Vejledende besvarelse af hjemmeopgave, forår 2018

Udleveret 12. februar, afleveres senest ved øvelserne i uge 10 (6.-9.marts)

I forbindelse med reagensglasbehandling blev 100 par randomiseret til to forskellige former for hormonstimulation. Herefter blev der udtaget et eller flere æg samt (i princippet) to ægblærevæsker, en fra hver æggestok.

Kvaliteten af de udtagne æg blev vurderet, og et antal hormoner blev målt i ægblærevæskerne, heriblandt Testosteron (nmol/l) og InhibinB (ng/ml).

For 4 kvinder lykkedes det ikke at opnå nogle æg eller ægblærer, og for visse af de 96 resterende var det (af forskellige årsager) ikke muligt at foretage alle hormonbestemmelserne.

På hjemmesiden
`http://staff.pubhealth.ku.dk/~lts/basal18_1/hjemmeopgave.html`
ligger oplysninger på 96 kvinder, med betegnelserne:

id: Løbenummer for kvinden

treat: Stimulationsmetoden (A eller B)

kvalitet: 1 for god kvalitet, 0 for mindre god

testo: Målinger af Testosteron, hhv. *left* og *right*

inhibin: Målinger af InhibinB, hhv. *left* og *right*

I de første 3 spørgsmål fokuserer vi udelukkende på **prediktion af testosteron målinger på højre side**, og **ignorerer**, at der er benyttet to forskellige stimulationsmetoder.

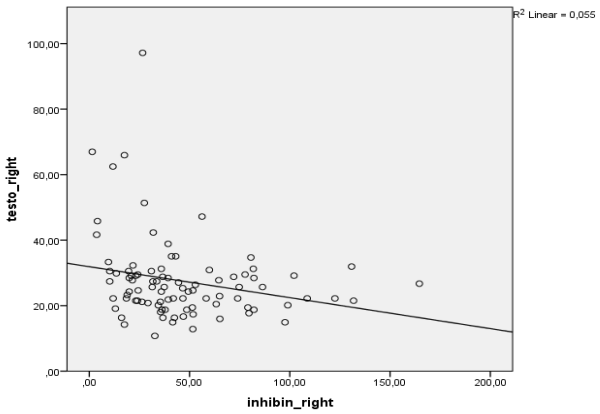
1. *Lav en figur til at illustrere sammenhængen mellem testosteron og inhibin, begge målt på højre side. Udfør på denne baggrund en lineær regressionsanalyse (***på passende skala***), med testosteron (højre side) som outcome og inhibin (højre side) som forklarende variabel. Angiv estimater for intercept og hældning med konfidensintervaller samt spredning omkring regressionslinien. Giv også, så vidt muligt, fortolkninger af disse.*

Her er det naturligt at starte med et scatterplot, med tilhørende regressionslinie (som støtte for vurderingen af rimeligheden af forudsætningerne for at foretage regressionsanalysen).

Vi går derfor ind i `Graph/Chart Builder/Scatter` og i den fremkomne boks trækkes `testo_right` over på Y-aksen, og `inhibin_right` over på X-aksen. Efterfølgende dobbeltklikkes på scatterplottet, man klikker `Add Fit Line at Total` og derefter afkrydser man i `Properties`-boksen `Linear`.

For at undgå boksen med regressionslinien hen over figuren, kan man evt. fjerne fluebenet i `Attach label to line`.

Vi får så figuren

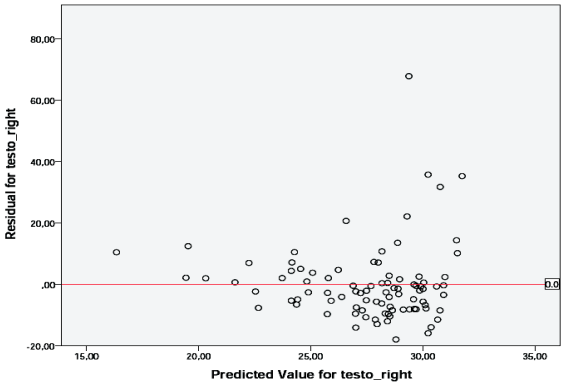


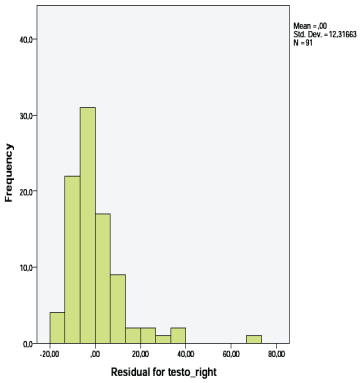
Denne figur giver ikke noget klart lineært indtryk, og residualerne fra en regressionsanalyse på disse utransformerede data ville have flere problemer

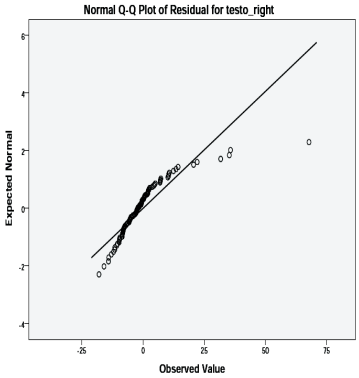
- Residualerne har ikke samme spredning for alle niveuer af outcome, men har derimod en tendens til at være større (numerisk, dvs. både store negative og store positive værdier) for de høje (forventede) værdier af outcome.

  Dette illustreres også ved det trompetagtige udseende af plottet af residualer mod forventede værdier for denne regressionsanalyse

på utransformeret skala







- Residualerne er ikke normalfordelte, men har en hale mod de store værdier. Dette illustreres bedst ved fraktildiagrammet i figuren ovenfor, der klart har den omvendte "hængekøjefacon".

I lyset af disse problemer, vil vi logaritme-transformere vores outcome, og vi holder os til 10-tals logaritmen. Der er ikke noget i de ovenstående problemer, der indikerer, at vi skal transformere vores kovariat, men alligevel kan det rent fortolkningsmæssigt være rimeligt at transformere kovariaten også, da der i begge tilfælde er tale om koncentrationer. Vi vil derfor tranformere begge med logaritmen i de efterfølgende analyser.

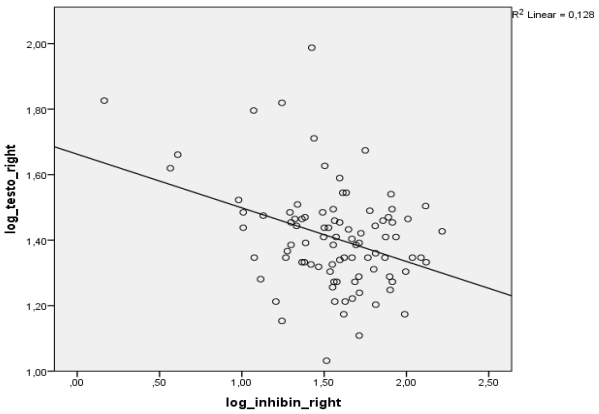
Vi definerer således (via `Transform/Compute`) alle de logaritmerede

værdier (vi får brug for dem efterfølgende):

```
log_testo_left=Lg10(testo_left)
log_testo_right=Lg10(testo_right)

log_inhibin_left=Lg10(inhibin_left)
log_inhibin_right=Lg10(inhibin_right)
```

Vi kan nu gentage figuren fra før, denne gang på logaritmisk skala, så vi går tilbage i `Graph/Chart Builder/Scatter` og udskifter de variable, hvorved vi får figuren:



Denne figur viser ingen særlige problemer med henblik på en regressionsanalyse, men vi ser nærmere på det ved modelkontrollen nedenfor. Der er dog nogle observationer svarende til de lave inhibinmålinger, der godt kan gå hen og blive indflydelsesrige. Dette kunne have været en grund til *ikke* at logaritmere inhibin....

Vi laver dernæst regressionsanalysen ved at gå ind i menuen `General Linear Model/Univariate`, hvor man sætter `log_testo_right` ind som `Dependent Variable` og `log_inhibin_right` som `Covariate(s)`, hvorefter man går ind i `Model` og sætter `log_inhibin_right` over i `Model`. Desuden går man ind i `Options` og afkrydser det ønskede, hvilket altid vil være `Parameter Estimates`, og evt også `Residual Plot`.

Vi finder så outputtet (beskåret):

Tests of Between-Subjects Effects

Dependent Variable: log_testo_right

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	,288[a]	1	,288	13,052	,001
Intercept	11,683	1	11,683	529,718	,000
log_inhibin_right	,288	1	,288	13,052	,001
Error	1,963	89	,022		
Total	182,520	91			
Corrected Total	2,251	90			

a. R Squared = ,128 (Adjusted R Squared = ,118)

Parameter Estimates

Dependent Variable: log_testo_right

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	1,662	,072	23,016	,000	1,519	1,806
log_inhibin_right	-,164	,045	-3,613	,001	-,254	-,074

Estimatet for interceptet ses at være 1.662 med tilhørende konfidensinterval $(1.519, 1.806)$, men dette er rimeligt meningsløst, idet det refererer til en inhibinværdi på 1 (fordi logaritmen til 1 er 0), og vi vil derfor ikke bruge krudt på at tilbagetransformere det.

Estimatet for hældningen er -0.164, med et konfidensinterval på $(-0.254, -0.074)$, hvilket viser den negative association mellem de to hormoner.

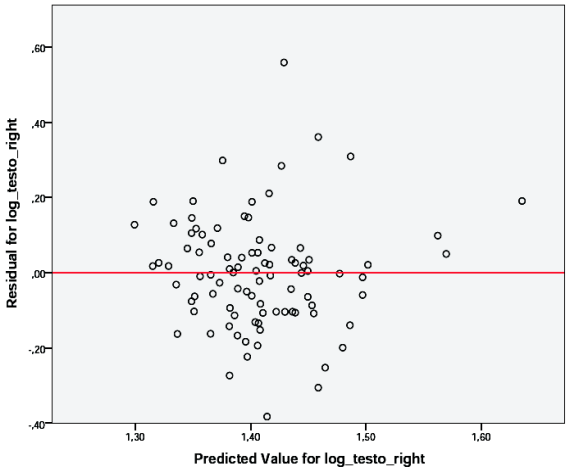
Når man tilbagetransformerer hældningen, får man effekten af at 10-doble inhibinværdien, $10^{-0.164} = 0.69$, altså at testosteron falder med 31% eller til 69% af hvad den var før. Konfidensintervallet for dette er $(10^{-0.254}, 10^{-0.074}) = (0.56, 0.84)$, svarende til et fald i testosteron på mellem 16 og 44%.

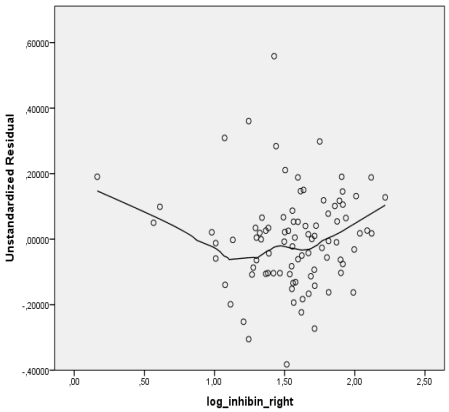
Interessant er det, at når begge variable (outcome såvel som kovariat)

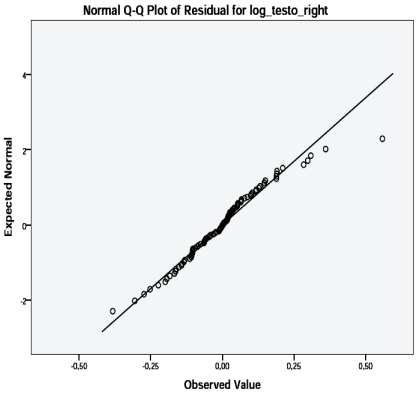
logaritmetransformeres, så får man samme hældning uanset hvilken logaritme, man benytter (blot man selvfølgelig benytter samme logaritme på begge). Det betyder, at man også direkte kan se effekten af en fordobling af inhibinværdien, nemlig en faktor $2^{-0.164} = 0.89$, på testosteron-værdien, altså et fald på ca. 11%. Konfidensintervallet for dette er $(2^{-0.254}, 2^{-0.074}) = (0.84, 0.95)$, svarende til et fald i testosteron på mellem 5 og 16%.

Estimatet for spredningen omkring regressionslinien
(`Standard Error of the Estimate`) kan findes som kvadratroden af `Mean Square Error`, altså $\sqrt{0.022}$, som med lidt flere decimaler giver 0.149, men det er ikke så let at fortolke. Det benyttes jo til at konstruere prediktionsintervaller, idet disse er givet ud fra linien ved at lægge $\pm 2 \times 0.149 = \pm 0.298$ til. Tilbagetransformerer vi, fås $10^{0.298} = 1.99$ og $10^{-0.298} = 0.504$. Det kan fortolkes som tilbagetransformerede prediktionsgrænser, der siger, at de individuelle testosteron-værdier kan være op til enten dobbelt eller ned til halvdelen af, hvad prediktionen angiver.

Vi skal også lige med et par residualplots sikre os, at vores model er rimelig



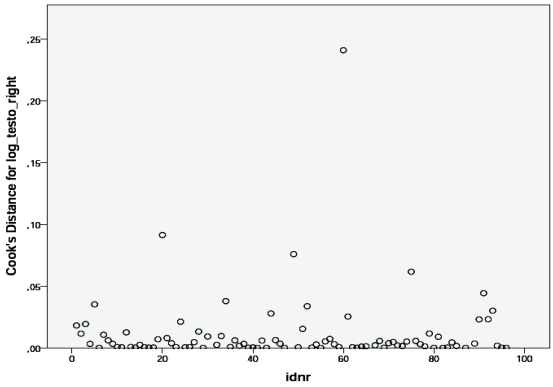




Det ses, at forudsætningerne ser fornuftige ud, idet residualerne har nogenlunde samme spredning uanset niveauet af outcome variablen, og at de er nogenlunde normalfordelte. Lineariteten ser dog ud til at halte lidt, hovedsageligt på grund af den meget lave inhibin-måling.

Der ses mindst en observation med ret stor Cook-værdi. Vi kan se nærmere på denne ved at gemme Cook-værdierne i datasættet (benyt `Save`-knappen) og bagefter se, hvor de store værdier er.

Vi kan plotte dem mod observationsnummeret:



Den rigtig store Cook-værdi (0.24) ses at tilhøre individ nr. 60, som har en dårlig æg-kvalitet og en meget lille værdi af inhibin på højre side, faktisk den mindste i datamaterialet. Det *kunne* jo være en fejl....

2. *Hvad er den forventede værdi af testosteron for en kvinde med en inhibin på 30? Og hvad er normalområdet for sådanne kvinder?*
Er det usædvanligt at se en kvinde med en testosteron-værdi på 80 og en inhibin-værdi på 30?

   For at få interceptet til at svare til en Inhibin-værdi på 30, skal vi sørge for, at vores kovariat `log_inhibin_right` netop er 0 for denne værdi. Dette opnås ved at fratrække $\log_{10}(30) = 1.477$ fra `log_inhibin_right`, altså ved at gå i `Transform/Compute` og definere `log_inhibin_right30=log_inhibin_right-Lg10(30)`. Herefter gentager vi regressionsanalysen med denne nye kovariat, hvorved vi får outputtet:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1,420	,016		88,925	,000
	log_inhibin_right30	-,164	,045	-,358	-3,613	,001

Coefficients[a]

Model		95,0% Confidence Interval for B	
		Lower Bound	Upper Bound
1	(Constant)	1,389	1,452
	log_inhibin_right30	-,254	-,074

a. Dependent Variable: log_testo_right

   Den predikterede testosteron værdi er her $10^{1.420} = 26.30$ nmol/l, og normalområdet på logaritmisk skala er $1.420 \pm 2 \times 0.149 = (1.122, 1.718)$. Da dette ikke indeholder $1.90 = log_{10}(80)$ er kombinationen usædvanlig. Vi kan også tilbagetransformere normalområdet til $10^{(1.122,1.718)} = (13.24, 52.24)$ nmol/l, som jo ikke indeholder 80.

   Bemærk i øvrigt, at det i dette spørgsmål er **vigtigt** om normalfordelingen og den konstante varians (spredning) holder, idet vi udtaler os om enkeltobservationer.

3. *Har vi nogen glæde af at kende de to inhibin målinger, hvis vi også udnytter information om testosteron niveauet på venstre side som forklarende variabel?*

   Ovenfor så vi, at højre sides inhibin målinger havde en signifikant evne til at prediktere testosteronmålingerne på højre side, omend vi kun forklarede ca. 13% af variationen. Det virker oplagt, at testosteron målingerne på venstre side skulle prediktere bedre, og vi afprøver nu dette ved at inddrage alle tre mulige kovariater i en regressionsmodel.

   Vi gå derfor tilbage til `General Linear Model/Univariate`, udskifter `log_inhibin_right30` med den oprindelige `log_inhibin_right` og indsætter desuden de to resterende kovariater: `log_testo_left` og `log_inhibin_left`.

   Vi får så

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,735[a]	,541	,524	,11151

a. Predictors: (Constant), log_inhibin_right, log_testo_left, log_inhibin_left

b. Dependent Variable: log_testo_right

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,172	3	,391	31,403	,000[b]
	Residual	,995	80	,012		
	Total	2,166	83			

a. Dependent Variable: log_testo_right

b. Predictors: (Constant), log_inhibin_right, log_testo_left, log_inhibin_left

Coefficients[a]

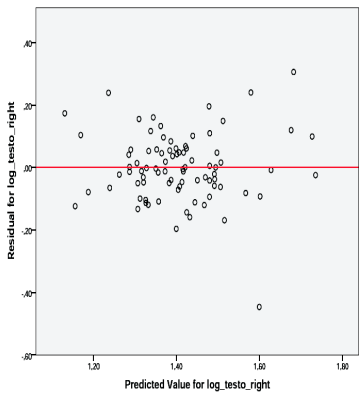
Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	,477	,160		2,986	,004
	log_testo_left	,684	,082	,723	8,309	,000
	log_inhibin_left	,056	,047	,121	1,197	,235
	log_inhibin_right	-,073	,044	-,158	-1,650	,103

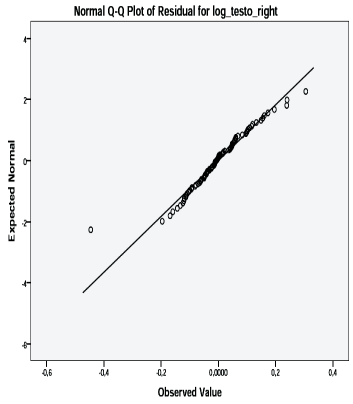
Coefficients[a]

Model		95,0% Confidence Interval for B	
		Lower Bound	Upper Bound
1	(Constant)	,159	,795
	log_testo_left	,520	,848
	log_inhibin_left	-,037	,148
	log_inhibin_right	-,161	,015

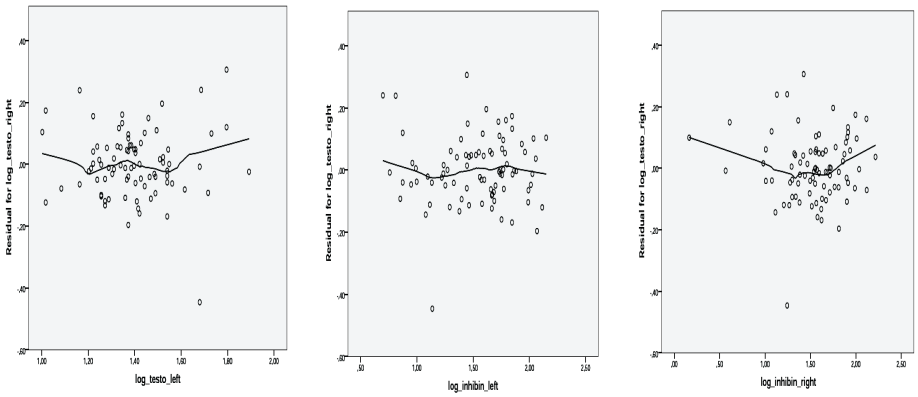
a. Dependent Variable: log_testo_right

For en ordens skyld skal vi lige overbevise os selv om, at den multiple regressionsmodel overhovedet var rimelig, dvs. vi skal udføre de sædvanlige modelkontroltegninger. Det drejer sig om residualer plottet mod forventede værdier af outcome samt fraktildiagram over residualerne.





Endvidere vil vi gerne se plots af residualer mod hver af de tre kovariater, med udglattede kurver, for at vurdere lineariteten:



Disse er ikke alle helt fine i kanten, men heller ikke så slemme, at vi ikke kan fortsætte med vores forehavende.

Vi vender derfor tilbage til det output, vi fandt fra vores analyse, og her bemærker vi, at vi i hvert fald ikke har meget glæde af at kende inhibinværdien på venstre side, så denne udelader vi, hvorved vi får

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,729[a]	,531	,520	,11128

a. Predictors: (Constant), log_inhibin_right, log_testo_left

b. Dependent Variable: log_testo_right

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,151	2	,576	46,472	,000[b]
	Residual	1,016	82	,012		
	Total	2,167	84			

a. Dependent Variable: log_testo_right

b. Predictors: (Constant), log_inhibin_right, log_testo_left

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,569	,141		4,045	,000
	log_testo_left	,649	,077	,687	8,412	,000
	log_inhibin_right	-,045	,038	-,097	-1,187	,239

Coefficients[a]

Model		95,0% Confidence Interval for B	
		Lower Bound	Upper Bound
1	(Constant)	,289	,849
	log_testo_left	,495	,802
	log_inhibin_right	-,119	,030

a. Dependent Variable: log_testo_right

Vi ser, at der heller ikke ligger signifikant information i at kende inhibin værdien på højre side.

Bemærk, at man ikke bør sammenligne R^2-værdierne for modeller, der ikke har det samme antal kovariater, da denne altid vil være mindre, når der er flere kovariater med. Så er residalspredningen, eller residualvariansen (`Residual Mean Square`) et bedre mål, og her ser vi, at modellen med udelukkende `log_testo_left` som forklarende variabel giver næsten helt den samme værdi, som hvis en eller begge inhibin-målinger er med i modellen.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,723[a]	,523	,517	,11156

a. Predictors: (Constant), log_testo_left

b. Dependent Variable: log_testo_right

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,134	1	1,134	91,086	,000[b]
	Residual	1,033	83	,012		
	Total	2,167	84			

a. Dependent Variable: log_testo_right

b. Predictors: (Constant), log_testo_left

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	,452	,101		4,490	,000
	log_testo_left	,683	,072	,723	9,544	,000

Coefficients[a]

Model		95,0% Confidence Interval for B	
		Lower Bound	Upper Bound
1	(Constant)	,252	,652
	log_testo_left	,541	,826

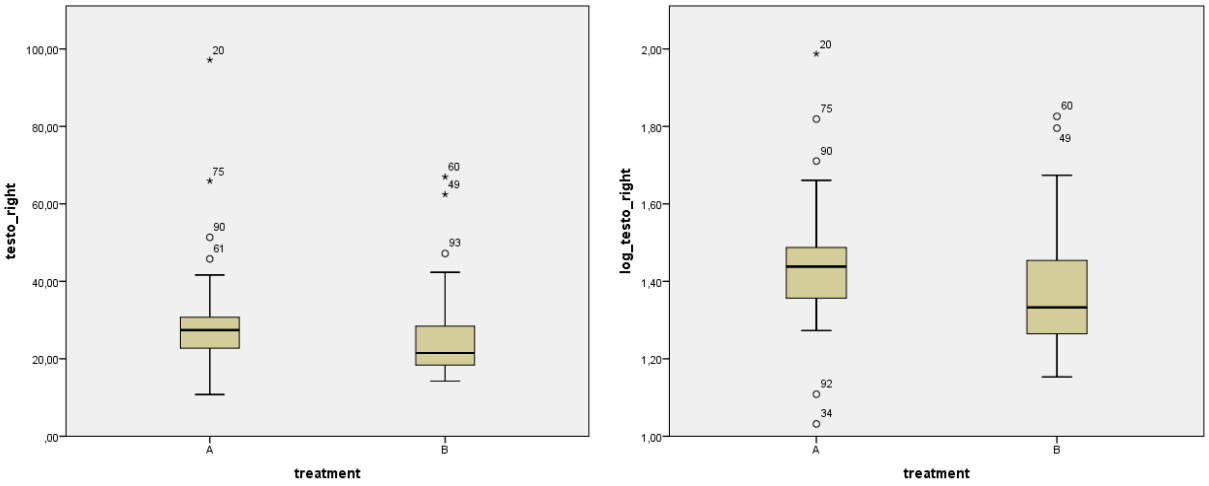
a. Dependent Variable: log_testo_right

Nu kommer et par spørgsmål, der fokuserer på
forskellen på stimulationsmetoderne:

4. *Lav en figur til illustration af forskellene på de to behandlingsgrupper forsåvidt angår testosteron på højre side. Udfør på baggrund af denne tegning en sammenligning af de to grupper* **på passende skala** *og kvantificer forskellen med konfidensinterval.*

   Vi starter med et simpelt Boxplot, som vi får ved at gå ind i `Analyze/Descriptive Statistics/Explore`, hvor vi sætter `testo_right` i `Dependent List`, `treatment` i `Factor List` samt sætter hak i `Plots`. Plottet ses til venstre i nedenstående figur.

   Man bemærker her (igen), at fordelingen i begge grupper er en anelse skæv, med en hale mod de høje værdier. Da vi er på vej mod en sammenligning af grupperne med et T-test, forsøger vi igen at transformere med logaritmen for at opnå symmetriske fordelinger (helst normalfordelinger). På figuren til højre er anvendt 10-tals logaritmer, men det spiller ingen rolle.



Ud fra den højre figur ser det fornuftigt ud at foretage et uparret T-test, idet

- Observationerne antages uafhængige (forskellige kvinder)
- Der synes at være nogenlunde samme variation i begge grupper
- Fordelingerne ser rimeligt symmetriske/normalfordelte ud

Man kan selvfølgelig supplere med fraktildiagrammer, men det skønnes ikke at være nødvendigt her, ikke mindst fordi fordelingsantagelsen ikke er voldsomt kritisk, hvis vi blot skal sammenligne middelværdier og ikke udtale os om enkeltindivider.

Vi udfører så det uparrede T-test ved at gå ind i `Analyze/Compare Means/Independent Samples T-test`, hvor vi sætter `log_testo_right` over i `Test Variable(s)` og `treat` i `Grouping Variable`.

Under `Define Groups` vælges `Group1` til A og `Group1` til B. Herved får vi outputtet:

Group Statistics

	treatment	N	Mean	Std. Deviation	Std. Error Mean
log_testo_right	A	48	1,4415	,15116	,02182
	B	43	1,3695	,15883	,02422

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
log_testo_right	Equal variances assumed	1,257	,265	2,213	89
	Equal variances not assumed			2,207	86,762

Independent Samples Test

		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
log_testo_right	Equal variances assumed	,029	,07193	,03251
	Equal variances not assumed	,030	,07193	,03260

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
log_testo_right	Equal variances assumed	,00734	,13653
	Equal variances not assumed	,00714	,13673

Her ser vi

- Der indgår i alt 91 personer i vores sammenligning, idet nogle kvinder har manglende værdier for testosteron på højre side.
- Spredningerne i de to grupper er meget ens (0.1512 hhv. 0.1588), P-værdi for test af identitet af disse er 0.265.
- Middelværdierne i de to grupper er signifikant forskellige, med P=0.029.

- Den estimerede forskel på de to behandlinger er (på logaritmisk skala) 0.0719 (stimulation A mod stimulation B), med et konfidensinterval på (0.00734, 0.13653). Tilbagetransformeres dette, fås et ratio-estimat på $10^{0.0719} = 1.18$, altså at stimulation A giver 18% højere værdier end stimulation B, med et konfidensinterval på (1.017, 1.369), dvs. svarende til fra 1.7% over til hele 36.9% over.

5. *Hvis to personer med forskellige stimulationsmetoder har opnået samme testosteron-værdi på venstre side, kan vi da forvente, at de også har nogenlunde samme værdi på højre side, eller kan afvigelserne komme op på f.eks. 50 nmol/l (eller 25%)?*

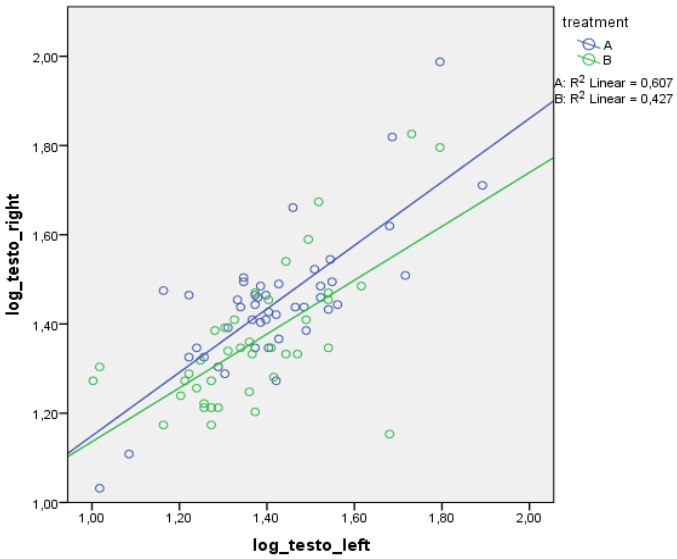
   Der er her lagt op til en kovariansanalyse, hvor vi ved sammenligning af `log_testo_right` for de to stimulationsmetoder korrigerer (betinger, justerer) for forskelle på venstre side, således at vi sammenligner værdier på højre side, *for fastholdt værdi* af værdi på venstre side.

   Først skal vi dog lige lave et plot ved at gå ind i `Graph/Chart Builder`, vælge `Scatter plot` (nr. 2 fra venstre), og i den fremkomne boks trække `log_testo_right` over på Y-aksen, `log_testo_left` over på X-aksen og `treatment` over i `Set Color`

   For at tegne linierne, dobbeltklikker man efterfølgende på grafen og klikker på ikonet `Add Fit Line at Subgroups` og derefter i `Properties`-boksen afkrydse `Linear` og klikke `Apply`.

   Man kan endvidere vælge, om man vil have liniens ligning skrevet på linien eller ej (flueben ved `Attach label to line` kan fjernes)

   Herved får vi figuren:



Selv om disse regressionslinier ikke er helt parallelle, vil vi fortsætte med en kovariansanalyse, da vi ikke på forhånd havde formodning om en interaktion her.

Vi benytter `Analyze/General Linear Model/Univariate`, hvor vi sætter `log_testo_right` over i `Dependent Variable`, `treatment` i `Fixed Factor(s)` og `log_testo_left` i `Covariate(s)`.

For at undgå at få interaktionen med i modellen, klikkes nu på `Model`, hvorefter man vælger `Custom`, markerer begge kovariater, skifter fra `Interaction` til `Main Effects` og klikker på pilen og `Continue`.

Husk også i `Options` at vælge `Parameter Estimates`

Vi får så outputtet

Tests of Between-Subjects Effects

Dependent Variable: log_testo_right

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1,198[a]	2	,599	50,705	,000
Intercept	,281	1	,281	23,759	,000
treatment	,064	1	,064	5,445	,022
log_testo_left	1,039	1	1,039	87,921	,000
Error	,969	82	,012		
Total	170,223	85			
Corrected Total	2,167	84			

a. R Squared = ,553 (Adjusted R Squared = ,542)

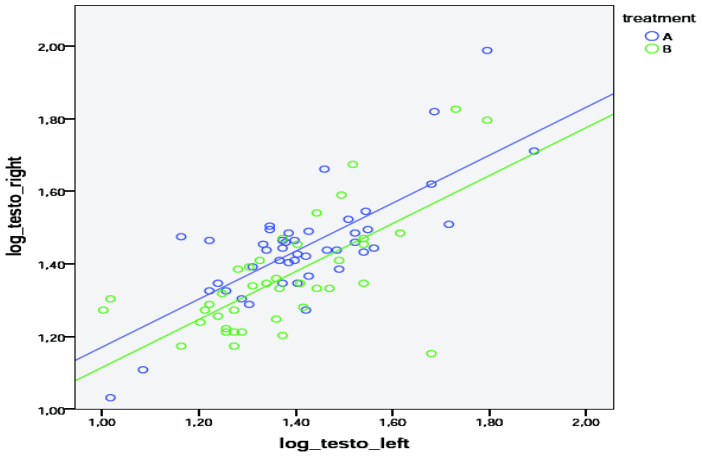
Parameter Estimates

Dependent Variable: log_testo_right

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	,455	,098	4,633	,000	,259	,650
[treatment=A]	,056	,024	2,334	,022	,008	,103
[treatment=B]	0[a]	.	.	.	.	.
log_testo_left	,660	,070	9,377	,000	,520	,801

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

Den model, vi har fittet, svarer til to parallelle linier, som skitseret nedenfor:

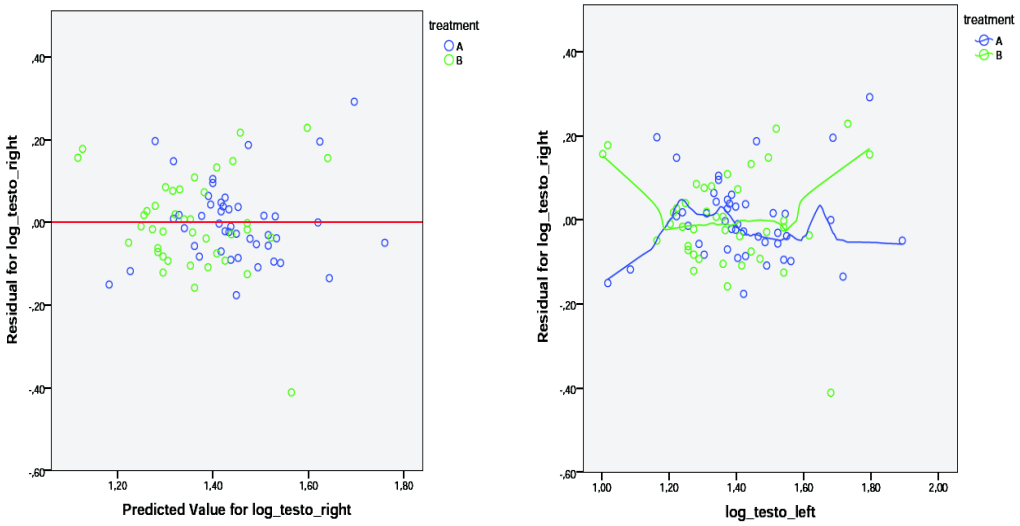


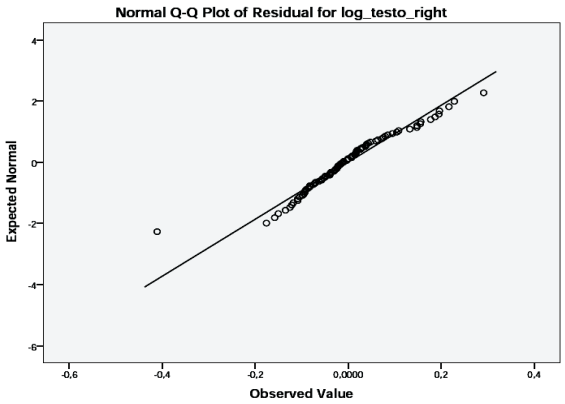
Denne figur er lidt besværlig at lave, idet man efter at have lavet selve scatterplottet, skal gå ind i menuen `Add a reference Line by Equation`, vælge farve under `Lines` (her blå for gruppe A og grøn for gruppe B, svarende til punkternes farve) og herefter under `Reference Line` skrive den relevante formel ind, dvs. for gruppe A'e vedkommende rette udtrykket `y=1*x+0` til liniens ligning `y=0,660*x+0,511`.

Herefter gentages for den anden gruppe, der har ligningen `y=0,660*x+0,455`.

Det ses af outputtet, at der stadig er signifikant forskel på de to stimulationsmetoder (P=0.022), omend forskellen er formindsket noget, fra 0.0719(0.0325) i spm. 4 til nu 0.056(0.024), på logaritmisk skala. Vi kvantificerer yderligere nedenfor.

De automatiske modelkontroltegninger er ikke så informative, men vi kan ligesom tidligere lave dem mere detaljerede ved at gemme residualer og predikterede værdier og selv tegne efterfølgende:





Vi skal nu udbygge kvantificeringen af forskellen på de to stimulationsmetoder for `log_testo_right`, for *fastholdt* værdi af `log_testo_left`, altså forskellen 0.056(0.024) på logaritmisk skala, dvs. (med en ekstra decimal): CI=(0.0082, 0.1031). Når vi tilbagetransformerer dette, får vi $10^{0.0557} = 1.137$, med CI=$(10^{0.0082}, 10^{0.1031}) = (1.019, 1.268)$. For fastholdt værdi af testosteron på venstre side, finder vi altså, at stimulationsmetode A stadig estimeres til at give 13.7% højere værdier på højre side end stimulationsmetode B, men det kan være helt op til 26.8% højere.

Da `testo_right`-værdierne ligger et sted mellem 10 og 60-70 stykker, kan forskellen **på middelværdierne** altså *ikke* komme helt op mod de 50 nmol/l, idet 26.8% af 60 kun er ca. 16.

Nu er der imidlertid spurgt til forskelle for to **personer**, og vi skal derfor over i noget med prediktionsintervaller i stedet for. Og det er ikke engang et helt almindeligt prediktionsinterval, fordi det drejer sig om en *differens* mellem to personer, ikke *en enkelt* person, som det plejer.

Vi indfører nu betegnelserne H_A, H_B, V_A, og V_B for `log_testo`, på hhv Højre og Venstre side, for stimulationsmetode A og B.

Vi har så her set på modellen med 2 parallelle regressionslinier:

$$H_A = \alpha_A + \beta V_A + \varepsilon_A$$
$$H_B = \alpha_B + \beta V_B + \varepsilon_B$$

og vi vil gerne finde prediktionsinterval for differensen $H_A - H_B$ under

forudsætning af, at $V_A = V_B$. Men så er

$$H_A - H_B = \alpha_A - \alpha_B + \varepsilon_A - \varepsilon_B$$

som har middelværdi $\alpha_A - \alpha_B$, men varians $2 \times \sigma^2$.

I prediktionsintervallet skal vi således gange `Mean Square Error` med 2, så vi finder:

$$0.056 \pm 2 \times \sqrt{2 \times 0.022} = (-0.252, 0.363)$$

og når vi tilbagetransformerer dette, får vi CI=$(10^{-0.252}, 10^{0.363}) = (0.560, 2.307)$, svarende til, at stimulationsmetode A kan være over dobbelt så høj som stimulationsmetode B, men også ned til ca. 44% lavere. Her er således rigelig plads til en forskel på 50 nmol/l.

6. *Kvantificer middelværdien af sideforskellene for testosteron, for hver stimulationsmetode for sig.*

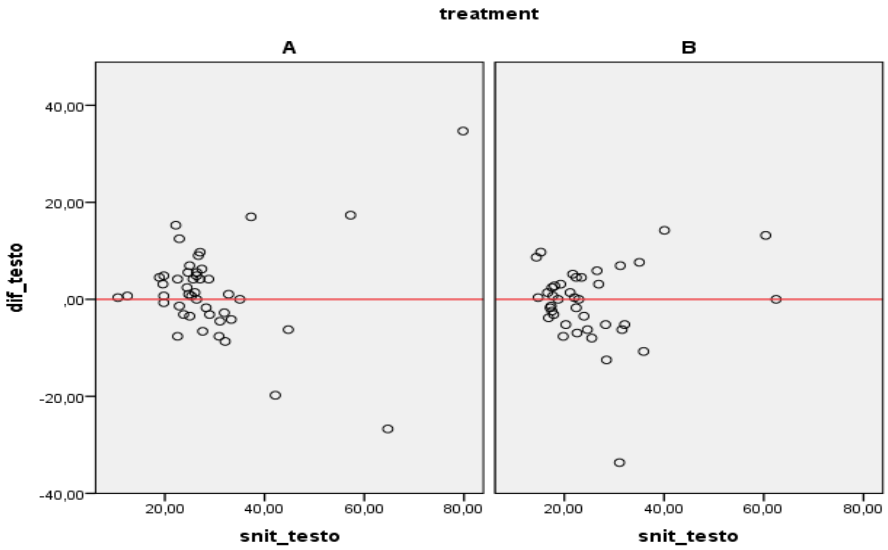
   Vi har her brug for at definere nogle nye variable, nemlig differenser og gennemsnit for testosteron på log-skala. Husk, at man skal lave differenser af logaritmer, **aldrig** logaritmer af differenser.

   I `Transform/Compute` definerer vi:

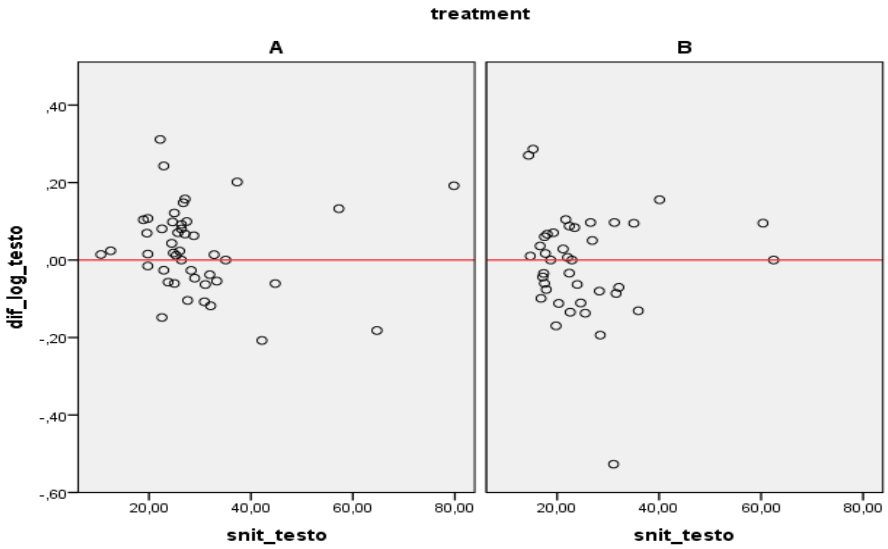
```
dif_log_testo=log_testo_right-log_testo_left
dif_testo=testo_left-testo_right
snit_testo=(testo_left+testo_right)/2
```

   Når vi skal se på sideforskelle, skal vi naturligvis stadig arbejde på logaritmisk skala, som det fremgår af nedenstående Bland-Altman plots, der tydeligt viser trompetfacon.

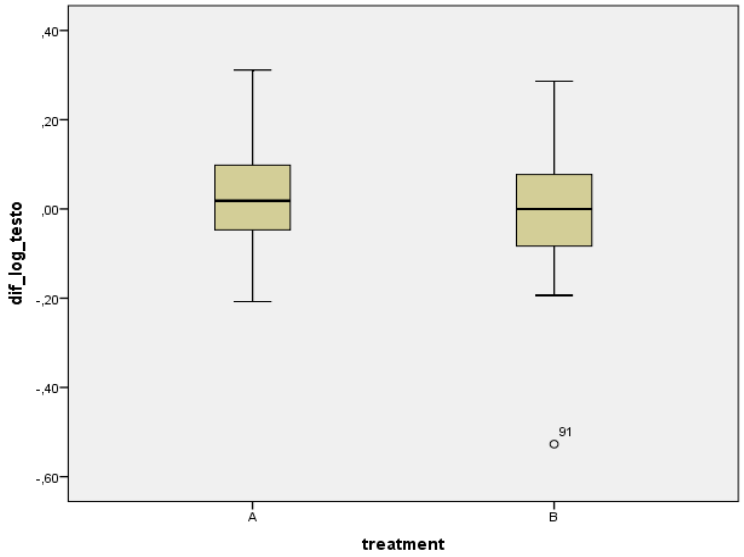
   Disse er dannet ved at gå i `Graph/Chart Builder/Scatter`, vælge det simple scatterplot, klikke på `Groups/Point ID`, afkrydse `Rows panel variable` og sætte `treatment` over i `Panel?`. Desuden `dif_testo` over på Y-aksen, og `snit_testo` over på X-aksen.



De tilsvarende Bland-Altman plots, for de logaritmerede værdier, bliver:



som ser noget bedre ud. Vi kvantificerer derfor differenserne på logskala, først med et Box-plot:



og så ved at udføre tests for middelværdi 0, dvs. ved at lave parrede T-tests for de to sider, for hver stimulationsmetode for sig:

Paired Samples Statistics

treatment			Mean	N	Std. Deviation	Std. Error Mean
A	Pair 1	log_testo_left	1,4184	45	,16924	,02523
		log_testo_right	1,4469	45	,15458	,02304
B	Pair 1	log_testo_left	1,3713	40	,16951	,02680
		log_testo_right	1,3602	40	,15655	,02475

Paired Samples Correlations

treatment			N	Correlation	Sig.
A	Pair 1	log_testo_left & log_testo_right	45	,779	,000
B	Pair 1	log_testo_left & log_testo_right	40	,653	,000

Paired Samples Test

treatment			Paired Differences: Mean	Std. Deviation	Std. Error Mean
A	Pair 1	log_testo_left - log_testo_right	-,02857	,10858	,01619
B	Pair 1	log_testo_left - log_testo_right	,01110	,13629	,02155

Paired Samples Test

treatment			Paired Differences: 95% Confidence Interval of the Difference: Lower	Upper	t	df
A	Pair 1	log_testo_left - log_testo_right	-,06119	,00405	-1,765	44
B	Pair 1	log_testo_left - log_testo_right	-,03249	,05469	,515	39

Paired Samples Test

treatment			Sig. (2-tailed)
A	Pair 1	log_testo_left - log_testo_right	,084
B	Pair 1	log_testo_left - log_testo_right	,609

Herved får vi P-værdierne 0.084 hhv. 0.609, og dermed ingen signifikante sideforskelle.

Den estimerede forskel (højre minus venstre) er (på logaritmisk skala) 0.0286 hhv -0.0111 for hhv stimulationsmetode A og B, dvs. sideforskellene går i forskellige retninger. Vi kan dog se af konfidensintervallerne, at ingen af dem er signifikant forskellige fra 0, dvs. der er ingen evidens for sideforskelle for nogen af stimulationsmetoderne.

For at forstå disse sideforskelle, skal vi tilbagetransformere til ratio'er, f.eks. for stimulationsmetode A: $10^{0.0286} = 1.07$, hvilket betyder, at værdierne på højre side i gennemsnit ligger ca. 7% over de tilsvarende i venstre side, med konfidensintervallet $(10^{-0.0041}, 10^{0.0612}) = (0.99, 1.15)$, altså fra mellem 1% under til 15% over.

Tilsvarende fås for stimulationsmetode B: $10^{-0.0111} = 0.97$, hvilket be-

tyder, at værdierne på højre side i gennemsnit ligger ca. 3% *under* de tilsvarende i venstre side, med konfidensintervallet $10^{-0.0547}, 10^{0.0325} = (0.88, 1.08)$, altså fra mellem 12% under til 8% over.

Ser det ud som om forskellen på stimulationsmetoderne er den samme på de to sider (for testosteron)?

Umiddelbart kunne det her se ud til, at der var lagt op til uparrede T-tests for hver af siderne, og da vi allerede i spm. 4 har set på højre side, supplerer vi her med venstre side:

Group Statistics

	treatment	N	Mean	Std. Deviation	Std. Error Mean
log_testo_left	A	48	1,4199	,16873	,02435
	B	42	1,3649	,16811	,02594

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
log_testo_left	Equal variances assumed	,041	,840	1,545	88
	Equal variances not assumed			1,546	86,505

Independent Samples Test

		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
log_testo_left	Equal variances assumed	,126	,05500	,03559
	Equal variances not assumed	,126	,05500	,03558

Independent Samples Test

		t-test for Equality of Means: 95% Confidence Interval of the Difference	
		Lower	Upper
log_testo_left	Equal variances assumed	-,01573	,12573
	Equal variances not assumed	-,01573	,12573

Sammenfattende finder vi kvantificeringerne af forskellen på de to stimulationsmetoder, for hver side (for A vs. B), på $\log_{10}$-skala:

Venstre side: 0.0550 (-0.0157, 0.1257),
tilbagetransformeret til 1.135 (0.965, 1.336), altså 13.5% højere for A end for B, med CI fra 3.5% under til 33.6% over.

Højre side: 0.0719 (0.0073, 0.1365),
tilbagetransformeret til 1.180 (1.017, 1.369), altså 18.0% højere for A end for B, med CI fra 1.7% over til 36.9% over.

For højre side er der således signifikant forskel, medens der på venstre side *ikke* er. Dette er imidlertid *absolut* ikke nok til at sige, at der er forskel på effekterne på de to sider! Ydermere er der et lille problem: Det er ikke (helt) de samme personer for de to sider, idet manglende observationer for enkelte individer til at ryge ud af analyserne.....

Men hvordan finder vi så en P-værdi for testet af, om de to forskelle på stimulationsmetoderne er ens??

Vi benytter her igen de tidligere indførte betegnelser, nemlig H_A, H_B, V_A, og V_B for `log_testo`, på hhv Højre og Venstre side, for stimulationsmetode A og B.

Vi vil gerne undersøge, om forskellen på A og B er ens for de 2 sider, dvs. om

$$H_A - H_B = V_A - V_B$$

men ved at bytte lidt rundt på disse led, ses dette at være det samme som at undersøge om

$$H_A - V_A = H_B - V_B$$

altså om side-differenserne `dif_log_testo` er ens for de to stimulationsmetoder.

Vi laver derfor et uparret T-test til sammenligning af stimulation A og B for sideforskellene `dif_log_testo=log_testo_left-log_testo_right`:

Group Statistics

	treatment	N	Mean	Std. Deviation	Std. Error Mean
dif_log_testo	A	45	,0286	,10858	,01619
	B	40	-,0111	,13629	,02155

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
dif_log_testo	Equal variances assumed	,581	,448	1,492	83
	Equal variances not assumed			1,472	74,422

Independent Samples Test

		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
dif_log_testo	Equal variances assumed	,140	,03967	,02659
	Equal variances not assumed	,145	,03967	,02695

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
dif_log_testo	Equal variances assumed	-,01322	,09257
	Equal variances not assumed	-,01402	,09337

Vi ser, at der ikke findes klare tegn på forskel her (P=0.14), så vi må konkludere, at det sagtens kan tænkes, at stimulationstyperne virker ens på de to sider.

Bemærk, at der i denne sidste analyse indgår endnu færre personer end i de tidligere, fordi nu kun personer med målte værdier for alle hormoner indgår.

I de to sidste spørgsmål ser vi på **ægkvaliteten**.

7. *Giv et estimat for sandsynligheden for høj ægkvalitet, for hver af de to stimulationsmetoder hver for sig. Ser de to estimater forskellige ud?*

Når vi skal udføre noget for hver stimulationstype for sig, starter vi med at lave en `Split File`, hvor vi afkrydser `Compare groups` og sætter `treatment` over i `Groups Based On`.

Dernæst går vi ind i `Generalized Linear Models/Generalized Linear Models`, hvor vi afkrydser `Custom`, sætter `Distribution` til `Binomial`, `Link function` til `Identity`, og `Response` til `kvalitet`. For at få procenterne for god ægkvalitet, kan man evt skifte `Reference Category` til `First`, men dette er ikke gjort her.

Vi finder så (blandt en hel del andet):

Parameter Estimates

				95% Wald Confidence Interval		Hypothesis Test
treatment	Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square
A	(Intercept)	,412	,0689	,277	,547	35,700
	(Scale)	1[a]				
B	(Intercept)	,333	,0703	,196	,471	22,500
	(Scale)	1[a]				

De to estimater **for dårlig ægkvalitet!**, med tilhørende (desværre kun approksimative) konfidensintervaller ses at være:

A: 0.41 (0.28, 0.55)

B: 0.33 (0.20, 0.47)

Hvis vi ønsker estimater for **god kvalitet** i stedet, skal vi bare trække fra 1:

A: 0.59 (0.45, 0.72)

B: 0.67 (0.53, 0.80)

De ser jo ikke særligt forskellige ud, og vi tester hypotesen om identitet nedenfor:

8. *Lav et test for om kvaliteten af de udtagne æg afhænger af stimulationsmetoden. Kvantificer forskellen, dels som en forskel på de to sandsynligheder fundet ovenfor og dels som en relativ risiko. Husk i begge tilfælde at supplere med et konfidensinterval. Kan vi udelukke, at der er dobbelt så stor sandsynlighed for at have god ægkvalitet i den ene gruppe i forhold til den anden?*

   Vi aflyser ny `Split File` og udfører et χ^2-testet for uafhængighed ved at gå ind i
   `Analyze/Descriptive Statistics/Crosstabs/`, og sætte `treatment` over i `Row(s)` og `kvalitet` i `Column(s)`

   Vi går derefter ind i `Statistics` og afkrydser `Chi-square`, `Somer` og `Risk`. Desuden går vi ind i `Exact` og afkrydser `Exact`, så vi får Fishers eksakte test med:

   og endelig går vi ind i `Cells` for at afkrydse, at vi, foruden de observerede antal (`Observed`), gerne vil se de forventede antal (`Expected`), samt under `Percentages`: `Row` for at få rækkeprocenter, dvs. sandsynlighed for god/dårlig kvalitet for hver af de to stimulationsmetoder.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
treatment * kvalitet	96	100,0%	0	0,0%	96	100,0%

treatment * kvalitet Crosstabulation

			kvalitet		
			0	1	Total
treatment	A	Count	21	30	51
		Expected Count	19,1	31,9	51,0
		% within treatment	41,2%	58,8%	100,0%
	B	Count	15	30	45
		Expected Count	16,9	28,1	45,0
		% within treatment	33,3%	66,7%	100,0%
Total		Count	36	60	96
		Expected Count	36,0	60,0	96,0
		% within treatment	37,5%	62,5%	100,0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,627[a]	1	,428	,527	,281
Continuity Correction[b]	,337	1	,561		
Likelihood Ratio	,629	1	,428	,527	,281
Fisher's Exact Test				,527	,281
N of Valid Cases	96				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 16,88.

b. Computed only for a 2x2 table

Directional Measures

			Value	Asymptotic Standard Error[a]	Approximate T[b]
Ordinal by Ordinal	Somers' d	Symmetric	,081	,101	,797
		treatment Dependent	,083	,104	,797
		kvalitet Dependent	,078	,098	,797

Directional Measures

			Approximate Significance	Exact Significance
Ordinal by Ordinal	Somers' d	Symmetric	,426	,527
		treatment Dependent	,426	,527
		kvalitet Dependent	,426	,527

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Risk Estimate

	Value	95% Confidence Interval Lower	Upper
Odds Ratio for treatment (A / B)	1,400	,608	3,223
For cohort kvalitet = 0	1,235	,729	2,094
For cohort kvalitet = 1	,882	,648	1,202
N of Valid Cases	96		

Vi ser af såvel χ^2-test og Fishers eksakte test, at der ikke er signifikant forskel på ægkvaliteten for de to behandlinger (P=0.43 hhv. 0.53). De forventede antal er alle et pænt stykke over de krævede 5 (minimum 16.9), så det er ikke nødvendigt at lave Fishers eksakte test, da chi-i-anden approksimationen er god nok.

De estimerede sandsynligheder for god ægkvalitet ses under `% within treatment` i kolonnen `kvalitet 1`. De er A: 58.8, og B: 66.7.

Den estimerede forskel på de to sandsynligheder for god ægkvalitet aflæses under **Directional Measures**, eller `Somer's d, kvalitet Dependent` til 0.078 ($\approx$ 0.667-0.588), altså ca. 8 procentpoint, med asymptotisk standard error på 0.098, svarende til konfidensgrænserne (-0.118, 0.274). Der kan altså med rimelighed være op til 27%-point større sandsynlighed for god ægkvalitet for treatment B.

Den relative "risiko" for god ægkvalitet for stimulationsmetode A vs. B udregnes nemt i hånden til

$$\frac{0.6667}{0.5882} = 1.13$$

men for at få konfidensgrænser på dette tal, må vi benytte SPSS. Desværre er 1.13 dog ikke blandt de tal, SPSS giver os, og det er fordi den angiver den relative risiko for treatment A vs. B i stedet. For `cohort kvalitet=1` er denne angivet som 0.882, og ved at invertere denne, fås

$$\frac{1}{0.882} = 1.13$$

Tilsvarende kunne vi invertere det tilhørende konfidensinterval:

$$\left(\frac{1}{1.202}, \frac{1}{0.0.648}\right) = (0.83, 1.54)$$

Man kan derfor *ikke* sige, at sandsynligheden for god ægkvalitet kan være dobbelt så stor i den ene gruppe i forhold til den anden, men bemærk, at man *godt* kan sige, at den ene gruppe (gruppe B) kan have dobbelt så store *odds* for god ægkvalitet.

Vi kunne få SPSS til at udregne disse inverteringer ved at bytte om på rækkefølgen af vores stimulationsformer, f.eks. ved at kalde dem `2:A` hhv. `1:B`: