

Phd-kursus i Basal Statistik, Opgaver til 2. uge

Opgave 1: Sædkvalitet

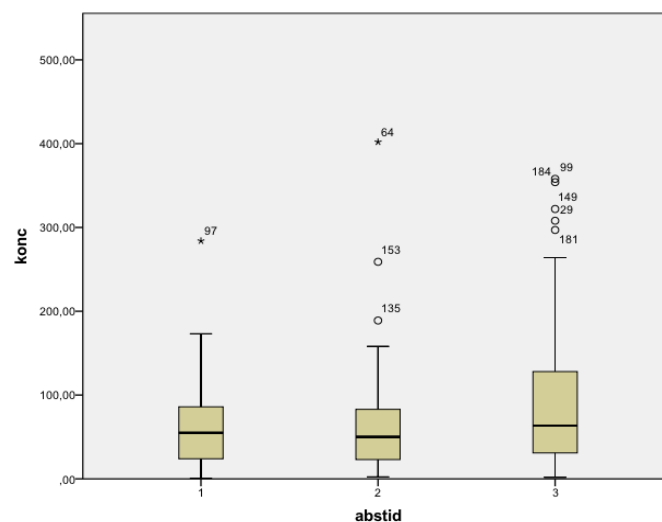
Filen `oeko.sav` på hjemmesiden indeholder datamateriale til belysning af forskellen i sædkvalitet mellem SAS-ansatte og mænd, der lever økologisk (i det følgende ofte blot omtalt som økologer). Variablene er (i den nævnte rækkefølge):

- `sas_ansat`: ansat i SAS (ja/nej)
- `abstid`: abstinensetid (1: kort, 2: medium, 3: lang)
(et mål for længden af seksuel afholdenhed)
- `konc`: sædkoncentrationen (mill/ml)

Formålet med opgaven er at undersøge, om der er forskel på de to populationsgruppers sædkoncentrationsniveau.

1. Lav en passende illustration af data.

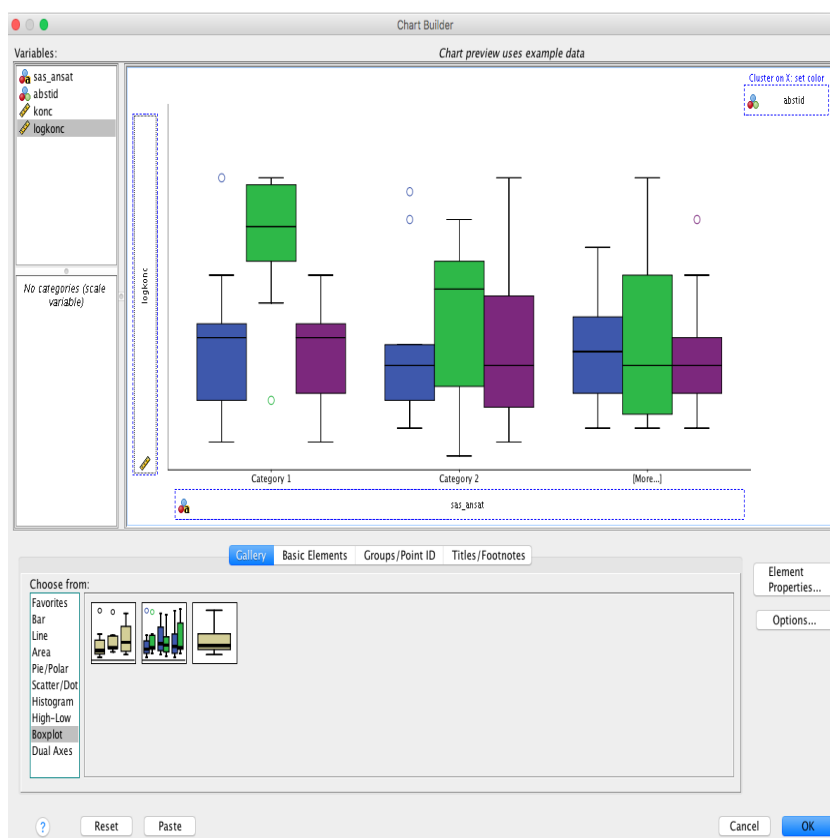
En umiddelbar optegning af sædkoncentration mod abstinensetid kunne være et box-plot:



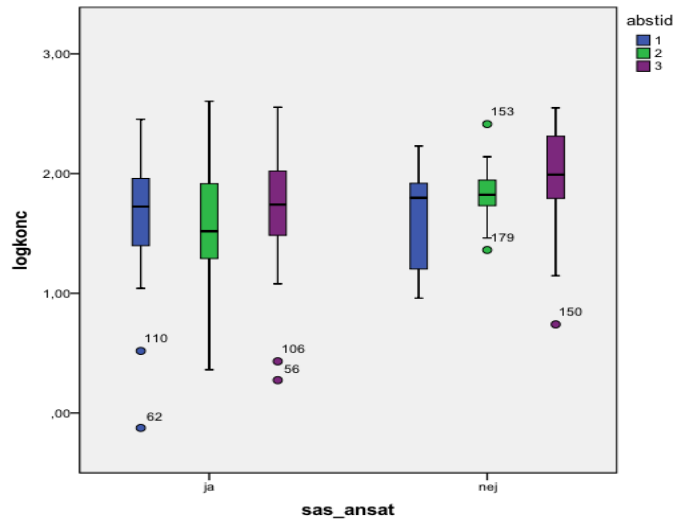
Heraf ses, at sædkoncentration næppe er normalfordelt inden for hver gruppe, og vi laver derfor en logaritmetransformation med **Compute Variable** (her er brugt $\log_{10}$).

```
COMPUTE logkonc=LG10(konc) .
EXECUTE.
```

Samtidig skiller vi observationerne ud i 2 x 3 grupper, så vi bedre kan se, hvad der er hvad ved at vælge **Clustered Boxplot**:



Figuren ser nu således ud:

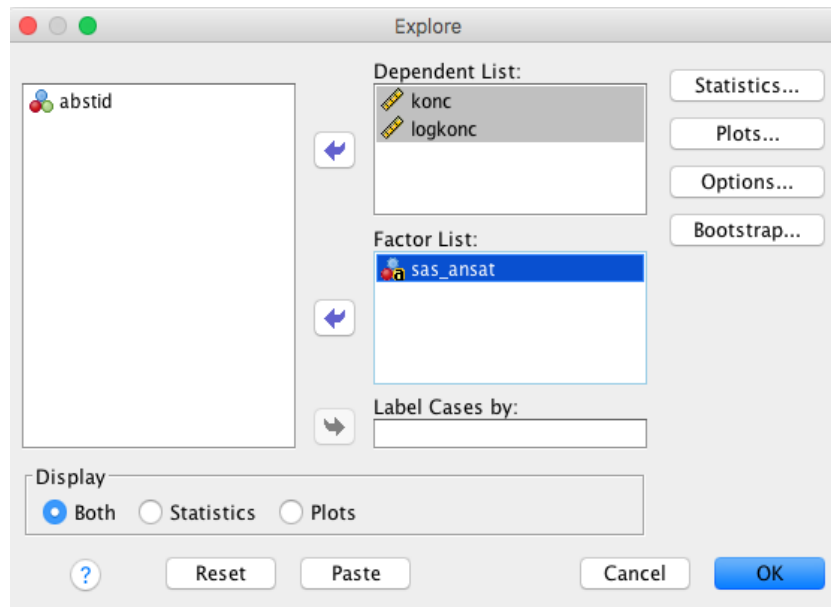


På denne skala ser både normalfordelingsantagelse og varianshomogenitet rimelig fornuftig ud, omend ikke perfekt.

2. Vi skal nu kvantificere niveauet af sædkoncentration for de to grupper af mænd og sammenligne disse niveauer, i første omgang uden at tage hensyn til abstinensiden. Overvej, om der skal logaritmetransformeres, når I svarer på nedenstående spørgsmål:

- (a) *Giv et estimat for niveauet af sædkoncentrationen for hver af de to grupper af mænd. Husk et 95% konfidensinterval.*

Som estimat for niveauet vil vi jo umiddelbart anvende gennemsnittet, men da fordelingen er skæv, vil det nok være mere passende at benytte medianen, eller at transformere til logaritmisk skala. Vi gør det hele med **Explore**:



De estimerede middelværdier (med tilhørende 95% konfidensinterval) ses at være

Data	Gruppe	gennemsnit	SEM	95% konfidensinterval (CI)
utransformeret	SAS	71.28	6.10	(59.22, 83.35)
	øko	100.95	11.94	(76.99, 124.92)
log10-transformeret	SAS	1.649	0.0406	(1.568, 1.729)
	øko	1.835	0.0577	(1.719, 1.950)
tilbagetransformeret	SAS	44.56	-	(36.98, 53.58)
	øko	68.39	-	(52.36, 89.13)

Til sammenligning kan det anføres, at medianerne i de to grupper er hhv. 48 (SAS-ansatte) og 69 (økologer), hvilket ses at passe en del bedre med de estimater, der fremkommer ved at tilbage-transforme gennemsnittene på logaritmisk skala, i forhold til de gennemsnit, der er lavet direkte på den utransformerede skala.

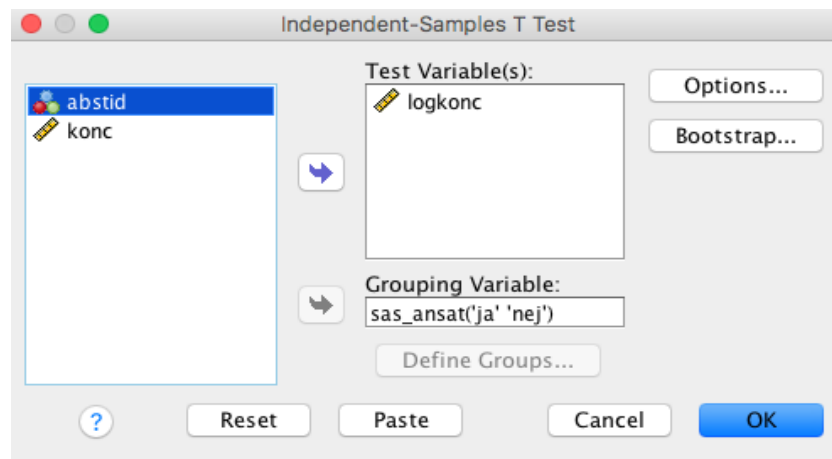
- (b) *Sammenlign de to estimater og de to tilhørende konfidensintervaller fundet ovenfor, og giv en intuitiv vurdering af, hvorvidt der er forskel på de to grupper eller ej.*

På såvel de utransformerede som de logaritmetransformerede gennemsnit ses, at økologerne har et højere niveau af sædkoncentrationen end de SAS-ansatte. Der er nogen overlap mellem de

tilhørende konfidensgrænser, men ikke ret meget for de logaritmetransformeredes vedkommende. Vi vil derfor nok forvente, at der er en faktisk forskel, men det ser vi på nedenfor.

- (c) *Foretag nu en sammenligning af de to grupper, og kvantificer forskellen i sædkoncentration for grupperne, igen med 95% konfidensinterval.*

Når vi skal sammenligne de to grupper uden hensyntagen til abstinensstiden, drejer det sig blot om et T-test. Antagelserne er bedst på log-skala, så det er den, vi benytter:



Group Statistics

	sas_ansat	N	Mean	Std. Deviation	Std. Error Mean
logkonc	ja	135	1,6486	0,47152	0,04058
	nej	53	1,8346	0,41978	0,05766

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
logkonc	Equal variances assumed	0,881	0,349	-2,506	186	0,013	-0,18594	0,07418	-0,33228 -0,03959
	Equal variances not assumed			-2,637	106,165	0,010	-0,18594	0,07051	-0,32573 -0,04615

Det ses, at SAS-ansatte har en signifikant lavere sædkoncentra-

tion end økologer ($P=0.013$). Forskellen på logaritmisk-skala er -0.186 , svarende til at de SAS-ansattes sædkoncentration udgør i gennemsnit kun udgør $10^{-0.186} = 0.65$, dvs. 65% af økologernes koncentrationsniveau.

95% sikkerhedsintervallet for denne forskel er angivet af SPSS til $(-0.332, -0.040)$, og når dette tilbagetransformeres, fås:

$$(10^{-0.332}, 10^{-0.040}) = (0.47, 0.91)$$

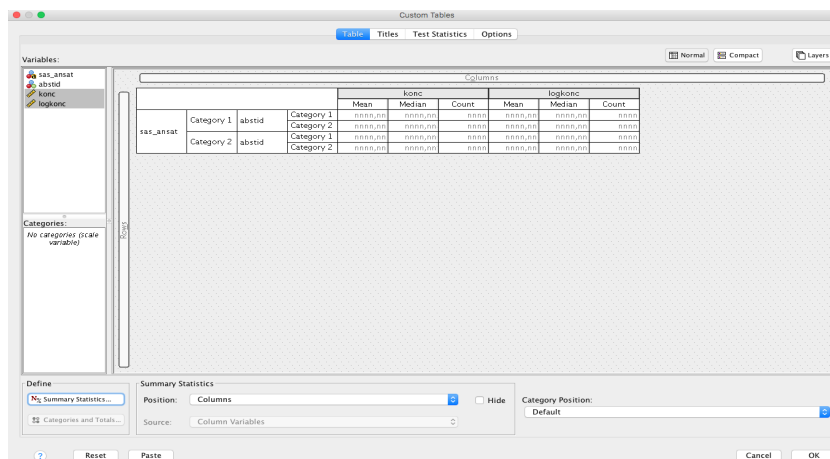
svarende til (47%; 91%) af økologernes koncentrationsniveau.

(d) *Er der signifikant forskel på de to gruppers sædkoncentration?*

Ja, på et sædvanligt 5% signifikansniveau er der forskel, da $P=0.013$.

3. Lav nu en underopdeling af de to grupper, baseret på længden af abstinensstiden og udregn passende summary statistics for de nu i alt 6 grupper.

Vi ser igen på nogle *summary statistics*, men for at minimere outputtet, tager vi denne gang kun gennemsnit og median (og antallet, som man altid bør have med) og bruger **Custom Tables**, hvor man trækker **konc** (og/eller **logkonc**), **abstid** og **sas_ansat** over til **Rows**:



Dette giver så outputtet

				konc			logkonc		
				Mean	Median	Count	Mean	Median	Count
sas_ansat	ja	abstid	1	67,04	53,00	25	1,60	1,72	25
			2	60,12	33,00	47	1,57	1,52	47
			3	81,29	55,00	63	1,73	1,74	63
	nej	abstid	1	59,01	63,00	12	1,61	1,80	12
			2	80,50	66,50	16	1,83	1,82	16
			3	134,18	98,00	25	1,94	1,99	25

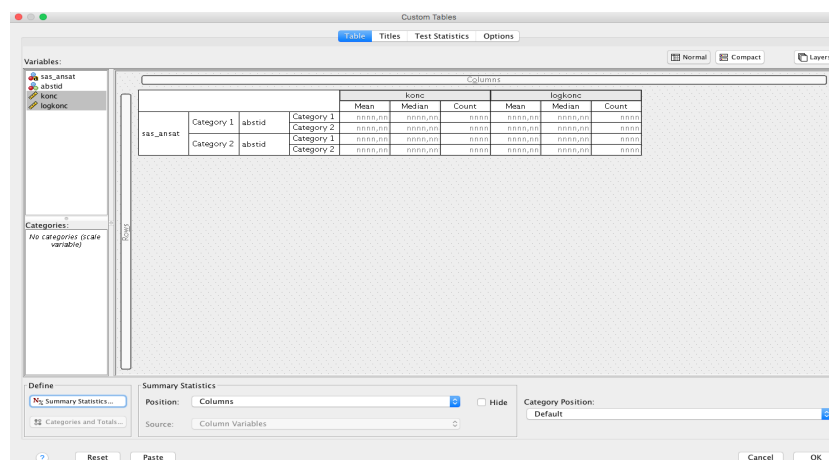
- (a) *Ser det ud som om abstinenstiden har indflydelse på sædkoncentrationen?*

Og i givet fald, ser denne indflydelse så ens ud i grupperne?

Mændene med den lange abstinenstid ses at have en noget højere sædkoncentration end dem med kort eller mellem abstinenstid. Forskellen på de to korte abstinenstider er lidt mere uklar, men noget kunne tyde på, at effekten af abstinenstid er mere udtalt for økologerne end for de SAS ansatte (altså at der *kunne være* en interaktion).

- (b) *Ser det ud som om fordelingen af abstinenstider er den samme i de to grupper?*

Dette spørgsmål vedrører slet ikke sædkoncentrationen, men udelukkende de to potentielle forklarende variable. Hvis disse to forklarende variable har relation til hinanden, altså hvis abstinenstiden i nogen grad afhænger af om man er SAS-ansat eller økolog, så kan abstinenstiden virke som en confounder for sammenligningen mellem de to grupper af mænd, således at vores estimat fra spørgsmål 2c bliver et misvisende udtryk for effekten af at leve økologisk. Vi vil lave en simpel tabel:



				konc			logkonc		
				Mean	Median	Count	Mean	Median	Count
sas_ansat	ja	abstid	1	67,04	53,00	25	1,60	1,72	25
			2	60,12	33,00	47	1,57	1,52	47
			3	81,29	55,00	63	1,73	1,74	63
	nej	abstid	1	59,01	63,00	12	1,61	1,80	12
			2	80,50	66,50	16	1,83	1,82	16
			3	134,18	98,00	25	1,94	1,99	25

I denne tabel ses antallene af mænd i hver af de 6 grupper, samt rækkeprocenterne, dvs. fordelingen af abstinensstider for hver af de to grupper mænd (SAS-ansatte og økologer). Der synes ikke at være nogen særlig forskel på disse fordelinger (man kunne lave et χ^2 -test for dette, det lærer I i næste uge).

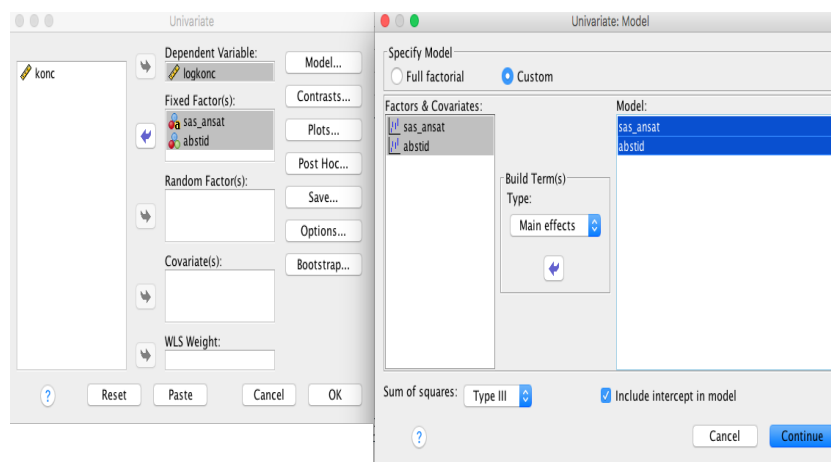
4. Benyt en variansanalysemodel til at besvare følgende:

- (a) Find et estimat for forskellen i sædkoncentration mellem de to populationer af mænd, for fastholdt værdi af abstinensstid.

Hvis abstinensstiden har en effekt på sædkoncentrationen (som det ser ud til, at den har) og hvis den også var relateret til SAS-ansat ja/nej (som det *ikke* ser ud til, at den er), så ville estimatet fra spørgsmål 2c som nævnt ikke være en rimelig sammenligning af de to grupper af mænd.

I så fald ville vi hellere sammenligne SAS-ansatte med økologer, under forudsætning af samme abstinensetid, og det er præcis hvad en (additiv) tosidet variansanalysemodel gør.

Selv om det kan virke oplagt at vælge One Way ANOVA fra Analyse/Compare Means skal vi bruge Univariate under Analyse/General Linear Model da vi har mere end en kategorisk prædiktor. Vi vælger at se parameterestimerer under Options og specificerer vores model under Model:



Outputtet bliver nu:

Between-Subjects Factors		
		N
sas_ansat	ja	135
	nej	53
abstid	1	37
	2	63
	3	88

Tests of Between-Subjects Effects					
Dependent Variable: logkonc					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2,572 ^a	3	0,857	4,184	0,007
Intercept	408,100	1	408,100	1991,860	0,000
sas_ansat	1,325	1	1,325	6,469	0,012
abstid	1,256	2	0,628	3,066	0,049
Error	37,699	184	0,205		
Total	584,267	188			
Corrected Total	40,271	187			

a. R Squared = 0,064 (Adjusted R Squared = 0,049)

Parameter Estimates						
Dependent Variable: logkonc						
Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	1,921	0,071	26,911	0,000	1,780	2,062
[sas_ansat=ja]	-0,187	0,073	-2,543	0,012	-0,332	-0,042
[sas_ansat=nej]	0 ^a	.	.	.	.	.
[abstid=1]	-0,187	0,089	-2,110	0,036	-0,362	-0,012
[abstid=2]	-0,146	0,075	-1,953	0,052	-0,293	0,002
[abstid=3]	0 ^a	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Begge kovariater ses at være signifikante, abstinenstiden dog kun lige akkurat ($P=4.9\%$). For **abstid** ses de to laveste abstinenstider at ligge nogenlunde på samme niveau, mens sædkoncentrationen er højere for mænd med lang abstinenstid, ganske som vi konkluderede ud fra gennemsnittene ovenfor. Endvidere ses, at SAS-ansatte har en signifikant lavere sædkoncentration end økologer *med samme abstinenstid*. Forskellen på logaritmisk-skala er -0.187 , svarende til at de SAS-ansattes sædkoncentration udgør i gennemsnit kun udgør $10^{-0.187} = 0.65$, altså 65% af økologernes koncentrationsniveau.

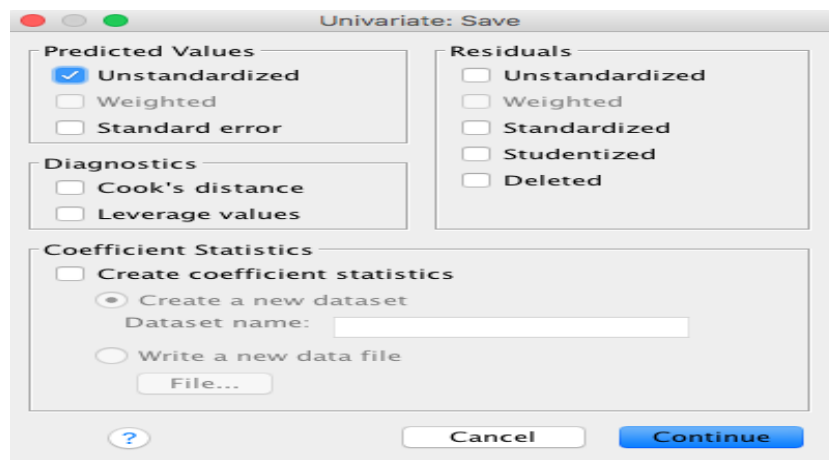
95% sikkerhedsintervallet for denne forskel er angivet til $(-0.332, -0.042)$, og når dette tilbagetransformeres, fås:

$$(10^{-0.332}, 10^{-0.042}) = (0.47, 0.91)$$

svarende til (47%; 91%) af økologernes koncentrationsniveau.

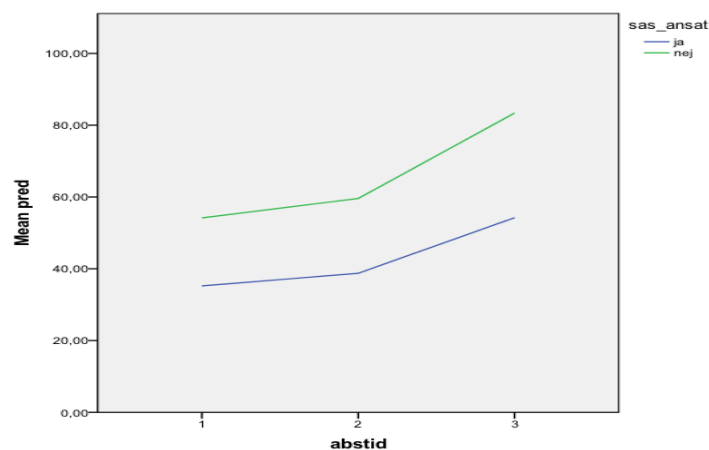
altså (med den valgte nøjagtighed) fuldstændig det samme som det ukorrigerede estimat fra spørgsmål 2c.

De tilhørende predikterede værdier for sædkoncentrationen ses i figuren nedenfor. De er dannet ved at vælge at gemme predicted values under **Save** :



Dette opretter en ny variabel **PRE_1** i datasettet. Denne kan man tilbagetransformere til den oprindelige skala med **Compute Variable** og herefter plotte dem med et multiple line plot:

```
COMPUTE pred=10 ** PRE_1.  
EXECUTE.
```



Bemærk, at de predikterede værdier er tilbagetransformeret til den oprindelige skala, og på denne skala er der *ikke* additivitet. Når

effekterne er additive på logaritmisk skala, er de multiplikative på den oprindelige skala.

Hvis man blot vil se de predikterede værdier på log-skala, kan man naturligvis lade være med at tilbagetransformere, men man kan da også benytte det automatiske plot i **General Linear Model**, hvor man går ind i **Plots**, sætter **abstid** i **Horizontal axis** og **sas_ansat** i **Separate Lines**.

Modelkontrollen.

- *Varianshomogenitet?*

Vi kan checke antagelsen om ens varians i alle 6 grupper ved at bruge Levenes test, som fås ved at vælge **Homogeneity Tests** under **Options** i GLM-dialogen. Samtidig vælger vi **Residual plot**:

og vi får så outputtet

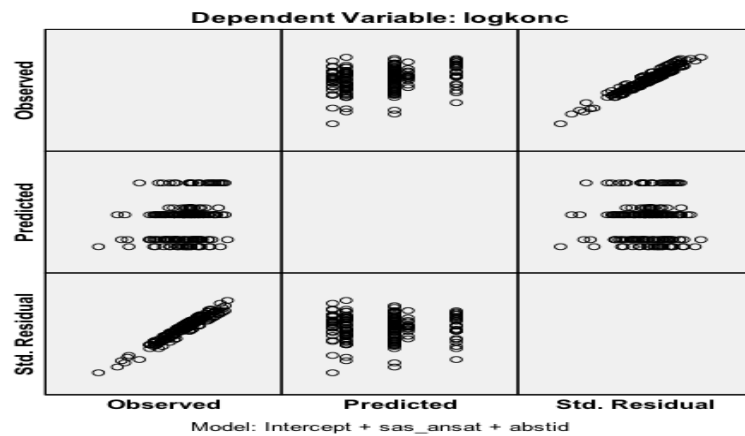
Levene's Test of Equality of Error Variances^a

Dependent Variable: logkonc

F	df1	df2	Sig.
1,450	5	182	0,209

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + sas_ansat + abstid



Antagelsen om ens varianser ser altså rimelig ud, idet Levenes test accepteres (P=21%)

Vi kunne også se på figuren. Det viser blandt andet et plot af residualer mod predikterede=forventede værdier, og det bør ikke udvise nogen form for struktur (bortset fra, at det jo ligner 6 søjler, da der kun er 6 forskellige predikterede værdier i denne model). Vi ser ingen tendens til trompetfacon eller anden form for struktur.

Tegnes histogrammer eller residual-plots (se nederst og i højre kolonne i figuren ovenfor) vil man opdage at logaritmetransformationen har bevirket en skævhed til 'den anden side', så normalfordelingsantagelsen er tvivlsom.

En bedre overensstemmelse kan opnås efter en kubikrodstransformation ($f(konc) = konc^{1/3}$). De overordnede konklusioner ændres dog ikke. Til gengæld kan parametrene i den nye model ikke direkte fortolkes (forskellene kan ikke kvantificeres på en enkel måde), så vi foretrækker at fortsætte på logaritmisk skala og glæde os over det rimeligt store datamateriale, der nedsætter behovet for en perfekt normalfordeling, så længe vi afholder os fra at lave normalområder.

- (b) *Sammenhold ovenstående estimat med det tilsvarende fra spørgsmål 2a og kommenter.*

Denne sammenligning er allerede kommenteret ovenfor. Der er ikke nævneværdig confounding at spore.

- (c) *Er der evidens for, at abstininstiden har en forskellig effekt på sædkoncentrationen i de to populationer?*

En model, der tillader effekten af abstinensid at afhænge af SAS-ansat ja/nej, er en model med et interaktionsled (vekselvirkningsled).

Vi tilføjer derfor et interaktionsled under `Model` ved at markere både `abstid` og `sas_ansat` og vælge `Interaction` i rullemenuen før de tilføjes til højre side:

Univariate: Model

Specify Model

☐ Full factorial ☒ Custom

Factors & Covariates:

sas_ansat
abstid

Build Term(s)

Type:

Interaction

Model:

sas_ansat
abstid
abstid*sas_ansat

Sum of squares: Type III

☒ Include intercept in model

Cancel Continue

som resulterer i nedenstående output:

Tests of Between-Subjects Effects

Dependent Variable: logkonc

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2,918 ^a	5	0,584	2,843	0,017
Intercept	402,445	1	402,445	1960,895	0,000
sas_ansat	0,913	1	0,913	4,448	0,036
abstid	1,251	2	0,625	3,047	0,050
sas_ansat * abstid	0,346	2	0,173	0,842	0,432
Error	37,353	182	0,205		
Total	584,267	188			
Corrected Total	40,271	187			

a. R Squared = 0,072 (Adjusted R Squared = 0,047)

Parameter Estimates

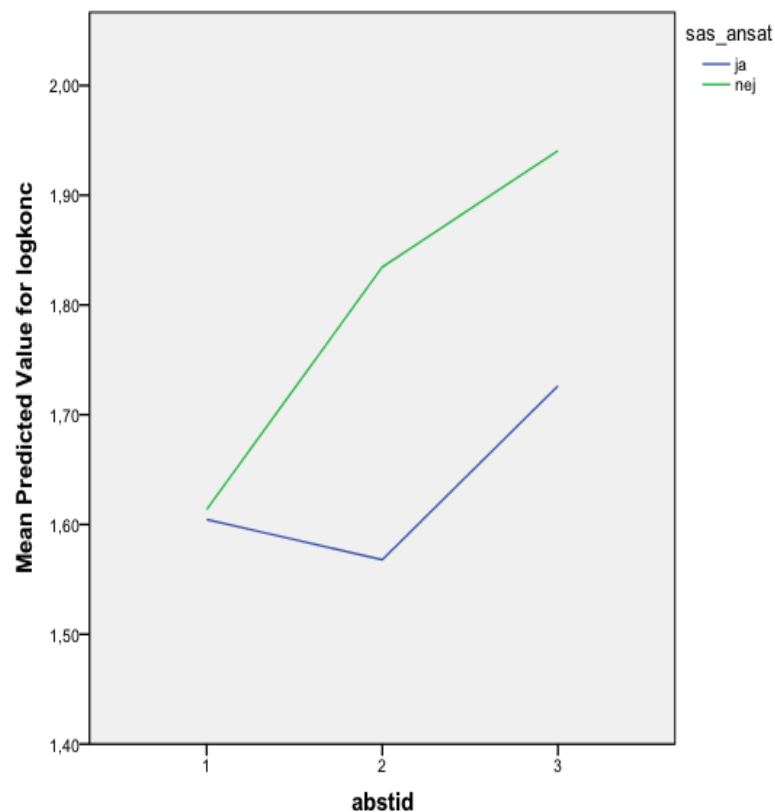
Dependent Variable: logkonc

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Intercept	1,941	0,091	21,418	0,000	1,762	2,119
[sas_ansat=ja]	-0,214	0,107	-2,001	0,047	-0,426	-0,003
[sas_ansat=nej]	0 ^a	.	.	.	.	.
[abstid=1]	-0,327	0,159	-2,055	0,041	-0,641	-0,013
[abstid=2]	-0,106	0,145	-0,731	0,466	-0,392	0,180
[abstid=3]	0 ^a	.	.	.	.	.
[sas_ansat=ja] * [abstid=1]	0,205	0,192	1,070	0,286	-0,173	0,584
[sas_ansat=ja] * [abstid=2]	-0,052	0,169	-0,309	0,757	-0,386	0,282
[sas_ansat=ja] * [abstid=3]	0 ^a	.	.	.	.	.
[sas_ansat=nej] * [abstid=1]	0 ^a	.	.	.	.	.
[sas_ansat=nej] * [abstid=2]	0 ^a	.	.	.	.	.
[sas_ansat=nej] * [abstid=3]	0 ^a	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Hypotesen om ingen vekselvirkning kan accepteres med $P = 43.2\%$.

De predikterede forløb ses i figuren nedenfor, og vi bemærker, at de ser meget anderledes ud end de tilsvarende for den additive model. Hvis vi bare skulle udtale os ud fra figuren, ville vi således gætte på, at der *var* interaktion mellem SAS-ansat og abstinentid, men figuren viser jo heller ikke usikkerhederne, og disse er altså så store, at vi ikke kan påstå at have evidens for en interaktion.



Til gengæld kan vi selvfølgelig heller ikke afvise, at der *kunne* være en interaktion, som vi bare ikke finder på grund af et for lille datamateriale.

Opgave 2: Space shuttle

Et studie involverer de 26 astronauter, der deltog på de første 8 rejser med space shuttle (Bungo et.al., 1985). På frivillig basis konsumerede 17 af disse astronauter store mængder af salt og væske inden landingen, i et forsøg på at imødegå 'space deconditioning' (salt=1). De 9 øvrige indtog intet specielt (salt=0). Filen "**space.sav**" fra hjemmesiden indeholder saltindtag og puls (slag pr. minut) før og efter flyvningen for hver af de 26 astronauter:

```
salt pre post
1 71 61
1 65 59
1 52 47
.....
.....
.....
0 52 77
0 54 80
0 52 79
```

Der er således i alt tale om 26 observationer, idet de to grupper lægges 'i forlængelse af hinanden' (kun oplysninger fra 1 person på hver linie!).

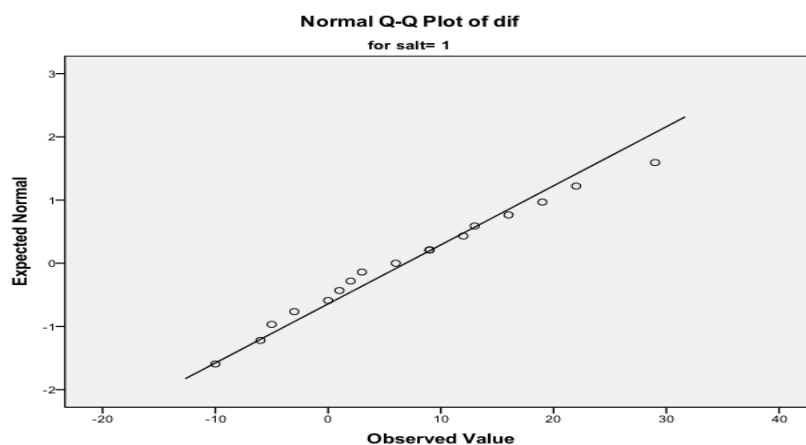
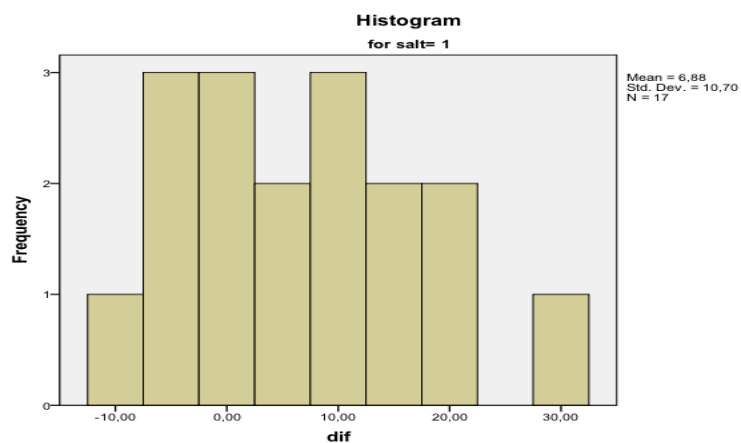
Start med, ved hjælp af **Compute Variable**, at oprette variablene **dif** og **snit**, som er henholdsvis gennemsnittet af og forskellen mellem de to målinger:

```
COMPUTE dif=post-pre.
EXECUTE.
COMPUTE snit=(post+pre)/2.
EXECUTE.
```

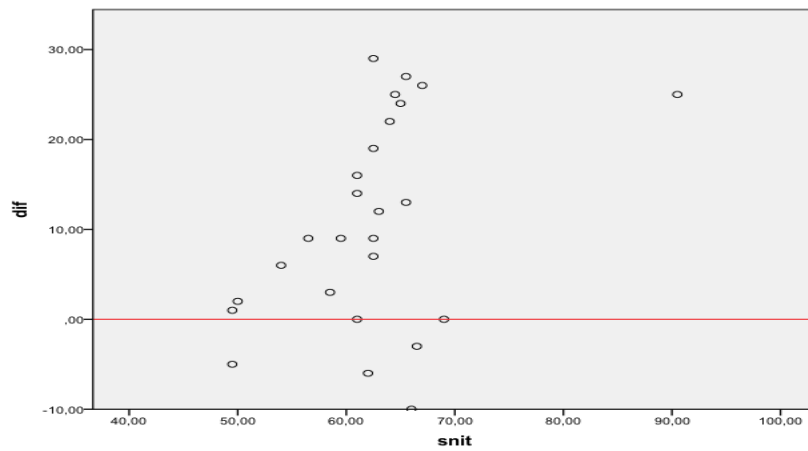
1. *Sammenlign før- og efter-målingerne i 'salt'-gruppen, og husk samtidig at lave passende illustrationer.*

Vi skal sammenligne før- og efter-målingerne i salt-gruppen. Selv om vi således i første omgang kun bliver bedt om at se på salt-gruppen, er det ligeså let at foretage sammenlignene for begge grupper på en gang, ved at benytte **Explore**, inden den relevante analyse foretages. Hvis vi skal foretage et parametrisk test (og det foretrækker vi, da det giver et

konfidensinterval), bliver der tale om et parret t-test. Forudsætningen for dette er rimelig normalitet for differenserne $\text{dif} = \text{post} - \text{pre}$, som er udregnet ovenfor. Et histogram og et fraktildiagram fås med Explore:



Vi ser her ingen særlige tegn på afvigelse fra normalfordelingen (men det betyder ikke, at vi har stor tiltro til denne antagelse, da der er tale om et ganske lille materiale). En anden vigtig forudsætning er, at differenserne er 'lige store' over hele skalaen, altså at der ikke er nogen sammenhæng mellem differenser og niveau. Dette undersøges ved et Bland-Altman plot, som simpelthen er et scatterplot af differenser mod gennemsnit:



Da dette heller ikke viser udprægede tegn på sammenhæng (eller gør det??), vil vi fortsætte med et parret t-test. Vi udfører t-testet for begge grupper på en gang ved først at bruge **Split File** og dernæst **Paired-Samples T test** i menuen **Analyse/Compare means** (husk at tilføje **post** først og dernæst **pre** for at få forskellen med korrekt fortegn, ligesom da du beregnede **dif**). Vi får resultatet:

Paired Samples Statistics

salt			Mean	N	Std. Deviation	Std. Error Mean
0	Pair 1	post	74,67	9	13,010	4,337
		pre	57,22	9	8,438	2,813
1	Pair 1	post	63,76	17	8,857	2,148
		pre	56,88	17	7,296	1,770

Paired Samples Correlations

salt			N	Correlation	Sig.
0	Pair 1	post & pre	9	0,629	0,069
1	Pair 1	post & pre	17	0,133	0,611

Paired Samples Test

			Paired Differences					t	df	Sig. (2-tailed)
			Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
Lower	Upper									
salt										
0	Pair 1	post - pre	17,444	10,113	3,371	9,671	25,218	5,175	8	0,001
1	Pair 1	post - pre	6,882	10,700	2,595	1,381	12,384	2,652	16	0,017

P-værdien for test af identiske middelværdier for puls før og efter flyvningen ses at være 0.017, hvilket er signifikant på et 5% niveau og altså viser, at der formentlig sker en pulsstigning.

Hvis vi føler os usikre på normalfordelingsantagelsen, kunne vi i stedet udføre et non-parametrisk test (Wilcoxon signed-rank test) ved at vælge **2 Related Samples** under menuen **Analyse/Nonparametric Tests/Legacy Dialogs**. Herved finder vi en P-værdi på 0.024, som understøtter konklusionen fra t-testet. Vi kunne også lave et test for normalfordelingen, men det giver ikke rigtig nogen mening på sådan et lille datamateriale.

Ranks					
salt			N	Mean Rank	Sum of Ranks
0	pre – post	Negative Ranks	8 ^a	4,50	36,00
		Positive Ranks	0 ^b	0,00	0,00
		Ties	1 ^c		
		Total	9		
1	pre – post	Negative Ranks	12 ^a	9,25	111,00
		Positive Ranks	4 ^b	6,25	25,00
		Ties	1 ^c		
		Total	17		

a. pre < post
 b. pre > post
 c. pre = post

Test Statistics ^a		
salt		pre – post
0	Z	-2,524 ^b
	Asymp. Sig. (2-tailed)	0,012
	Exact Sig. (2-tailed)	0,008
	Exact Sig. (1-tailed)	0,004
	Point Probability	0,004
1	Z	-2,225 ^b
	Asymp. Sig. (2-tailed)	0,026
	Exact Sig. (2-tailed)	0,024
	Exact Sig. (1-tailed)	0,012
	Point Probability	0,001

a. Wilcoxon Signed Ranks Test
 b. Based on positive ranks.

2. Huskede du at give en kvantificering af effekten af flyvning på pulsen i 'salt'-gruppen? Med konfidensinterval!

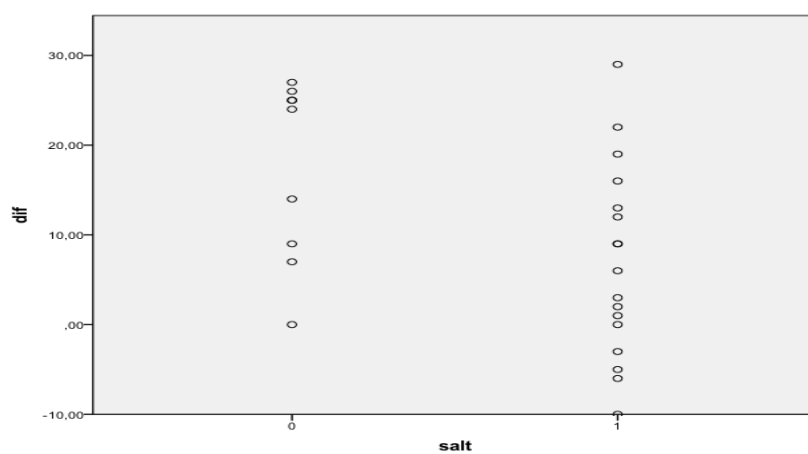
Udregning af et konfidensinterval for middelværdien af pulsstigningen fås automatisk ved at udføre t-testet. Det ses under betegnelsen 95% CL Mean, dvs. (1.3810, 12.3837).

Udfra ovenstående resultater kvantificeres stigningen i puls altså til 6.88 (med en standard error på 2.60), dvs. med 95% konfidensintervallet (1.38,12.38), altså ganske bredt. Testet gav signifikans på et 5% niveau, svarende til, at 0 ikke er inkluderet i konfidensintervallet. Vi er altså noget usikre på, hvor stor denne pulsstigning er, men den er næppe af afgørende betydning.

3. Sammenlign effekten af flyvning i de to grupper. Hvilken konklusion opnås for effekten af saltindtagelse? Husk konfidensinterval!

Vi skal nu se på en sammenligning af differenserne i de to grupper.

I kontrolgruppen har vi kun 9 personer, hvilket simpelthen er for lidt selv til grafiske illustrationer af fordelingen. Vi tillader os derfor (i hvert fald til en start) at gå ud fra, at differenserne **post-pre** er 'ligeså' normalfordelte i denne gruppe som i 'salt'-gruppen. En illustration af differenserne i de to grupper gøres bedst ved et scatterplot, da der er så få observationer:



På trods af den ikke så pæne fordeling i kontrolgruppen, fortsætter vi alligevel med at basere en sammenligning af de to gruppers differenser på et **uparret** t-test:

Group Statistics

	salt	N	Mean	Std. Deviation	Std. Error Mean
dif	0	9	17,4444	10,11325	3,37108
	1	17	6,8824	10,69978	2,59508

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
dif	Equal variances assumed	0,004	0,949	2,438	24	0,023	10,56209	4,33169	1,62193	19,50226
	Equal variances not assumed			2,483	17,260	0,024	10,56209	4,25425	1,59671	19,52747

Vi ser, at P-værdien for sammenligning af middelværdierne for differenserne i de to grupper er 0.023, svarende til, at de er signifikant forskellige på et 5% niveau. Det betyder, at de astronauter, der ikke traf nogen foranstaltninger havde en mere udtalt pulsøgning end de, der gjorde noget. Denne øgede stigning er estimeret til 10.56, med et 95% konfidensinterval på (1.62,19.50). Ikke særligt overbevisende, men alligevel ...

Det tilsvarende non-parametriske test fås (for differenser og før-målinger på en gang, se forklaring under spørgsmål 5), via **2 Unrelated Samples** in the **Analyse/Nonparametric Tests/Legacy Dialogs'** menu:

Mann-Whitney Test

Ranks				
	salt	N	Mean Rank	Sum of Ranks
dif	0	9	17,94	161,50
	1	17	11,15	189,50
	Total	26		

Test Statistics^a

	dif
Mann-Whitney U	36,500
Wilcoxon W	189,500
Z	-2,158
Asymp. Sig. (2-tailed)	0,031
Exact Sig. [2*(1-tailed Sig.)]	0,029 ^b
Exact Sig. (2-tailed)	0,030
Exact Sig. (1-tailed)	0,015
Point Probability	0,001

a. Grouping Variable: salt

b. Not corrected for ties.

Vi finder altså også en signifikans i det non-parametriske test (P-værdien er ca. 3-4%).

4. To astronauter deltog i to forskellige flyvninger og optræder altså i datamaterialet to gange. Spiller det nogen rolle?

Vi ved ikke hvilke par af observationer, der stammer fra samme astronauter, så helt konkret kan vi ikke stille noget op med vores viden. Men hvis vi havde kunnet identificere dem, ville det nok være klogest kun at benytte første flyvetur for disse. Hvis pulsøgningen er meget person-specifik skaber det nemlig problemer for antagelsen om uafhængighed mellem observationerne, at der er flere målinger for hver person.

Herudover kunne man tænke sig

- at det er nogle selekterede personer, der tager afsted flere gange
- at personer, der allerede har været afsted en gang, er blevet varigt ændret, så de anden gang adskiller sig fra de øvrige

Den konkrete betydning for analyseresultaterne er svær at sige ret meget om. Det afhænger f.eks. af om personerne er med i samme gruppe begge gange:

- Hvis de er med i samme gruppe, bliver variationen indenfor grupper for lille, og dermed kan man lettere finde en (måske ikke-eksisterende) forskel på de to grupper (type 1 fejl).
 - Hvis de er med i hver sin gruppe, bliver grupperne for ens, og vi får dermed sværere ved at se en evt. forskel (type 2 fejl).
5. *Kommenter frivilligheden i opdelingen i de to grupper og hvordan dette kunne tænkes at påvirke fortolkningen af resultaterne.*

Frivilligheden i gruppeopdelingen kan tænkes at skabe problemer, som kan gå begge veje

- Måske er det de overforsigtige/velovervejede, der tager deres forholdsregler, og hvis disse samtidig er i fysisk bedst form, kan de tænkes i forvejen at ville opleve en mindre pulsstigning
- eller måske er det dem med en kendt risiko for pulsstigning, der vælger at tage forholdsregler, og så er det sandsynligt, at forskellen på de to grupper bliver mindre udtalt.

For at få en valid sammenligning, burde grupperne have været randomiseret.

En lille indsigt i en evt. skævvridning kan fås ved at sammenligne **pre**-værdierne i de 2 grupper. Bemærk, at et t-test nu vil kræve normalitet af disse **pre**-målinger i hver gruppe og ikke som tidligere kun af differenserne. Vi finder

Mann-Whitney (Kruskal-Wallis) test: $P=0.94$
T-test, med ens varianser: $P=0.92$
T-test, med forskellige varianser: $P=0.92$

Der er altså *ikke* her nogen indikation af selektion.