

Praktikum 3



Tänase praktikumi teema on andmetabelite filtreerimine ja kokkuvõtivate tabelite loomine, juttu tulebka mõningatest pisut nutikamatest funktsioonidest keskmiste ja vaatluste arvu arvutamisel.

SORTEERIMINE JA FILTREERIMINE

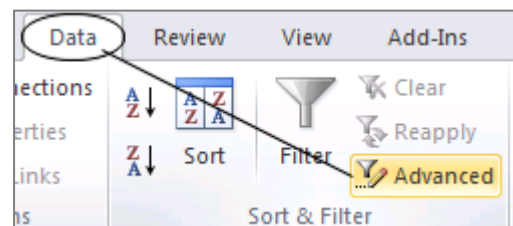
Märksõnad:

andmeid mitte blokki võtta

sorteerimine ridades

Rakendades mingit funktsiooni (näiteks keskmise arvutamise funktsiooni AVERAGE) filtreeritud andmetele, kasutab Excel funktsiooni argumentidena siiski kõiki, mitte üksnes näha olevaid andmeid!

Advanced Filter (edumeelne filter?)

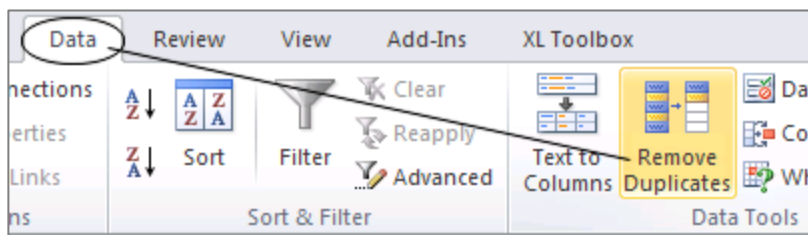


Tulemus teisele töölehele!

Unique values

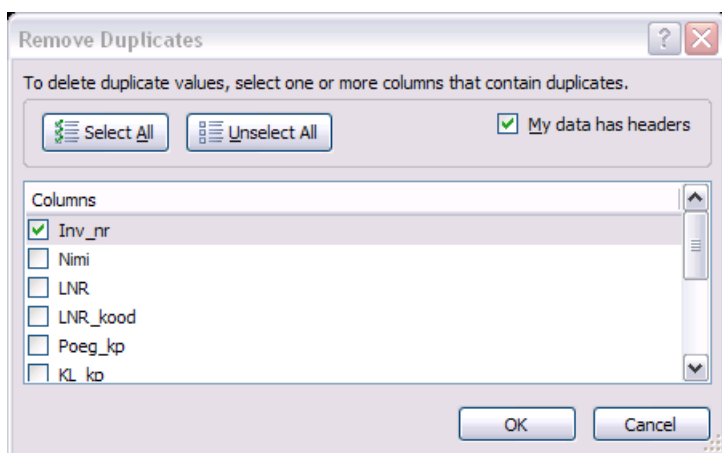
KORDUVATE RIDADE/VÄÄRTUSTE ERALDAMINE

Eemaldamiseks andmetabelist korduvate väärtustega rida tuleb rakendada *Data*-sakilt käsku *Remove Duplicates* (seejuures peab kursor olema mistahes lahtris andmetabeli sees).



Nipid, märkused, soovitused.

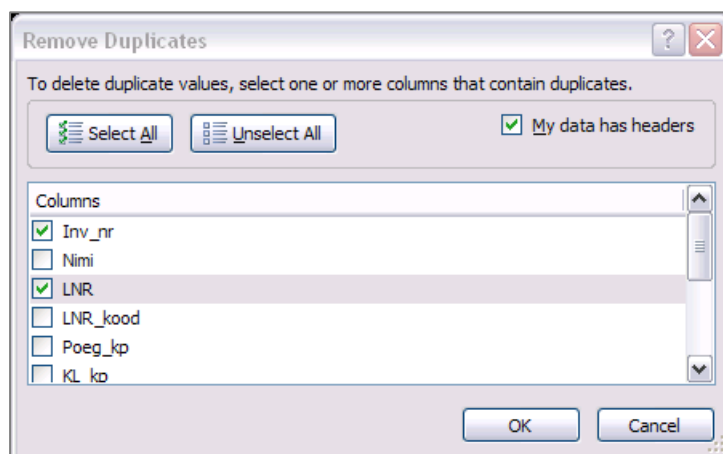
- Et käsu *Remove Duplicates* tulemusena korduvate väärtustega read andmetabelist tõepoolest kustutatakse, oleks enne selle käsu rakendamist soovitatav teha andmetabelist uuele töölehele (või lausa uude faili) koopia ja rakendada korduvate väärtustega ridade eraldamise käsku seal (kõik see ei ole muidugi vajalik, kui te oma le parisis täpselt aru annate, mida te teete ☺).
- Korduvate väärtuste olemasolul jäetakse neist alles vaid esimene (selles järjekorras, nagu nad parasjagu tabelis on).



Kuidas tuleks toimida, kui soovite alles jätta vaid iga lehma viimasele kontroll-lüpsile vastavat rida, aga seda iga laktatsiooni kohta?

Vastus:

- esmalt tuleks andmetabel sorteerida lehma numbril ning selle siseselt laktatsioonil numbril ja seejärel ka kahanevalt laktatsioonipäeva järgi,
- misjärel tuleks rakendada korduvate väärtuste eraldamise käsku, kus korduvate väärtuste all mõistetakse kõiki ridu, milles nii lehma number kui ka laktatsiooni-number korduvad.



Kopeerige andmetabel (käsuga *Paste Special* → *Values*) uude faili ja proovige.

--- Pivot Table'i kasutamine ---

Kui aga on küll soov saada tulemuseks tabelit, kus iga lehm oleks küll ühekordselt, aga talle vastava esimese rea asemel oleks kirjas toodangunäitajate keskmised, tuleb kasutada Pivot Table abi.

Ülesanne.

Tekitage kontroll-lüpside andmebaasi alusel tabel, kus igale lehmale ja laktatsioonile vastaks vaid üks rida ning selles reas oleks kirjas lehma kontroll-lüpside arv laktatsioonil ning keskmine piimatoodang ja rasva- ja valguprotsent.

	A	B	C	D	E	F
2						
3	Inv_nr	LNR	Count of Piim	Average of Piim2	Average of Rasva_pr	Average of Valgu_pr
4	1125	1	9	25,04444444	4,471111111	3,622222222
5	1125	2	11	33,77272727	4,299090909	3,737272727
6	1125	3	8	37,475	3,72	3,53875
7	1125	4	10	39,64	4,291	3,599
8	1125	5	9	40,31111111	4,075555556	3,582222222
9	1125	6	15	32,32	4,750666667	3,854666667
10	1176	1	23	24,96086957	4,629565217	3,616956522
11	3177	1	10	36,11	3,916	3,47
12	3177	2	10	31,38	4,505	3,679
13	3177	3	13	34,46153846	4,24	3,673846154
14	3177	4	9	31,08888889	4,426666667	3,571111111
15	3195	1	13	42,03076923	2,840769231	3,501538462
16	3195	2	11	40,81818182	3,383636364	3,540909091
17	3195	3	8	33,7125	4,2	3,99125
18	3197	1	13	33,2	4,758461538	3,498461538
19	3197	2	13	40,01538462	5,006923077	3,577692308

TINGLIKUD JA KAALUTUD KESKMISED, SUMMAD JA LOENDUSED

Funktsioonid

AVERAGEIF

AVERAGEIFS

Ülesanne.

Arvutage nii kõigi kui ka vaid esimese laktatsiooni lehmade

- o keskmine esimese kontroll-lüpsi kaugus poegimisest ja
- o keskmine kahe kontroll-lüpsi vaheline aeg.

Analoogselt töötavad ka funktsioonid SUMIF ja SUMIFS ning sarnaselt funktsioonid COUNTIF ja COUNTIFS – esimesed kaks siis vastavalt väärtuste summa ja teised kaks väärtuste arvu leidmiseks.

Ülesanne.

Leidke,

- kui palju on andmetabelis esimesi laktatsioone (mitte kontroll-lüpe vaid justnimelt laktatsioone), aga kui palju teisi ja näiteks viiendaid;
- kui palju on kontroll-lüpe, kus somaatiliste rakkude arv (SRA) on üle 2000 (st, et tegelikult üle 2 miljoni – Jõudluskontrolli Keskus esitab somaatiliste rakkude arvu lihtsalt tuhandetes)?

Alternatiiv mistahes arvkarakteristiku arvutamiseks ette antud tingimustele vastavaist ridadest:

	A	B	C	D	E	F	G	H	I	J	
1	Inv_nr	E_numk	Nimi	LNR	LNR_kood	Poeg_kp	KL_kp	Lpäev	Lärk	Piim	Rasv
8115	5370	E5370	ÕKA	1	1	21.11.2010	13.01.2011	53	1	39,5	
8116	5376	E5376	RILLA	1	1	21.12.2010	13.01.2011	23	1	31,1	
8117	5448	E5448	LUSTI	1	1	30.12.2010	13.01.2011	14	1	33,4	
8118											
8119											
8120											
8121			Keskmine piimatoodang lehmadel, kelle nimi on PURRU								
8122			34,527		=AVERAGE(IF(C2:C8117="PURRU";J2:J8117;""))						
8123											
8124											
8125			34,527		=AVERAGEIF(C2:C8117;"PURRU";J2:J8117)						
8126											
8127			34,527		=AVERAGEIFS(J2:J8117;C2:C8117;"PURRU")						

NB! Tegu on massiivfunktsiooniga, see tähendab, et funktsiooni rakendamiseks tuleb **Enter**-klahvile vajutades hoida all klahve **Ctrl** ja **Shift**!

Funktsiooni AVERAGE asemel võib kasutada ka teisi kirjeldava statistika funktsioone (MEDIAN, STDEV.S, MIN, MAX jne).

Kaalutud keskmine

=SUMPRODUCT(B4:B13;C4:C13)/SUM(C4:C13)

Erinevate väärtuste arvu leidmine

Kui mitu erinevat lehma?

Kui puuduvaid väärtuseid pole, töötab ka funktsioon

=SUM(1/COUNTIF(A2:A8117;A2:A8117))

A...		fx {=SUM(IFERROR(1/COUNTIF(A2:A8117;A2:A8117);0))}									
	A	B	C	D	E	F	G	H			
1	Inv_nr	E_numk	Nimi	LNR	LNR_kood	Poeg_kp	KL_kp	Lpäev			
8116	5376	E5376	RILLA	1	1	21.12.2010	13.01.2011	2			
8117	5448	E5448	LUSTI	1	1	30.12.2010	13.01.2011	1			
8118											
8119			Erinevate lehmade arv								
8120			411		=SUM(IFERROR(1/COUNTIF(A2:A8117;A2:A8117);0))						

NB! Tegu on jällegi massiivfunktsiooniga (st, Ctrl+Shift+Enter).

ÜMARDAMINE

Trükkige Exceli töölehele mõned ümardamist vajavad arvud, näiteks π , $-\pi$ ja $15*\pi$ – muideks, arvu π leidmiseks on Excelis funktsioon PI() – ja uurige, kuidas töötavad järgmised funktsioonid (arv tähendab ümerdatavat arvu):

ROUND(arv;2)

ROUNDDOWN(arv;1)

ROUNDUP(arv;1)

INT(arv)

TRUNC(arv)

ODD(arv)

EVEN(arv)

CEILING(arv;5)

FLOOR(arv;5)

MROUND(arv;5)

	A	B	C	D
1		=PI()	=-PI()	=15*PI()
2		3,141593	-3,14159	47,12389
3	ROUND			
4	ROUNDDOWN			
5	ROUNDUP			
6	INT			
7	TRUNC			
8	ODD			
9	EVEN			
10	CEILING			
11	FLOOR			
12	MROUND			

VEEL NÄITEID TEKSTIFUNKTSIOONIDE KASUTAMISEST

Olgu lahtrisse B2 kirjutatud lause „Täna paistab päike, homme sajab vihma“, lahtrisse C2 nimi „Jüri Jänes“ ja lahtrisse D2 sõna „lill“:

	A	B	C	D
1	Tekst			
2		Täna paistab päike, homme sajab vihma	Jüri Jänes	lill

Rakendage järgnevaid funktsioone sisestatud tekstidele ja püüdke aru saada, mida need funktsioonid teevad ("tekst" tähendab konkreetset teksti või seda sisaldava lahtri aadressi).

LEFT("tekst";3)

RIGHT("tekst";4)

MID("tekst";2;3)

LEN("tekst")

FIND("a";"tekst")

SUBSTITUTE("tekst";"a";"@";2)

--- Lahtri väärtusest esimese sõna eraldamine ---

Olgu lahtrisse B2 kirjutatud lause „Täna paistab päike, homme sajab vihma“, lahtrisse C2 nimi „Jüri Jänes“ ja lahtrisse D2 sõna „lill“:

	A	B	C	D
1	Tekst			
2		Täna paistab päike, homme sajab vihma	Jüri Jänes	lill

Lause esimese sõna (sümbolid kuni esimese tühikuni) välja kirjutamiseks saab kasutada funktsiooni kujul

$$=LEFT(B2;FIND(" ";B2)-1)$$

- Funktsioon FIND(" ";B2) leiab, kui mitmendal positsioonil asub lahtris B2 paiknevas tekstis esimene tühik – käsu FIND(" ";B2)-1 tulemus on siis seega teksti vasakpoolseima sõna viimase sümboli positsioon (üks sümbol enne tühikut);
- funktsioon LEFT aga võtab lahtris B2 paiknevast tekstist vasakult poolt välja funktsiooniga FIND()-1 leitud sümbolite arvu.

Aga vaid üht sõna sisaldava lahtri puhul annab taoline lihtne funktsioon veateate, sest otsitavat tühikut ei leidu:

	A	B	C	D	E	F	G
1	Tekst						
2		Täna paistab päike, homme sajab vihma	Jüri Jänes	lill		Valem lahtri B2 jaoks	
3							
4	Esimene sõna						
5		Täna	Jüri	#VALUE!		=LEFT(B2;FIND(" ";B2)-1)	

Lahenduseks on kasutada Excel 2010-s funktsiooni IFERROR, mille esimene argument on teostatav operatsioon ja teine argument operatsioon siis, kui esimese operatsiooni teostamine lõppeb veateatega (kui tekst on ühesõnaline, lasemegi Excelil väljastadaa vaid selle ühe sõna):

$$=IFERROR(LEFT(B2;FIND(" ";B2)-1);B2)$$

Kopeerides selle funktsiooni ka ülejäänud veergudesse on tulemus järgmine:

	A	B	C	D	E	F	G	H	I
1	Tekst								
2		Täna paistab päike, homme sajab vihma	Jüri Jänes	lill		Valem lahtri B2 jaoks			
3									
4	Esimene sõna								
5		Täna	Jüri	#VALUE!		=LEFT(B2;FIND(" ";B2)-1)			
6		Täna	Jüri	lill		=IFERROR(LEFT(B2;FIND(" ";B2)-1);B2)			

Et Excel 2003-s funktsioon IFERROR puudub, tuleb seal rakendada IF-funktsiooni kombinatsioon funktsiooniga ISERR (viimase väärtus on tõene, kui argumentina ette antud operatsiooni tulemus annab veateate, ning väär veateate puudumisel):

$$=IF(ISERR(FIND(" ";B2));B2;LEFT(B2;FIND(" ";B2)-1))$$

Tekst							
	Täna paistab päike, homme sajab vihma	Jüri Jänes	lill		Valem lahtri B2 jaoks		
Esimene sõna							
	Täna	Jüri	#VALUE!		=LEFT(B2;FIND(" ";B2)-1)		
	Täna	Jüri	lill		=IFERROR(LEFT(B2;FIND(" ";B2)-1);B2)		(Excel 2010)
	Täna	Jüri	lill		=IF(ISERR(FIND(" ";B2));B2;LEFT(B2;FIND(" ";B2)-1))		(Excel 2003)

--- Lahtri väärtusest esimese sõna järgse tekstiosa eraldamine ---

$$=RIGHT(B2;LEN(B2)-FIND(" ";B2;1))$$

--- Lahtri väärtusest viimase sõna eraldamine ---

=RIGHT(B2;LEN(B2)-FIND(" ";SUBSTITUTE(B2;" ";"*";LEN(B2)-LEN(SUBSTITUTE(B2;" ";"")))))

Ülesanne.

Tekitage pullide tabelisse lisaveerg, mis sasaldab üksnes pulli nime ilma sellele järgneva sidekriipsuta ja tähtedeta ET (näiteks LANCELOT-ET → LANCELOT), kui pulli nimi ei sisalda laiendit -ET, tuleb nimi jätta samaks.

Soovitav tulemus:

	A	B
1	Isa_nimi	Pull
2	DYNASTY-ET	DYNASTY
3	OLIVER	OLIVER
4	MARION-ET	MARION
5	AUGUSTINE-ET	AUGUSTINE
6	APOLLO-ET	APOLLO
7	COMBAT	COMBAT
8	JANOS	JANOS
9	LANCELOT-ET	LANCELOT
10	DORADO-ET	DORADO
11	WIZZARD-ET	WIZZARD
12	RAMOS	RAMOS
13	JOSE	JOSE

--- Isikukoodist soo ja sünniaja arvutamine ---

Oletame, et isikukood paikneb tabeli lahtris A2. Leidmaks selle alusel kõrvallahtrisse koodi omaniku sugu, võib kasutada valemit kujul

=IF(MOD(LEFT(A2;1);2)=0;"naine";"mees").

- Siin funktsioon LEFT(A2;1) võtab isikukoodist välja vasakult poolt esimese numbril, funktsioon MOD(LEFT(A2;1);2) leiab väljavõetud numbril kahega jagamisel tekkiva jäägi ning
- funktsioon IF kontrollib, kas jääk on 0 (siis on isikukoodi esimese numbril näol tegu paarisarvuga) või mitte (siis on isikukoodi esimese numbril näol tegu paaritu arvuga) ning väljastab vastavalt väärtuse naine (isikukood algab 2, 4 või 6-ga) või mees (isikukood algab 1, 3 või 5-ga).

Sünniaja arvutamiseks isikukoodi järgi võib kasutada valemit

=DATE(MID(A2;2;2);MID(A2;4;2);MID(A2;6;2))

- Siin funktsioon MID võtab isikukoodist välja teise argumandiga määratud kohast kaks sümbolit ja
- funktsioon DATE ühendab aastaarvu, kuu numbril ja päeva numbril aastaarvuks.

Ülesanne.

Kontrollige, kas taolised valemid kehtivad ka teie isikukoodi puhul.

	A	B	C
1	Isikukood	Sugu	Sünniaeg
2	37112172714	mees	17.12.1971
3		=IF(MOD(LEFT(A2;1);2)=0;"naine";"mees")	=DATE(MID(A2;2;2);MID(A2;4;2);MID(A2;6;2))

Kas kirjeldatud sünniaja arvutamise valem annab alati õige tulemuse? Aga 21. sajandil sündinud isiku puhul? Proovige (muutes enese isikukoodis sisalduvad sünniaastat näitavad numbrid näiteks '05'-ks).

Lahendus?

Näiteks kasutada kahte IF-funktsiooni üksteise sees määramaks isikukoodi esimese numbriga järgi sajandit. Proovige kirja panna!

Aga saab ka nii:

=DATE(((ROUNDUP(LEFT(A10;1)/2;0)+18)*100)+MID(A10;2;2);MID(A10;4;2);MID(A10;6;2))

Püüdke loogikast aru saada!

LISAÜLESANNE FILTREERIMISEST, KESKMISTEST, DISPERSIOONANALÜÜSIST JA JOONISTEST

1. Tekitage andmetabel, mis sisaldab vaid iga lehma esimese laktatsiooni esimese kontroll-lüpsi tulemusi.
2. Viige läbi dispersioonanalüüs testimaks pullide vahelise erinevuse statistilist olulisust (nende tütarde keskmiste esimese laktatsiooni esimese kontroll-lüpsi tulemuste alusel, ma mõtlen).
3. Illustreerige analüüsi tulemusi tulpdiagrammiga, kus
 - igale pullile vastab üks tulp,
 - tulba kõrgus näitab pulli tütarde keskmist piimatoodangut (see väärtus kirjutage ka tulba sisse),
 - veajooned näitavad standardviga ning
 - lisaks on iga pulli kohta ära toodud ka tema kõigi tütarde piimatoodangud.
4. Teades, et pullide paarikaupa võrdluse (Tukey test) tulemustel olid pulli Janos tütreid statistiliselt oluliselt erinevad ($p < 0,05$) pullide Lancelot-ET ja Ramos tütardest ning muid statistiliselt olulisi erinevusi pullide vahel ei ilmenud, lisage joonisele tähed visualiseerimaks pullide vaheliste erinevuste statistilist olulisust.

Lahendus.

1. Iga lehma esimese laktatsiooni esimese kontroll-lüpsi andmestiku võib tekitada
 - jättes laktatsiooninumbri ja selle siseselt lüpsipäeva või kontroll-lüpsi järjekorranumbri alusel sorteeritud andmetabelist käsu *Remove Duplicates* abil alles vaid soovitud read (kuna andmebaas on sobivalt sorteeritu, jäävadki iga lehma kohta alles vaid esimese laktatsiooni esimesed kontroll-lüpsid),
 - kopeerides vajalikele tingimustele vastavad read sobivalt sorteeritud andmetabelist teise faili või töölehele (sorteerida tuleb siis jällegi laktatsiooninumbri ja selle siseselt lüpsipäeva või kontroll-lüpsi järjekorranumbri alusel),
 - filtreerides välja vaid vastavatele tingimustele vastavad read ja kopeerides need teise faili või töölehele.
2. *Data-sakk* → *Data Analysis* → *Anova: Single Factor*

Juhendi tarvis vt näiteks http://www.eau.ee/~ktanel/VL_0781/VL1_praks8_2013k.pdf

3.-4. Eesmärk:

