelis olevad suhtelised aretusväärtused ja poegimiskerguse ja surnultsündide aretusväärtused:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Inv_nr	Synnikp	Isa	lehmaSPA	EHF_veres	Isa_nimi	Saasta	SPAV	SSAV	SVAV	SKAV	SGAV	STAV
2	E1125	15.09.2002	65303	93	0,887								
3	E1176	22.10.2002	65303	85	0,875								

ja seejärel lisage kontroll-lüpside lehel igasse ritta lehma sünnikuupäev, lehma SPAV, EHF-veresus, isa nimi ning isa suhteline piimajõudluse aretusväärtus SPAV ja suhteline üldaretusväärtus SKAV.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Inv_nr	Nimi	LNR	Poeg_kp	KL_kp	Piim	Rasva_pr	Valgu_pr	SRA	Synnikp	lehmaSPA	EHF_veres	Isa_nimi	SPAV	SKAV
2	1125	PURRU	1	24.09.2004	16.10.2004	26,2	4,14	3,47	31						
3	1125	PURRU	1	24.09.2004	20.11.2004	24,7	4,56	3,66	23						

Selle ülesande lahendamiseks tuleb

- ühendada lehmade tabel pullide tabeliga isa numbril alusel ning seejärel
- kontroll-lüpside tabel lehmade tabeliga lehma numbril alusel.

Kahe andmetabeli ühendamine mingi ühiseid väärtuseid sisaldava veeru alusel on Excelis teostatav funktsioonide LOOKUP, VLOOKUP või INDEX ja MATCH abil.

Vaatame nende funktsioonide tutvustamiseks kahte näitetabelit, kus

- esimeses on kirjas tudengite eesnimed ja erinevate eksamite hinded,
- teises tudengite eesnimed, kursus ja õppejõudude subjektiivne hinnang:

	A	B	C	D	E	F	G	H	I	J
1	Nimi	Hinne					Nimi	Kursus	Hinnang	
2	Mari	A					Jüri	I	Loru	
3	Mari	B					Mari	I	Usin	
4	Jüri	C					Mati	II	Rumal, aga edasipüüdlik	
5	Mati	B					Siim	III	Oivik	
6	Mati	C								
7	Mati	A								
8	Siim	A								
9	Siim	A								

Järgnevate näidete läbi tegemiseks võite selle tabeli alla laadida siit:

http://www.eau.ee/~ktanel/Exceli_koolitus_EMYs_2014/tabelite_yhendamine.xlsx

Eesmärgiks on saada esimesse tabelisse igasse ritta tudengi kursus ja õppejõudude subjektiivne hinnang. Selleks tuleb otsida vastava tudengi nime teisest tabelist ning võtta sealt välja sellele nimele vastav kursus ja hinnang.

Järgnevad näited on tehtud vaid kursuse numbril lisamiseks, teie võite proovida lisada ka õppejõudude hinnangut.

--- Funktsioon LOOKUP ---

Funktsioon LOOKUP leiab otsitava väärtuse ühest veerust ja tagastab vastava (samas reas paikneva) väärtuse teisest veerust. Argumentidena tuleb määrata

- otsitav väärtus,
- veerg, kust otsida, ja
- veerg, kust võtta funktsiooni poolt tagastatav väärtus:

	A	B	C	D	E	F	G	H	I	J
1	Nimi	Hinne	Kursus				Nimi	Kursus	Hinnang	
2	Mari	A	=LOOKUP(A2;G2:G5;H2:H5)				Jüri	I	Loru	
3	Mari	B					Mari	I	Usin	
4	Jüri	C					Mati	II	Rumal, aga edasipüüdlik	
5	Mati	B					Siim	III	Oivik	
6	Mati	C								
7	Mati	A								
8	Siim	A								
9	Siim	A								

Selleks, et funktsiooni saaks kopeerida ilma valemil muutmata ka järgmistesse ridadesse, tuleb ära fikseerida otsingu veerg (nimede veerg teises tabelis kujul \$G\$2:\$G\$5) ning tagastatavate väärtuste veeru reanumbrid (H\$2:H\$5) – jättes veeru fikseerimata, saab sama valemil kopeerida ka paremale poole, misjärel lisatakse esimesele tabelile automaatselt ka järgmine veerg teisest tabelist (sellisel juhul tuleb muidugi ära fikseerida ka otsitava väärtuse veerg (a'la \$A2):

	A	B	C	D	E	F	G	H	I	J
1	Nimi	Hinne	Kursus				Nimi	Kursus	Hinnang	
2	Mari	A	=LOOKUP(\$A2;\$G\$2:\$G\$5;H\$2:H\$5)				Jüri	I	Loru	
3	Mari	B	I				Mari	I	Usin	
4	Jüri	C	I				Mati	II	Rumal, aga edasipüüdlik	
5	Mati	B	II				Siim	III	Oivik	
6	Mati	C	II							
7	Mati	A	II							
8	Siim	A	III							
9	Siim	A	III							

Nipid, märkused, soovitused.

Funktsiooni LOOKUP eeldus on, et otsingu aluseks olev tabel on sorteeritud kasvavalt otsingu aluseks oleva veeru järgi (antud näite puhul tähendab see, et teine, tudengite kursuseid ja õppejõudude hinnanguid sisaldav tabel peab olema sorteeritud tudengite nimede järgi).

Kui see nii ei ole, võib tulemus olla vale. Näiteks sorteerides antud näites teise tabeli tudengi nime asemel hoopis hinnangute veeru alusel, on tabelite ühendamise tulemus funktsiooniga LOOKUP vale, Matile on ekslikult omistatud I kursus:

Nimi	Hinne	Kursus				Nimi	Kursus	Hinnang	
Mari	A	I				Jüri	I	Loru	
Mari	B	I				Siim	III	Oivik	
Jüri	C	I				Mati	II	Rumal, aga edasipüüdlik	
Mati	B	I				Mari	I	Usin	
Mati	C	I							
Mati	A	I							
Siim	A	III							
Siim	A	III							

--- Funktsioon VLOOKUP ---

Funktsioon VLOOKUP leiab otsitava väärtuse määratud piirkonna **esimesest veerust(!)** ning tagastab vastava (samal reas paikneva) väärtuse teisest, **järjekorranumbriga määratud** veerust. Argumentidena tuleb ette anda

- otsitav väärtus,
- piirkond, mille esimesest veerust huvipakkuvat väärtust otsida ning mille mõnest edasisest veerust tuleb võtta funktsiooni poolt tagastatav väärtus,
- veeru number määratud piirkonnas, kust võtta funktsiooni poolt tagastatav väärtus,
- mittekohustusliku argumentina otsimisviisi tõeväärtuse kujul
 - kui TRUE (so vaikumisi väärtus), otsitakse täpset või suuruselt lähimat väärtust altpoolt (**see valik eeldab sarnaselt funktsioonile LOOKUP, et tabel on sorteeritud kasvavalt, vastasel juhul võib tulemus olla vale**),
 - kui FALSE, otsitakse vaid täpset väärtust ning selle mitteleidmisel väljastatakse veateade #N/A.

	A	B	C	D	E	F	G	H	I	J
11										
12	Nimi	Hinne	Kursus				Nimi	Kursus	Hinnang	
13	Mari	A	=VLOOKUP(\$A13;\$G\$13:\$I\$16;2;FALSE)				Jüri	I	Loru	
14	Mari	B	I				Mari	I	Usin	
15	Jüri	C	I				Mati	II	Rumal, aga edasipüüdlik	
16	Mati	B	II				Siim	III	Oivik	
17	Mati	C	II							
18	Mati	A	II							
19	Siim	A	III							
20	Siim	A	III							

Funktsiooniga VLOOKUP analoogselt töötab funktsioon **HLOOKUP**, ainult veergude asemel mööda ridu.

Nipid, märkused, soovitused.

Kui te ei soovi saada otsitava väärtuse mitteleidmisel tulemuseks veateadet vaid lihtsalt tühja lahtrit, saab kasutada Excel 2010 uut funktsiooni IFERROR – sellel funktsioonil on kaks argumenti:

- esimene määrab ära soovitava tegevuse ja
- teine selle, mida teha siis, kui esimese tegevuse tulemuseks oli veateade:

	A	B	C	D	E	F	G	H	I	J
11										
12	Nimi	Hinne	Kursus				Nimi	Kursus	Hinnang	
13	Mari	A	I				Jüri	I	Loru	
14	Mari	B	I				Mari	I	Usin	
15	Jüri	C	I				Mati	II	Rumal, aga edasipüüdlik	
16	Mati	B	II				Siim	III	Oivik	
17	Mati	C	II							
18	Mati	A	II							
19	Siim	A	III							
20	Siim	A	III							
21	Anu		#N/A	=VLOOKUP(\$A21;\$G\$13:\$I\$16;2;FALSE)						
22	Anu			=IFERROR(VLOOKUP(\$A22;\$G\$13:\$I\$16;2;FALSE);"")						

--- Funktsioonid MATCH ja INDEX ---

Funktsioon MATCH tagastab otsitava väärtuse järjenumbriga antud piirkonnas (**ükskõik kas reas või veerus!**). Argumentidena tuleb määrata

- otsitav väärtus,
- piirkond (rida või veerg), kust otsida, ja
- mittekohustusliku argumentina otsimisviisi võimalike väärtustega
 - 1 (so vaikumisi väärtus), mille puhul otsitakse täpset või suuruselt lähimat väärtust altpoolt (see valik eeldab, et **tabel on sorteeritud kasvavalt**, vastasel juhul võib tulemus olla vale),
 - 0, mille puhul otsitakse vaid täpset väärtust ning selle mitteleidmisel väljastatakse veateade #N/A,
 - 1, mille puhul otsitakse täpset või suuruselt lähimat väärtust ülaltpoolt (see valik eeldab, et **tabel on sorteeritud kahanevalt**, vastasel juhul võib tulemus olla vale).

Näiteks annab järgmine funktsioon Mari järjekorranumbriks teises tabelis arvu kaks:

	A	B	C	D	E	F	G	H	I
24									
25	Nimi	Hinne	Jrk. nr.	Kursus			Nimi	Kursus	Hinnang
26	Mari	A	=MATCH(\$A26;\$G\$26:\$G\$29;0)				Jüri	I	Loru
27	Mari	B					Mari	I	Usin
28	Jüri	C					Mati	II	
29	Mati	B					Siim	III	

Nimi	Hinne	Jrk. nr.	Kursus
Mari	A	2	
Mari	B	2	
Jüri	C	1	

Funktsioon INDEX tagastab väärtuse määratud piirkonna ette antud järjekorranumbriga positsioonilt (**ükskõik kas reas, veerus või tabelis – esimesel kahel juhul tuleb ette anda vaid vastavalt veeru või rea number, viimasel juhul mõlemad!**). Argumentidena tuleb määrata

- piirkond (rida või veerg või tabel), kust võtta funktsiooni poolt tagastatav väärtus, ja
- veeru, rea või mõlema numbrid, mis määravad ära tagastatava väärtuse asukoha ja mis on võimalik määrata funktsiooniga MATCH.

Seega, kombineerides omavahel funktsioone INDEX ja MATCH, on võimalik luua universaalne tabelite ühendamisel kasutatav funktsioon:

	A	B	C	D	E	F	G	H	I
25	Nimi	Hinne	Jrk. nr.	Kursus			Nimi	Kursus	Hinnang
26	Mari	A	2	I			Jüri	I	Loru
27	Mari	B	2	I			Mari	I	Usin
28	Jüri	C	1	I			Mati	II	Rumal, aga ec
29	Mati	B	3	II			Siim	III	Oivik
30	Mati	C	3	II					
31	Mati	A	3	=INDEX(H\$26:H\$29;MATCH(\$A31;\$G\$26:\$G\$29;0))					
32	Siim	A	4	III					
33	Siim	A	4	III					

Otsitava väärtuse mitteleidmisel tulemuseks oleva veateate #N/A asemel tühja lahtri saamiseks võib rakendada juba eelnevalt kirjeldatud funktsiooni IFERROR.

Ja nüüd võib asuda pullide, lehmade ja kontroll-lüpside andmete ühendamise juurde.

Seejuures pange tähele, et lehmade andmetabelis on lehmade numbrite ees täht E, mida kontroll-lüpside andmetabelis ei ole. Nende kahe tabeli lehma numbri järgi ühendamiseks tuleb kas tekitada ühes tabelis uus ilma E-täheta lehmanumbrite veerg või siis lisada teise tabelisse E-tähga lehmanumbrite veerg.

Et teine variant on lihtsam, siis olgu siinkohal räägitud vaid sellest:

- tekitage kontroll-lüpside tabelisse lehma numbrite järel uus tühi veerg (näiteks nimega E_number),
- kirjutage sinna teise ritta valem ="E"&A2 ja
- kopeerige see valem tabeli lõpuni (vt järgmist nippide kasti!).

	A	B
1	Inv_nr	E number
2	1125	= "E"&A2

Nipid, märkused, soovitused.

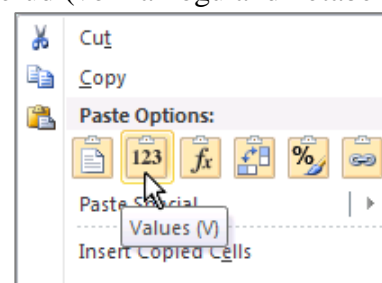
Suurte andmebaaside korral on sisestatud valemi kopeerimine kõigisse järgnevasse ridadesse lihtsalt valemit sisaldavat lahtrit alumisest paremast nurgast lohistades aeganõudev protsess. Lihtsam on

- kopeerida valemiga lahter arvuti vahemällu (Ctrl+C), võtta blokki kõik ülejäänud samas veerus paiknevad lahtrid (klikkides neist esimesel, kerida kerimisriba abil viimase lahtrini ja klikkida Shift-klahvi all hoides seal) ning kleepida valem sinna (Ctrl+V),
- võtta blokki kõik antud veeru lahtrid alates valemist sisaldavast lahtrist ning valida *Home*-sakilt käsk *Fill* → *Down*,
- **teha topeltklõps valemiga lahtri alumisel vasakul nurgal** (samal kohas, kust lohistades saab valemist kopeerida järgnevasse lahtritesse) – **Excel kopeerib sama valemi automaatselt kõigisse lahtritesse kuni andmetabeli lõpuni**, sama nipp töötab ka enam kui ühe veeru ja valemi puhul!

Lõpuks, kui kõik vajalikud veerud on ühendatud, oleks mõistlik kaotada andmetabelist ära valemid ja jätta alles üksnes nende tulemused. Selleks

- võtate blokki andmetabelite ühendamise tulemusena saadud veerud (või ka kogu andmetabeli)
- vajutate Ctrl+V ja
- seejärel valite käsu *Paste Special* → *Values*:

Taoline tegevus on mõttekas vältimaks andmete muutumist mingite andmetabeleis tehtud muudatuste tõttu ning hoidmaks kokku aega, sest igal andmetabeli avamisel ja ka lihtsalt sorteerimisel arvutab Excel kõigi lahtrite väärtused uuesti, mis suurte andmetabelite puhul võib olla üsna töömahukas.



UUTE TUNNUSTE ARVUTAMINE

- Lisage kontroll-lüpside tabelisse laktatsiooni numbri (LNR) järele uus veerg, andke sellele nimeks LNR_kood ja arvutage väärtused lähtudes eeskirjast:

$$\text{LNR_kood} = \begin{cases} \text{LNR, kui } \text{LNR} \leq 3 \\ 4, & \text{kui } \text{LNR} > 3 \end{cases}$$

... X ✓ fx =IF(C2<4;C2;4)			
	B	C	D
1	Nimi	LNR	LNR kood
2	PURRU	1	=IF(C2<4;C2;4)

- Lisage kontroll-lüpsi kuupäeva (KL_kp) järele uus veerg ja arvutage sinna lüpsipäeva (LPäev) väärtus kui vahe kontroll-lüpsi kuupäeva ja poegimiskuupäeva vahel.

... X ✓ fx =F2-E2				
	D	E	F	G
1	LNR_kood	Poeg_kp	KL_kp	Lpäev
2	1	24.09.2004	16.10.2004	=F2-E2

Kuna Excel tõlgendab kahe kuupäeva vahet vaikumisi samuti kuupäevana, tuleb vahe päevade arvuna kuvamiseks muuta lahtri vormingut:

Wrap Text		Date	→		Wrap Text		General	
Merge & Center					Merge & Center			
G	H	I	J		G	H	I	J
Lpäev	Ljark	Piim	Rasva		Lpäev	Ljark	Piim	Rasva
22.01.1900	1	26,2			22	1	26,2	

Muide, Excel tõlgendab kõiki kuupäevi tegelikult päevade arvudena oma ajaarvamise algusest 1.01.1900. St, et vormindades numbri 1 kuupäevana saate tulemuseks justnimelt 1. jaanuari aastal 1900. Proovige! Arv 22 vastab siis aga kuupäevale 22.01.1900 nagu eelnevalt jooniseltki näha.

- Arvutage lüpsipäeva alusel laktatsioonijärk (Ljark), mille väärtuseks lugege
 - 1, kui lüpsipäev on 60 päeva või alla selle,
 - 2, kui lüpsipäev on 61-150 päeva,
 - 3, kui lüpsipäev on üle 150 päeva.

Kui te aru ei saa, kuidas need mitu IF-käsku üksteise sees toimivad, küsige.

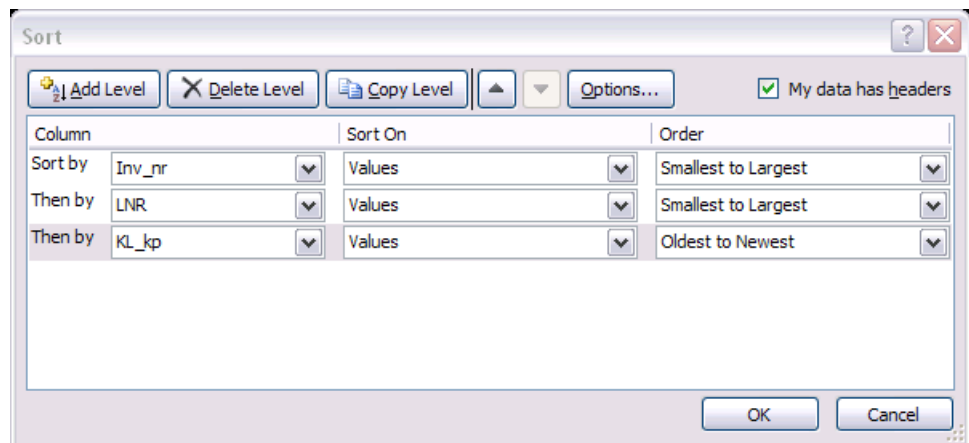
... X ✓ fx =IF(G2<=60;1;IF(G2<=150;2;3))			
	F	G	H
1	KL_kp	Lpäev	Ljark
2	16.10.2004	22	=IF(G2<=60;1;IF(G2<=150;2;3))

- Arvutage igasse ritta ka somaatiliste rakkude skoori (SRS) väärtused valemist $SRS = \log_2(SRA/100) + 3$ (siin SRA on somaatiliste rakkude arv, mis antud andmebaasis on esitatud tuhandetes, st et väärtus 31 tähendab tegelikult 31000-t).

... ▾ X ✓ fx =LOG(L2/100;2)+3				
	K	L	M	N
1	Valgu_pr	SRA	SRS	
2	3,47	31	=LOG(L2/100;2)+3	

- Arvutage igasse ritta konkreetse lehma kontroll-lüpsi number laktatsioonil ning kahe järjestikuse lüpsipäeva vahe. Kui tegu on lehma esimese kontroll-lüpsiga laktatsioonil, arvutage kaugus poegimisest.

Selle ülesande lahendamiseks tuleb kontroll-lüpside andmestik sorteerida esmalt lehma numbril, seejärel laktatsiooninumbri ning seejärel lüpsipäeva või kontroll-lüpsi kuupäeva järgi:



Lehma kontroll-lüpsi numbril ja lüpsipäevade vahe arvutamiseks tuleb mõista, et tänu andmete vastavale sorteeritusele on alati,

- kui eelmises andmebaasi reas oli teise numbriga lehm või teine laktatsiooninumber, tegu vaatlusaluses reas paikneva lehma ja laktatsiooni esimese kontroll-lüpsiga ning lüpsipäevade vahe arvutamise asemel tuleb leida vahe poegimise ja kontroll-lüpsi kuupäevade vahel,
- kui aga eelmises andmebaasi reas on sama lehma sama laktatsiooni kontroll-lüps, siis on vaatlusaluse rea kontroll-lüpsi number ühe võrra suurem kui eelmises reas ja kontroll-lüpside vahe on leitav vaatlusaluse ja eelmise rea kontroll-lüpsi kuupäevade vahena.

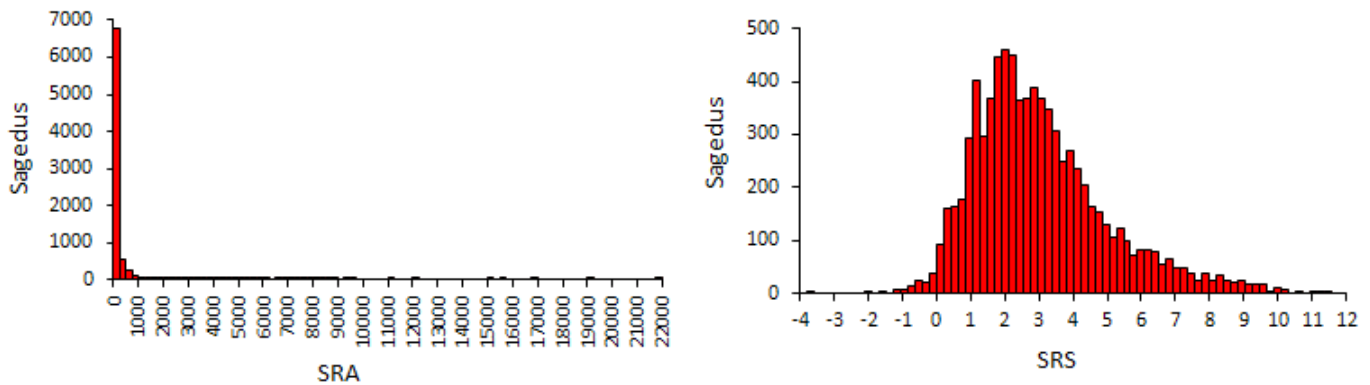
	A	B	C	D	E	F	G	U	V	W	X	Y
1	Inv_nr	E_numk	Nimi	LNR	LNR_k	Poeg_kp	KL_kp	KL_nr	KL vahe			
2	1125	E1125	PURRU	1	1	24.09.2004	16.10.2004	=IF(AND(A2=A1;D2=D1);U1+1;1)	=IF(AND(A2=A1;D2=D1);G2-G1;G2-F2)			
3	1125	E1125	PURRU	1	1	24.09.2004	20.11.2004					

Peale valemite kopeerimist andmetabeli lõpuni oleks jällegi mõistlik teha nendes veergudes paiknevatest väärtustest koopia (st kaotada ära väärtuste taga olevad valemid), sest nende valemite tulemused sõltuvad ridade järjestusest andmetabelis ja lähevad valeks, kui andmetabelit teisiti sorteerida või mõni rida vahelt ära kustutada.

JA LÕPETUSEKS – HISTOGRAMMI KONSTRUEERIMINE JA EXCELI LISAMOODULITE KASUTAMINE

Saamaks teada, kas teostatud teisendus muutis somaatiliste rakkude arvu jaotuse sümmeetriliseks (enam normaaljaotusele sarnanevaks), konstrueerige sagedustabelid nii somaatiliste rakkude arvule (SRA) kui ka somaatiliste rakkude skoorile (SRS) ning illustreerige neid histogrammiga.

Soovitav tulemus:



Ülesande võite lahendada nii eelmise praktikumi tööjuhendi alusel kui ka kasutada internetist tasuta alla laaditavat Exceli lisamoodulit *Better Histogram* (<http://www.treeplan.com/download-free-better-histogram-add-in.htm>).

Soovides kasutada nimetat lisamoodulit (ja näitena taolistest tööriistadest võiks seda teha),

- laadige alla vastav zip-fail,
- pakkige see fail lahti,
- ning võtke lisamoodul



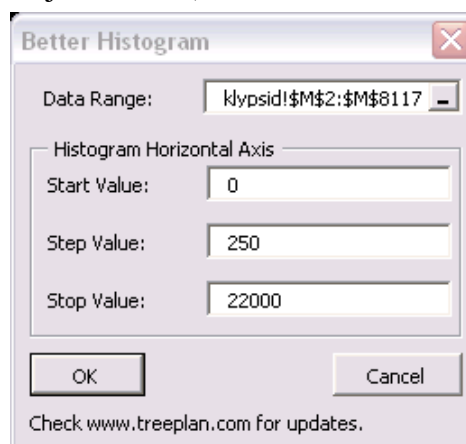
Excelis kasutusse

(File → Options → Add-Ins → Manage |Excel Add-ins| [Go...] → Browse... → *Better-Histogram-2007.xla*).

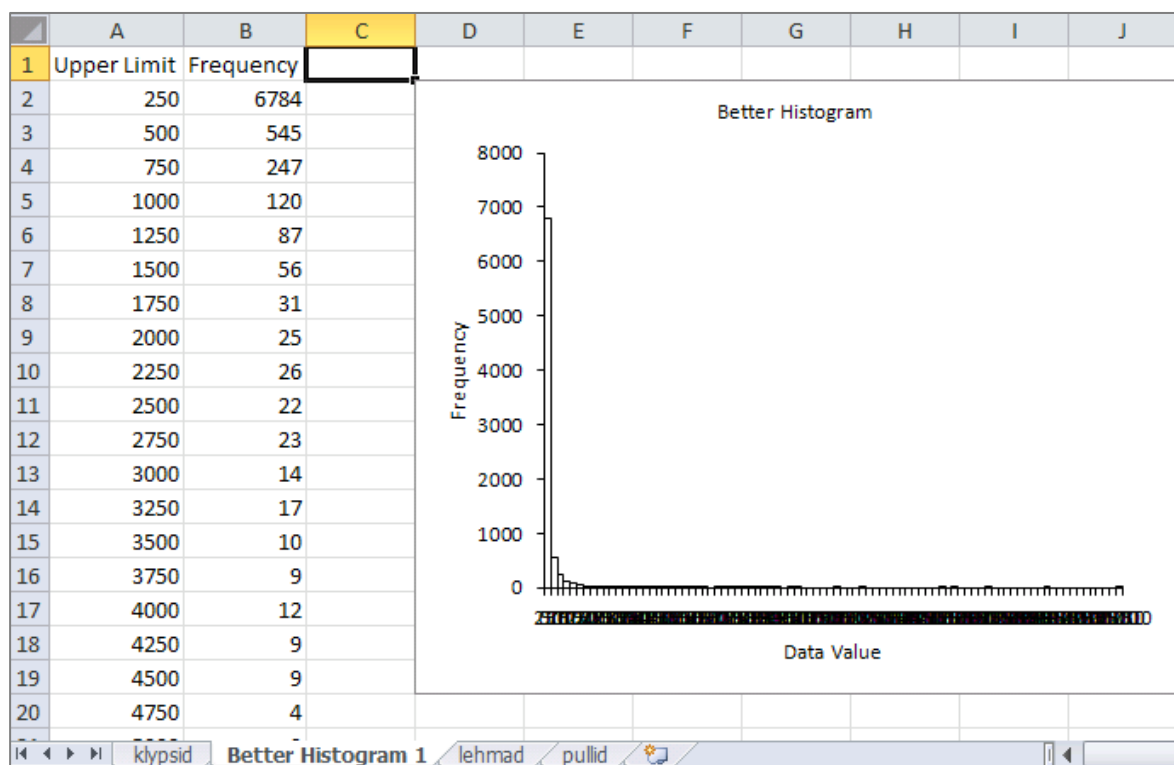
Tulemusena tekib Exceli menüüribale *Add-Ins*-sakk, mille all ongi kasutamiskvalmis vastav lisatööriist.

Et *Better Histogram* tahab kasutajalt teada, mis väärtusest mis väärtuseni ja millise sammuga tuleb sagedustabel ja selle alusel histogramm konstrueerida, on mõttekas need arvud eelnevalt paika panna. Lihtsaim viis on võtta blokki vastav veerg andmetabelis, mis võib olla selekteeritud ka tööriista *Better Histogramm* rakendamise hetkel – siis on hea vaadata, mis piirides tunnuse väärtused muutuvad, kui palju neid on ja teha selle alusel otsus nõutud väärtuste osas:

Somaatiliste rakkude arvu miinimum on 1 ja maksimum 21938. Vaatluste arv 8116 ütleb, et klasse võiks teha suurusjärku 100 tükki või pisut alla selle. Kui teha klassid 0-st 22000-ni sammuga 250, siis tuleb neid 88 tükki – võiks ju sobida (kui ei sobi, saab tööriista *Better Histogramm* abil kiirelt uue analüüsi teha):



Tulemuseks on uuele töölehele konstrueeritud sagedustabel ja vastav histogramm:



Jäänud on veel kujundada joonis ilusaks (nagu on esitatud eelmisel leheküljel) ja teha sama analüüs ka somaatiliste rakkude skooriga.