

Praktikum 5

Salvestage kursuse kodulehelt omale arvutisse andmestik *sead.xls*.

Kommentaari andmestiku kohta

Tegu on 2003. aastal teostatud katsega, kus 80-st seast 40 peeti uues külmaudas ja 40 vanas nõukogudeaegses sigalas, mõlemas sigalas söödeti pooli sigu (so 20 tk) välismaise söödaga ja pooli kodumaise söödaga, kõigist 20-sealistest gruppidest pooled loomad tapeti kohalikus tapamajas (nö üle õue), aga pooltele korraldati stressirohke reis kitsas autokastis Eestimaa teises otsas paiknevasse tapamajja. Kõigi rümpade puhul mõõdeti hulk lihakvaliteedinäitajaid.

Ülesanded

Praktikumi tehniline pool püüab anda juhiseid, kuidas võimalikult optimaalselt teostada suurt hulka sarnaseid teste ja kuidas saadud suurest hulgast tulemustest Exceli tingimusvormingu (*conditional formatting*) abil visuaalselt välja tuua statistiliselt olulisi (või mõnda muud tingimust rahuldavaid) tulemusi.

Statistiliste analüüside poole pealt käsitletakse selliseid klassiklalisi andmeanalüüsimeetodeid nagu t-test, korrelatsioonanalüüs ja regressioonanalüüs.

1. Võrrelge tavapärasel ja külmaudas peetud sigade lihakvaliteedinäitajaid.
 - Arvutage kõigi lihakvaliteedinäitajate kohta keskmised, minimaalsed ja maksimaalsed väärtused ning standardhälbed sõltuvalt pidamiskeskonnast (tavaline või külmlaut).
 - Teostage t-testid selgitamiseks keskmiste kvaliteedinäitajate erinevuse statistilist olulisust (NB! Õige t-testi valimiseks tuleb eelnevalt teostada dispersioonide võrdlus F-testiga).
 - Kasutades Exceli tingimusvormindamist, värvige kõik statistiliselt olulistele erinevustele vastavaid p-väärtuseid sisaldavad lahtrid – kui $p < 0,001$, siis punaseks, $p < 0,01$ korral oranžiks ja $p < 0,05$ korral kollaseks.
2. Millised on erinevate lihakvaliteedinäitajate vahelised seosed sigadel?
 - Arvutage uuele töölehele kõigi lihakvaliteedinäitajate vahelised lineaarsed korrelatsioonikordajad (*Data-sakk → Data Analysis... → Correlation*).
 - Kasutades Exceli tingimusvormindamist värvige oranžiks kõik tugevad seosed ($|r| \geq 0,7$) ja kollaseks kõik keskmise tugevusega seosed ($|r| \geq 0,3$), samuti proovige nõ sujuvat vormingut, kus korrelatsioonikordaja väärtusele -1 vastab sinist, väärtusele 0 valget ja väärtusele 1 punast värvi lahter (tehke korrelatsioonikordajate tabelist koopia – kopeerige vaid väärtused – ja rakendage sujuvat vormingut seal).
 - Arvutage korrelatsioonikordajate statistilist olulisust väljendavad p-väärtused (analoogsesse tabelisse nagu korrelatsioonikordajadki);
 - vormindage p-väärtuste tabel kasutades eelmisel töölehel paiknevate t-testi tulemuste vormingut (*Copy → Paste Special → Formats*),
 - seejärel tehke koopia algse korrelatsioonikordajate tabeli väärtustest (st ärge kopeerige vormingut) ning vormindage see hoopis p-väärtustele tuginedes, värvides (ikka Exceli tingimusvormindamist kasutades) punaseks kõik korrelatsioonikordajad, mille korral $p < 0,001$, oranžiks korrelatsioonikordajad, mille korral $p < 0,01$, ja kollaseks korrelatsioonikordajad, millele vastav $p < 0,05$.
 - Lõpetuseks uurige, kas näiteks tunnuste 'Temp 45min' ja 'Temp 24h' vaheline seos sõltub sigade pidamiskeskonnast – leidke nimetatud tunnuste vahelised lineaarsed korrelatsioonikordajad nii tavapärasel kui ka külmaudas peetud sigadel ning illustreerige

seost hajuvusdiagrammiga, kus erinevatele pidamistingimustele vastavad väärtused on tähistatud erinevalt (lisaks võite seoste enam esile toomiseks lisada punktiparvele regressioonisirged).

Järgnevalt võib funktsiooni F.TEST kopeerida jällegi kõigi veergude alla. Funktsiooni T.TEST nii lihtsalt kopeerida ei saa, kuna selle arvutuseeskiri sõltub F-testi väärtusest – kui t-testi funktsioon kopeerida, tuleb vajadusel ise muuta funktsiooni viimast argumenti (kas 2-ks või 3-ks).

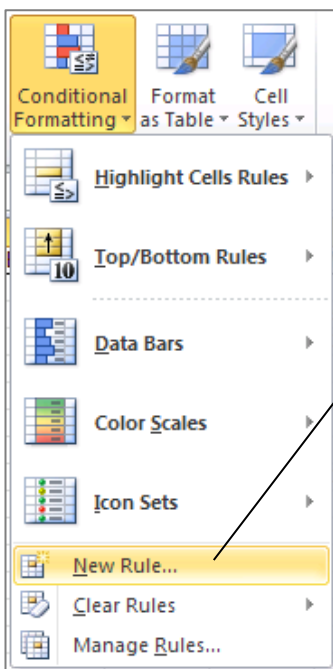
Alternatiiv on kasutada loogikafunktsiooni IF koos funktsioonidega F.TEST (mis määrab tingimuse) ja T.TEST (mille tüüp valitakse automaatselt vastavalt F-testi tulemusele):

F-test	0,0159493
t-test	0,057251
t-test	<code>=IF(F.TEST(E2:E41;E42:E81)<=0,05;T.TEST(E2:E41;E42:E81;2;3);T.TEST(E2:E41;E42:E81;2;2))</code>
	<code>IF(logical_test; [value_if_true]; [value_if_false])</code>

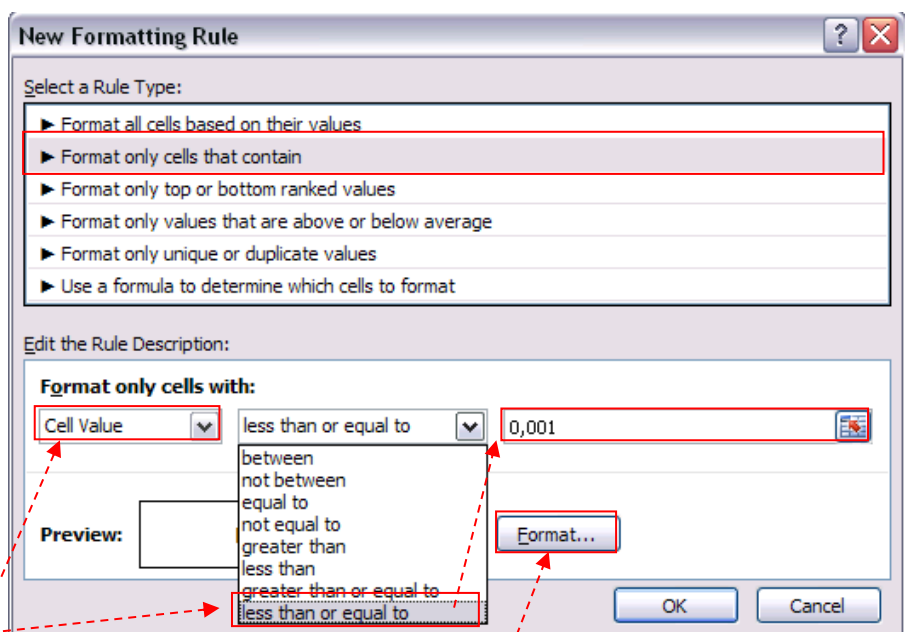
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Place	id_od	Feed	LWDobs14	WCW	CCW	dress %	pH45min	temp45mi	pH24h	temp24h	BackFat1	BackFat2	BackFat3	BackFat4	Meatpct	Moisture	Protein	Fat	Ash
80	LP	Out-door	Import	118	89,7	87,4	74	5,7	40,4	5,8	4,8	9	10	11	39	60,3	67,62	22,2	8,09	1,12
81	LP	Out-door	Import	118	86,1	84	71,1	6,4	39,1	5,8	4,9	16	16	13	35	57,5	67,07	22,8	4,73	1,19
82																				
83																				
84	ine	Keskmine		108,925	73,115	71,415	71,4825	6,1175	38,3925	5,835	2,7025	13,775	14,675	12,9	31,1	57,0725	69,6535	22,6225	5,80525	1,1635
85	ional)	Standardhälve		8,60646	5,1003	5,06256	2,54457	0,24588	1,2970158	0,11447	0,393855	2,69365	2,92108	2,95088	5,96055	3,81065	1,9091	0,86424	2,13605	0,09206
86		Min		90	63	60,6	65,8	5,7	35,1	5,6	1,9	8	9	8	20	47,9	63,8	20	2,31	1
87		Max		124	83,4	82	76,4	6,8	40,3	6,1	3,5	20	20	20	45	63,7	72,63	24	12,63	1,57
88																				
89		Keskmine		112,1	83,1975	81,2625	72,6025	5,96	37,8825	5,905	4,1725	16,55	17,1	25,25	26,075	57,74	69,9195	22,07	5,4745	1,137
90	r)	Standardhälve		5,80804	4,85003	4,77723	2,05669	0,30365	1,7366173	0,11972	0,6417075	4,78754	3,84841	12,5734	12,1937	3,29963	1,70609	1,10829	1,67247	0,07432
91		Min		99	75,4	73,6	68,7	5,4	33,9	5,6	3,2	9	10	8	10	48,1	65,46	19	2,56	0,98
92		Max		124	94	92	78,9	6,6	41,9	6,1	5,4	32	25	50	46	64,9	72,6	23,8	11,14	1,3
93																				
94																				
95		F-test			0,015949															
96		t-test			0,057251															
97																				
98		t-test			0,057															

3. Kasutades Exceli tingimusvormindamist, värvige kõik statistiliselt olulistele erinevustele vastavaid p-väärtusi sisaldavad lahtrid – kui $p < 0,001$, siis punaseks, $p < 0,01$ korral oranžiks ja $p < 0,05$ korral kollaseks

t-test	0,05725	8,2E-14	1,4E-13	0,03344	0,01275	0,1407497	0,00916	1,129E-18	0,00221	0,00215	3E-07	0,02275	0,40486	0,51307	0,01504	0,44299	0,1606
--------	---------	---------	---------	---------	---------	-----------	---------	-----------	---------	---------	-------	---------	---------	---------	---------	---------	--------

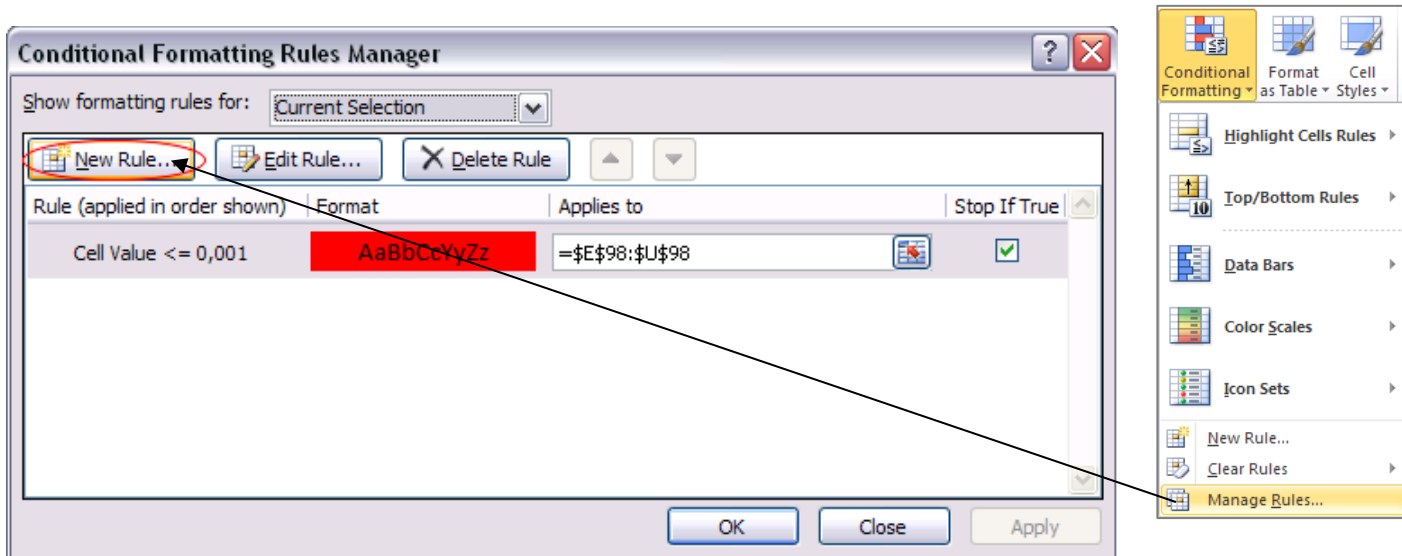


Home-sakk → Conditional Formatting → New Rule...

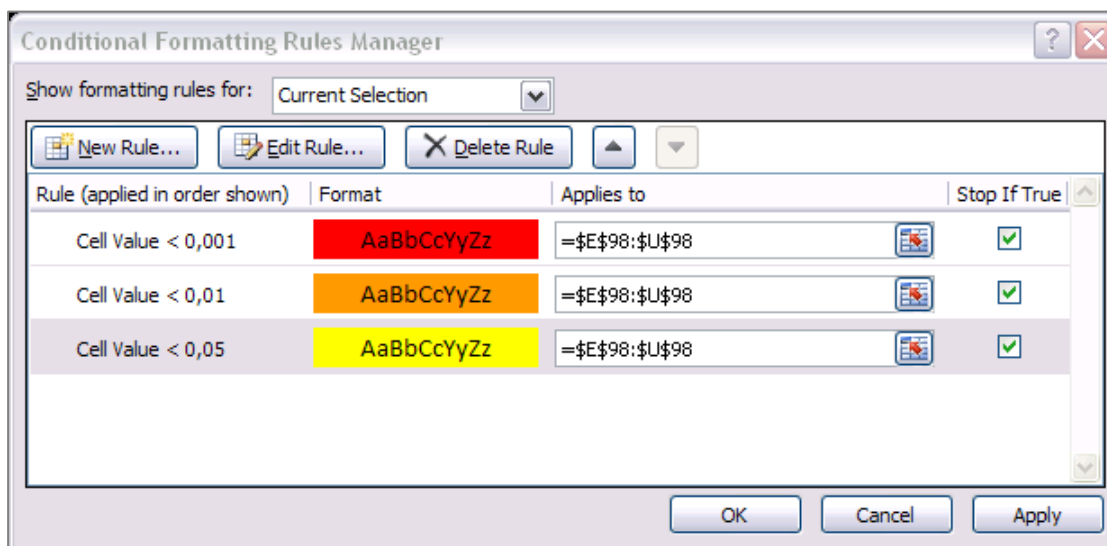


Tingimus, mille täidetuse korral selekteeritud lahtrid vormindatakse soovitud viisil.

Täiendava tingimuse lisamiseks: *Home*-sakk → *Conditional Formatting* → *Manage Rules...*



Lisage vormindamise reeglid nii $p < 0,01$ kui ka $p < 0,05$ tarvis:



Nipid, märkused, soovitused.

Teades ette, et soovite rakendada mitut erinevat vormindamise reeglit, on mõttekas valida koheselt *Home*-sakk → *Conditional Formatting* → *Manage Rules...* Sellest aknast saate

- lisada, muuta ja kustutada vormindamise reegleid ja tingimusi nii parajasti aktiivsete lahtrite kui ka teiste töölehtede ja andmepiirkondade tarvis,
- vaadata eelvaadet vormindamisreeglite rakendamise tulemusest (nupp *Apply*),
- muuta vormindamistingimuste järjekorda (nupud) – viimase muutmine võib osutuda vajalikuks, kui reeglite kirjapanekul on meelest läinud see, et Excel täidab vormindamise reegleid alt ülespoole, ehk järjekorras eespool (kõrgemal) paiknevad reeglid kirjutavad allpool olevad üle.

Tulemus peale kolme reegli rakendamist:

Place	id_od	Feed	LWDbst1.4	WCW	CCW	dress.%	pH45min	temp45mi	pH24h	temp24h	BackFat1	BackFat2	BackFat3	BackFat4	Meatpct	Moisture	Protein	Fat	Ash
t-test			0,057	0,000	0,000	0,033	0,013	0,141	0,009	0,000	0,002	0,002	0,000	0,023	0,405	0,513	0,015	0,443	0,161

2. Värvige oranžiks kõik tugevad seosed ($|r| \geq 0,7$) ja kollaseks kõik keskmise tugevusega seosed ($|r| \geq 0,3$), va peadiagonaalil paiknevad ühtesid sisaldavad lahtrid.

Kuigi seda ülesannet saab lahendada ka eelnevalt kirjeldatud viisil, määrates igale piirväärtusele oma vormingu (kokku viis tingimust; miks viis?), on tegelikult kiirem ja lihtsam moodus anda vormindamistingimused ette valemiga.

Näiteks antud juhul on vaja, et oranžiks värvitaks lahtrid, mis on kas 0,7-st suuremad või -0,7-st väiksemad ja mis ei võrdu ühega (neid peadiagonaalil paiknevaid ühtesid pole mõtet esile tuua, kuna need ei kujuta enesest informatiivseid väärtusi). Kaks esimest tingimust saab kokku võtta kontrollides, kas korrelatsioonikordaja absoluutväärtus on 0,7-st suurem (või võrdne). Kõik tingimused kokku saab ette anda valemiga

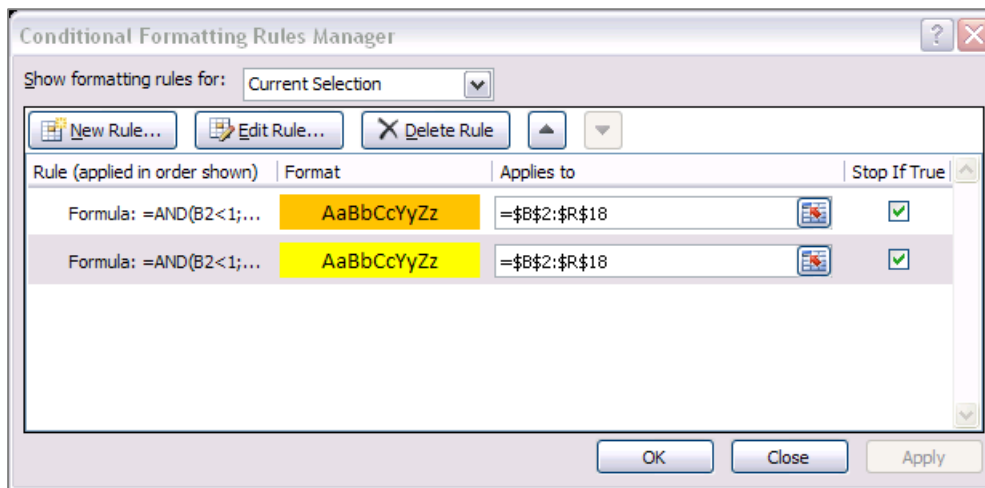
$$=AND(B2<1;ABS(B2)>=0,7)$$

- Nagu ikka, peab valem Excelis algama võrdusmärgiga (üksnes siis tõlgendab Excel järgnevat käsuna);
- funktsiooni AND argumentidena määratud tingimused peavad vormingu kinnitamiseks olema kõik (antud juhul kaks tk) rahuldatud;
- funktsioon ABS leiab absoluutväärtuse;
- lahter, millele funktsiooni on rakendatud, peab olema selekteeritud lahtribloki vasak ülemine lahter – Excel alustab tingimuse täidetuse kontrolli just nimelt sealt ja järgnevate lahtrite juurde edasi (vasakule või alla) liikudes muudab vastavalt ka valemis sisalduvat lahtriaadressi (st käitub analoogselt töölehele sisestatud valemite kopeerimisega).

The image shows the 'Conditional Formatting Rules Manager' dialog box with the 'New Rule...' button circled in red. Below it, the 'New Formatting Rule' dialog box is open, showing the 'Use a formula to determine which cells to format' option selected. The formula '=AND(B2<1;ABS(B2)>=0,7)' is entered in the 'Format values where this formula is true:' field. The preview shows a yellow cell with the text 'AaBbCcYyZz'. The 'Apply' button is highlighted in the 'Conditional Formatting' ribbon.

Analoogselt lisage reegel ka keskmise tugevusega seoste ($|r| \geq 0,3$) kollaseks värvimise tarvis.

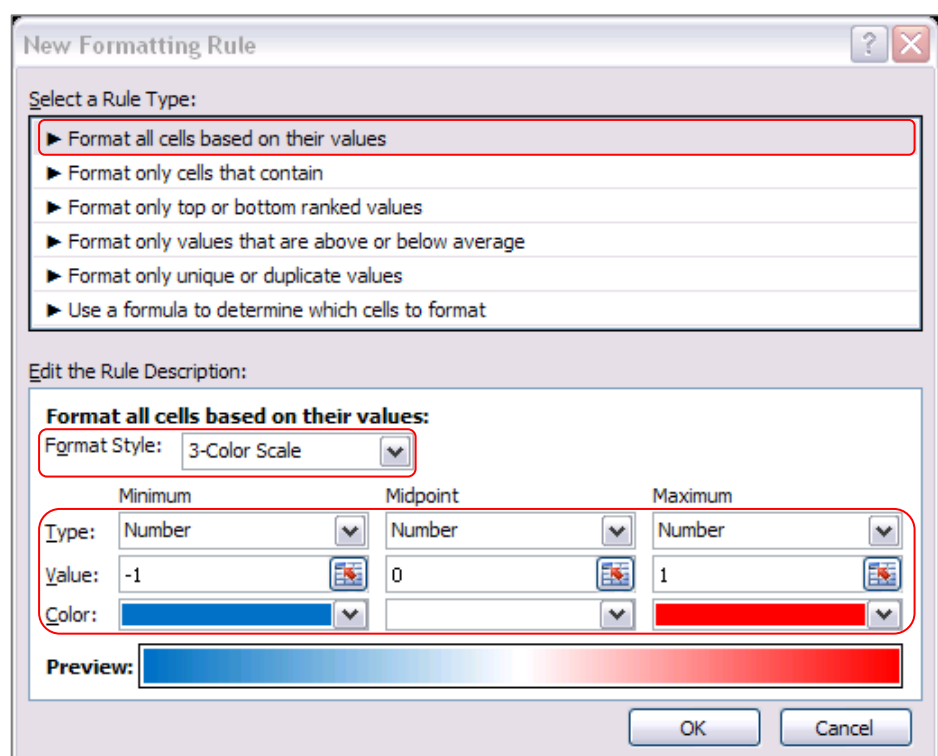
Kontrollige klikkides nupul *Apply*, kas rakendatud reeglid ikka vormindavad tabeli nii nagu soovitud. Kui mitte, tehke parandused (näiteks muutke reeglite järjekorda).



Tulemus:

	LWDbs1.4	WCW	CCW	dress.%	pH45min	temp45mi	pH24h	temp24h	BackFat1	BackFat2	BackFat3	BackFat4	Meatpot	Moisture	Protein	Fat	Ash
LWDbs1.4	1																
WCW	0,49774	1															
CCW	0,47436	0,99758	1														
dress.%	0,36691	0,43463	0,42297	1													
pH45min	0,0232	-0,2047	-0,2034	0,2562	1												
temp45m	0,16458	-0,2032	-0,232	0,02224	0,22356	1											
pH24h	-0,0972	0,19051	0,2215	-0,2581	-0,3232	-0,4516	1										
temp24h	0,11233	0,60954	0,60769	-0,0282	-0,1737	0,09011	0,15798	1									
BackFat1	0,06828	0,24259	0,25294	-0,0603	-0,0357	-0,1871	0,16207	0,32969	1								
BackFat2	-0,0176	0,27119	0,2783	0,05083	-0,175	-0,3496	0,16825	0,26409	0,5544	1							
BackFat3	0,12636	0,42679	0,43586	0,23051	-0,386	-0,6052	0,42671	0,14223	0,29809	0,50115	1						
BackFat4	-0,1431	-0,1389	-0,144	-0,1896	0,17727	0,49417	-0,3471	0,19688	0,06107	-0,0661	-0,6808	1					
Meatpot	0,11632	0,13705	0,13505	0,0397	0,00412	0,19111	0,00366	0,20179	-0,0206	0,09189	-0,0814	0,19111	1				
Moisture	-0,2646	-0,0019	0,0132	-0,0938	-0,1544	-0,3106	0,26209	-0,0783	-0,0516	0,09977	0,31662	-0,3192	0,00894	1			
Protein	-0,2313	-0,1027	-0,0954	-0,1239	0,1307	-0,118	-0,1136	-0,1256	-0,1632	-0,1053	-0,2348	0,15449	-0,0283	-0,0301	1		
Fat	0,19551	0,03041	0,02521	0,20862	0,04101	-0,0223	-0,0999	-0,1622	-0,0714	-0,1082	-0,0436	-0,0402	0,03845	-0,612	-0,2407	1	
Ash	-0,0904	-0,0636	-0,0573	-0,3519	0,08694	0,20251	-0,0339	0,1338	-0,151	-0,0662	-0,3523	0,32762	-0,0725	0,19843	0,1961	-0,4122	1

Tehke korrelatsioonikordajate tabelist (vaid väärtustest) koopia ja proovige seal ka teisi tingimusvormindamise variante – näiteks kasutage sujuvat vormingut, kus korrelatsioonikordaja väärtusele -1 vastab sinist, väärtusele 0 valget ja väärtusele 1 punast värvi lahter:



Tulemus:

	LWDbs1.4	WCW	CCW	dress.%	pH45min	temp45mi	pH24h	temp24h	BackFat1	BackFat2	BackFat3	BackFat4	Meatpct	Moisture	Protein	Fat	Ash
LWDbs1.4	1																
WCW	0,49774	1															
CCW	0,47436	0,99758	1														
dress.%	0,36691	0,43463	0,42297	1													
pH45min	0,0232	-0,20469	-0,20337	0,2562	1												
temp45mi	0,16458	-0,20319	-0,23201	0,02224	0,22356	1											
pH24h	-0,09716	0,19051	0,2215	-0,25809	-0,32317	-0,45163	1										
temp24h	0,11233	0,60954	0,60769	-0,02816	-0,17374	0,09011	0,15798	1									
BackFat1	0,06828	0,24259	0,25294	-0,06027	-0,03566	-0,18708	0,16207	0,32969	1								
BackFat2	-0,01757	0,27119	0,2783	0,05083	-0,17502	-0,3496	0,16825	0,26409	0,5544	1							
BackFat3	0,12636	0,42679	0,43586	0,23051	-0,38602	-0,60516	0,42671	0,14223	0,29809	0,50115	1						
BackFat4	-0,14313	-0,13893	-0,14399	-0,18962	0,17727	0,49417	-0,34705	0,19688	0,06107	-0,06605	-0,68078	1					
Meatpct	0,11632	0,13705	0,13505	0,0397	0,00412	0,19111	0,00366	0,20179	-0,02062	0,09189	-0,08144	0,19111	1				
Moisture	-0,26463	-0,00191	0,0132	-0,09382	-0,1544	-0,31062	0,26209	-0,07831	-0,05164	0,09977	0,31662	-0,31924	0,00894	1			
Protein	-0,23131	-0,10269	-0,09541	-0,12385	0,1307	-0,118	-0,11356	-0,12563	-0,16324	-0,1053	-0,2348	0,15449	-0,02828	-0,03014	1		
Fat	0,19551	0,03041	0,02521	0,20862	0,04101	-0,02228	-0,09991	-0,16217	-0,07145	-0,10824	-0,04357	-0,04024	0,03845	-0,61196	-0,2407	1	
Ash	-0,0904	-0,0636	-0,05729	-0,35189	0,08694	0,20251	-0,03388	0,1338	-0,15103	-0,06618	-0,35231	0,32762	-0,07248	0,19843	0,1961	-0,41217	1

3. Korrelatsioonikordajate statistilise olulisuse testimiseks Excelis sisseehitatud vahendeid ei ole, siiski on p-väärtused leitavad mõistes nende olemust ja teades arvutusvalemit.

Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

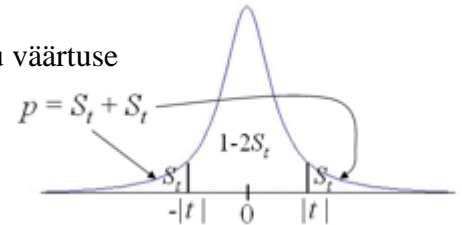
Testimaks korrelatsioonikordaja erinevust nullist (st testimaks seose statistilist olulisust) MS Excelis, tuleb esmalt arvutada teststatistiku (mis on nullhüpoteesi kehtides t-jaotusega) väärtus valemist

$$t = r\sqrt{n-2} / \sqrt{1-r^2} \sim t_{n-2},$$

H_0

suurus r selles valemis on arvatud korrelatsioonikordaja väärtus ja n on vaatluspaaride arv (ehk nende andmebaasi ridade arv, mille puhul olid mõlema tunnuse väärtused teada – puuduvate väärtusteta andmestiku puhul on siis tegu andmestiku suurusega).

Seose statistilise olulisuse üle otsustamiseks vajalik olulisuse tõenäosus p kujutab enesest leitud teststatistiku väärtuse kohalt ära lõigatud t-jaotuse sabade osakaalu (joonisel pindalade S_t summa).



Excel 2010-s on p-väärtus leitav funktsiooniga

T.DIST.2T(ABS(t);n-2),

Exceli varasemates versioonides aga valemiga TDIST(ABS(t);n-2;2).

Soovides arvutada p-väärtuseid kõigile korrelatsioonimaatriksis sisalduvatele korrelatsioonikordajatele on mõistlik viia arvutused läbi analoogselt tabelis.

- Selleks tuleb teha korrelatsioonikordajate tabelist (väärtustest) koopia ja kustutada ära tabeli sisu.

- Aga, kasutades välja arvutatatud p-väärtuseid võib vormindada ka hoopis korrelatsioonikordajate tabeli. Selleks tehke veelkord koopia korrelatsioonikordajate tabelist (väärtustest), võtke kopeeritud tabeli sisu blokki ja rakendage Exceli tingimusvormindamist, määrates lahtri vormingu vastavalt p-väärtuste tabelis samal kohal paiknevale arvule.

New Formatting Rule

Select a Rule Type:

- ▶ Format all cells based on their values
- ▶ Format only cells that contain
- ▶ Format only top or bottom ranked values
- ▶ Format only values that are above or below average
- ▶ Format only unique or duplicate values
- ▶ Use a formula to determine which cells to format

Edit the Rule Description:

Format values where this formula is true:

=AND(B42<0,001;B42>0)

Preview: AaBbCcYyZz

OK Cancel

Lahtri address p-väärtuste tabelist

Analoogselt tuleb defineerida korrelatsioonikordajate vormingud ka $p < 0,01$ ja $p < 0,05$ tarvis.

Tulemus:

r	LWDbs1.4	WCW	CCW	dress.%	pH45min	temp45mi	pH24h	temp24h	BackFat1	BackFat2	BackFat3	BackFat4	Meatpot	Moisture	Protein	Fat	Ash
LWDbs1.4	1																
WCW	0,497737	1															
CCW	0,474365	0,997576	1														
dress.%	0,366909	0,434626	0,422971	1													
pH45min	0,023202	-0,204694	-0,20337	0,2562	1												
temp45mi	0,164584	-0,203199	-0,232015	0,02224	0,223562	1											
pH24h	-0,09716	0,190514	0,221501	-0,25809	-0,32317	-0,45183	1										
temp24h	0,112325	0,609537	0,607693	-0,02818	-0,17374	0,09011	0,157983	1									
BackFat1	0,068278	0,242589	0,252939	-0,06027	-0,03586	-0,18708	0,162072	0,329688	1								
BackFat2	-0,01757	0,271189	0,278301	0,050826	-0,17502	-0,3496	0,168249	0,26409	0,554396	1							
BackFat3	0,126364	0,426795	0,435861	0,230514	-0,38802	-0,80516	0,426712	0,142231	0,296088	0,501145	1						
BackFat4	-0,14313	-0,13893	-0,14399	-0,18962	0,177273	0,494174	-0,34705	0,196878	0,06107	-0,06605	-0,58078	1					
Meatpot	0,116321	0,137052	0,135053	0,0397	0,004117	0,19111	0,003658	0,201787	-0,02062	0,091886	-0,08144	0,191111	1				
Moisture	-0,26463	-0,00191	0,013199	-0,09382	-0,1544	-0,31062	0,262094	-0,07831	-0,05164	0,099789	0,316617	-0,31924	0,008941	1			
Protein	-0,23131	-0,10269	-0,09541	-0,12385	0,130702	-0,118	-0,11356	-0,12563	-0,16324	-0,1053	-0,2346	0,154494	-0,02828	-0,03014	1		
Fat	0,195513	0,030413	0,025211	0,208616	0,041013	-0,02228	-0,09991	-0,16217	-0,07145	-0,10824	-0,04357	-0,04024	0,038456	-0,01196	-0,2407	1	
Ash	-0,0904	-0,0636	-0,05729	-0,35189	0,086938	0,202512	-0,03388	0,133796	-0,15103	-0,06618	-0,35231	0,327616	-0,07248	0,198433	0,196104	-0,41217	1

Need lahtrid on vormindatud lihtsalt nupu abil.

4. Lõpetuseks uurige, kas näiteks tunnuste 'Temp 45min' ja 'Temp 24h' vaheline seos sõltub sigade pidamiskeskonnast – leidke nimetatud tunnuste vahelised lineaarsed korrelatsioonikordajad eraldi nii tavapärasel kui ka külmaõhus peetud sigadel ning illustreerige seost hajuvusdiagrammiga, kus erinevatele pidamistingimustele vastavad väärtused on tähistatud erinevalt (lisaks võite seoste erinevuse selgemaks esile toomiseks lisada punktiparve regressioonisirged).

Eesmärk:

