

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

Opgave 1: Sædkvalitet

Filen `oeko.dat` er en let modificeret udgave af `oeko.txt` på hjemmesiden, blot med variabelnavnet `sas.ansat` i stedet for `sas_ansat`. Den 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`: abstinenstid (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.

Vi indlæser data i data framen `oeko`:

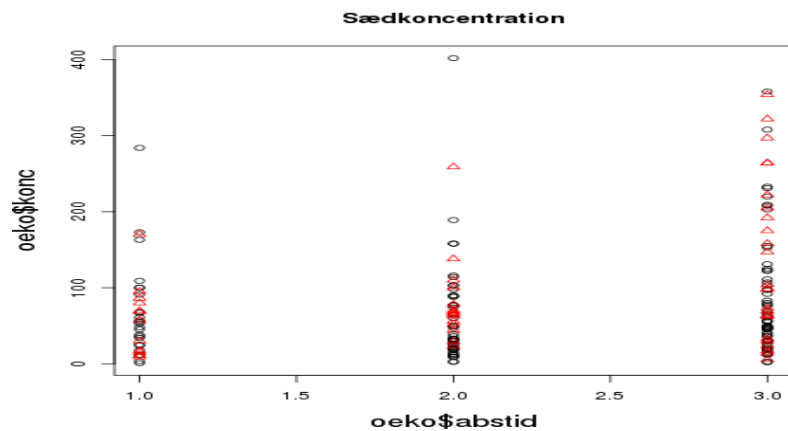
```
oeko <- read.table("oeko.dat", header=T)
```

1. Lav en passende illustration af data.

En umiddelbar optegning af sædkoncentration mod abstinenstid, med farveangivelse for grupperne kan udføres ved at skrive som nedenfor:

```
plot(oeko$abstid,oeko$konc,pch=as.numeric(oeko$sas.ansat),  
     col=as.numeric(oeko$sas.ansat), main="Sædkoncentration",  
     cex.lab=1.5)
```

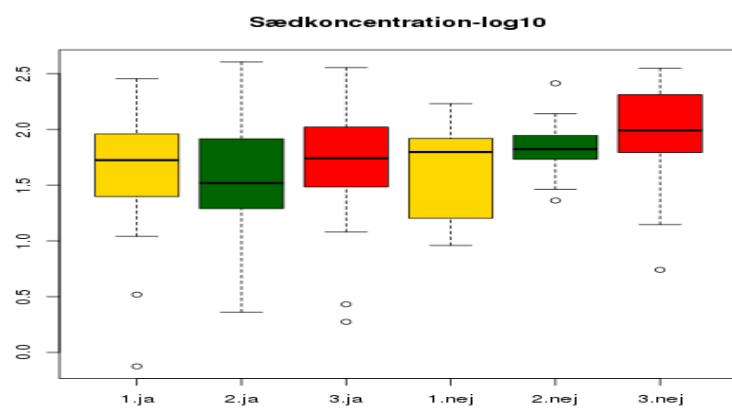
hvorved vi får figuren



Heraf ses, at sædkoncentration næppe er normalfordelt inden for hver gruppe, og vi laver derfor en logaritmetransformation (her er brugt $\log_{10}$). Samtidig skiller vi observationerne ud i 6 grupper, så vi bedre kan se, hvad der er hvad, og benytter nu Boxplots til at illustrere fordelingerne:

```
oeko$lkonc=log10(oeko$konc)
```

```
boxplot(lkonc~abstid*sas.ansat, data=oeko,
        col=(c("gold","darkgreen","red")), main="Sædkoncentration")
```



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 abstinensperioden. 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, så de nedenstående beregninger er foretaget på den logaritmetransformerede sædkoncentration ($\log_{10}$):

```
gennemsnit=tapply(oeko$lkonc,oeko$sas.ansat,mean)
median=tapply(oeko$lkonc,oeko$sas.ansat,median)
sd=tapply(oeko$lkonc,oeko$sas.ansat,sd)
sem=tapply(oeko$lkonc,oeko$sas.ansat,sd)/
      sqrt(tapply(oeko$lkonc,oeko$sas.ansat,length))
df=tapply(oeko$lkonc,oeko$sas.ansat,length)-1
lower=gennemsnit-qt(0.975,df)*sem
upper=gennemsnit+qt(0.975,df)*sem

cbind(gennemsnit,median,sd,sem, df, lower,upper)
```

hvorved vi får

	gennemsnit	median	sd	sem	df	lower	upper
ja	1.648640	1.681241	0.4715158	0.04058162	134	1.568377	1.728903
nej	1.834576	1.838849	0.4197763	0.05766070	52	1.718872	1.950281

Tilsvarende kunne vi have gjort for de utransformerede værdier,

og sammenfattes i en tabel ser det således ud

Data	Gruppe	gennemsnit	SEM	95% konfidensinterval
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.

Bemærk pointen her:

Der kan **godt** være signifikant forskel på to gruppers middelværdi, selv om konfidensintervallerne overlapper (lidt).

Den omvendte vej gælder det dog, at hvis konfidensintervallerne *ikke* overlapper, så er der signifikant forskel.

- (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 abstinentstiden, drejer det sig blot om et T-test. Antagelserne er bedst på log-skala, så det er den, vi benytter. Vi laver også lige først et test for varianshomogenitet:

```
var.test(lkonc~sas.ansat,data=oeko)
```

F test to compare two variances

```
data: lkonc by sas.ansat
F = 1.2617, num df = 134, denom df = 52, p-value = 0.3414
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7806967 1.9465644
sample estimates:
ratio of variances
      1.261702
```

Med en P-værdi på 0.34 er der ikke tegn på forskel på spredningerne i de to grupper, men et test siger ikke alt, så i praksis vil man ofte benytte Welch test, eller i hvert fald sammenligne til dette. Når man laver T-test i R, får man i øvrigt Welch test som default, hvis man ikke specificerer `var.equal=T`, se begge versioner nedenfor:

```
t.test(lkonc~sas.ansat,data=oeko)
```

Welch Two Sample t-test

```
data: lkonc by sas.ansat
t = -2.637, df = 106.17, p-value = 0.009619
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.32572642 -0.04614659
sample estimates:
mean in group ja mean in group nej
      1.648640      1.834576
```

```
t.test(lkonc~sas.ansat,data=oeko,var.equal=T)
```

Two Sample t-test

```
data: lkonc by sas.ansat
t = -2.5065, df = 186, p-value = 0.01305
```

```

alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.33228274 -0.03959027
sample estimates:
mean in group ja mean in group nej
      1.648640      1.834576

```

Fra det sidste T-test (med antagelse om ens varianser), ses det, at SAS-ansatte har en signifikant lavere sædkoncentration end økologer ($P=0.013$). Forskellen på logaritmisk-skala er $1.649 - 1.835 = -0.186$, 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 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 abstinenstiden og udregn passende summary statistics for de nu i alt 6 grupper.

Vi arbejder videre på log-skala, og skal nu dele op efter 2 variable. Ved at benytte en *add-on* pakke ved navn *Rmisc*, kan dette gøres ret simpelt:

```

install.packages("Rmisc")
library(Rmisc)

summarySE(oeko, measurevar="lkonc", groupvars=c("sas.ansat","abstid"))

```

	sas.ansat	abstid	N	lkonc	sd	se	ci
1	ja	1 25	1.604562	0.5515682	0.11031364	0.2276762	

2	ja	2	47	1.567940	0.4564178	0.06657537	0.1340093
3	ja	3	63	1.726336	0.4430979	0.05582509	0.1115928
4	nej	1	12	1.613619	0.4193392	0.12105279	0.2664354
5	nej	2	16	1.834595	0.2482210	0.06205526	0.1322677
6	nej	3	25	1.940624	0.4759862	0.09519723	0.1964774

- (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 selve effekten af at leve økologisk.

Vi vil lave en simpel tabel, så som:

```
> tabel=table(oeko$sas.ansat,oeko$abstid)
> tabel
```

	1	2	3
ja	25	47	63
nej	12	16	25

og tilføje rækkeprocenter, så vi får fordelingen på de 3 abstinens-tider i hver af de to grupper:

```
prop.table(tabel,1)*100
```

	1	2	3
ja	18.51852	34.81481	46.66667
nej	22.64151	30.18868	47.16981

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 abstinens-tid.*

Hvis abstinens-tiden 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 godt nok give en sammenligning af de to grupper af mænd, men det ville ikke give et retvisende indtryk af selve effekten af den økologiske levevis.

For at vurdere effekten af selve den økologiske levevis, skal vi sammenligne SAS-ansatte med økologer, under forudsætning af samme abstinens-tid, og det er præcis hvad en (additiv) tosidet variansanalysemodel gør.

Vi udfører denne ved at benytte proceduren `lm`, men da vi skal udnytte resultaterne på forskellig vis bagefter, giver vi først et navn til modellen og ser bagefter på udvalgte dele af denne:

```
> anova2=lm(lkonc~as.factor(abstid)+as.factor(sas.ansat),data=oeko)
> summary(anova2)
```

Call:

```
lm(formula = lkonc ~ as.factor(abstid) + as.factor(sas.ansat),
    data = oeko)
```


Residuals:

	Min	1Q	Median	3Q	Max
	-1.67182	-0.25675	0.03867	0.29830	1.01603

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.54688	0.07814	19.797	<2e-16 ***
as.factor(abstid)2	0.04131	0.09389	0.440	0.6605
as.factor(abstid)3	0.18723	0.08874	2.110	0.0362 *
as.factor(sas.ansat)nej	0.18691	0.07348	2.543	0.0118 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4526 on 184 degrees of freedom

Multiple R-squared: 0.06387, Adjusted R-squared: 0.0486

F-statistic: 4.184 on 3 and 184 DF, p-value: 0.006803

For at få konfidensgrænser på parameterestimerne, skriver vi

```
> confint(anova2)
```

	2.5 %	97.5 %
(Intercept)	1.39272131	1.7010409
as.factor(abstid)2	-0.14393631	0.2265621
as.factor(abstid)3	0.01215977	0.3623081
as.factor(sas.ansat)nej	0.04192633	0.3318850

og for at få det overordnede test af abstinenstidseffekten (med 2 frihedsgrader), skriver vi

```
> anova(anova2)
```

Analysis of Variance Table

Response: lkonc

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(abstid)	2	1.246	0.62325	3.0420	0.05015 .
as.factor(sas.ansat)	1	1.325	1.32547	6.4694	0.01180 *
Residuals	184	37.699	0.20488		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

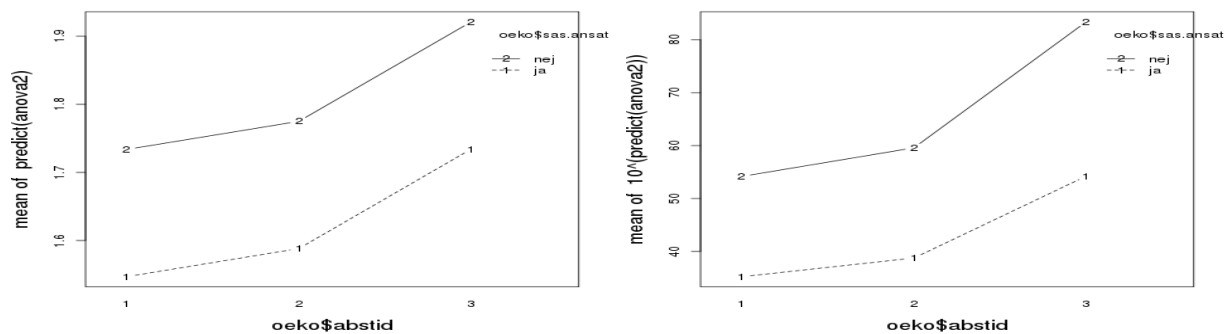
Begge kovariater ses at være signifikante, abstinensiden dog lige på kanten ($P=5.0\%$). For `abstid` ses de to laveste abstinensider at ligge nogenlunde på samme niveau (niveau 1 er reference, og svarer altså til et 0), mens sædkoncentrationen er højere for mænd med lang abstinensid, ganske som vi konkluderede ud fra gennemsnittene ovenfor. Endvidere ses, at SAS-ansatte har en signifikant lavere sædkoncentration end økologer *med samme abstinensid*. Forskellen på logaritmisk-skala er -0.187 , svarende til at de SAS-ansattes sædkoncentration udgør i gennemsnit kun udgør $10^{-0.187} = 65\%$ af økologernes koncentrationsniveau.

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

$$(10^{-0.332}, 10^{-0.042}) = (47\%; 91\%)$$

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, først på logaritmeskala, og bagefter tilbagetransformeret.



Figureerne er dannet ved at skrive

```
interaction.plot(oeko$abstid,oeko$sas.ansat,predict(anova2),
  type="b",cex.lab=1.5)
```

```
interaction.plot(oeko$abstid,oeko$sas.ansat,10^(predict(anova2)),
  type="b", cex.lab=1.5)
```

Bemærk, at de predikterede værdier på den højre figur er tilbage-transformeret 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.

Modelkontrollen.

- *Varianshomogenitet?*

Vi kan checke antagelsen om ens varians i alle 6 grupper ved at bruge Bartletts test, som om vi havde at gøre med en ensidet variansanalyse (one-way ANOVA), der sammenligner alle de 6 grupper under et:

```
> bartlett.test(oeko$lkonc~interaction(as.factor(oeko$abstid),
                                       as.factor(oeko$sas.ansat)))
```

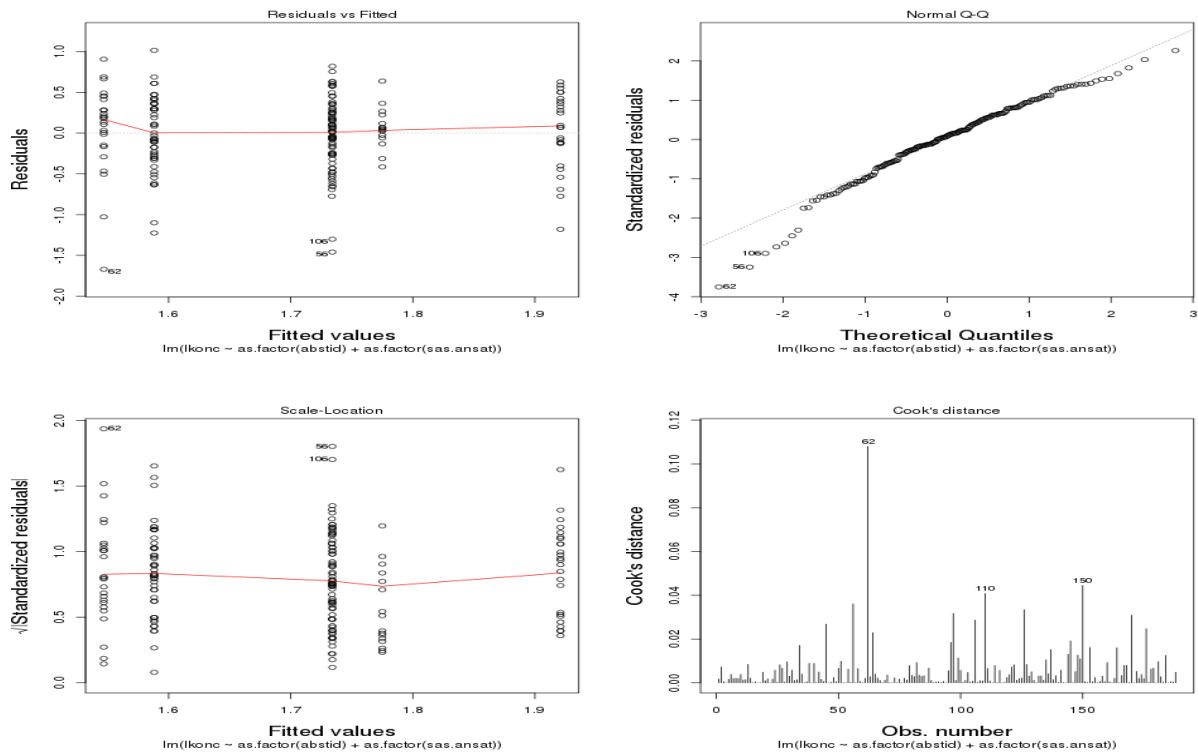
Bartlett test of homogeneity of variances

```
data:  oeko$lkonc by interaction(as.factor(oeko$abstid),
                                as.factor(oeko$sas.ansat))
Bartlett's K-squared = 9.8139, df = 5, p-value = 0.08068
```

Antagelsen om ens varianser ser altså ikke helt urimelig ud (P=0.08).

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

Varianshomogeniteten kan også vurderes ud fra plottet nedest til venstre, hvor den udglattede kurve ikke bør vise nogen trend (det lærer I om senere).



Normalfordelingsantagelsen?

QQ-plottet ovenfor viser, at logaritmetransformationen har bevirket en skævhed til 'den anden side', så normalfordelingsantagelsen er tvivlsom.

Teknisk note:

En bedre overensstemmelse kan opnås efter en kubikrodstransformation ($f(\text{konc}) = \text{konc}^{1/3}$). 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.*

Kan man sige, at forskelle i sædkoncentration i de to grupper

skyldes forskelle i deres abstinensstider?

Denne sammenligning er allerede kommenteret ovenfor. Der er ikke nævneværdig confounding at spore, så vi må konkludere, at forskellen i sædkvalitet *ikke* kan forklares ud fra forskelle i abstinensstid, men må søges andetsteds, evt. i selve den økologiske levevis.

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

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

Koden bliver derfor nu udbygget til

```
> veksle=lm(lkconc~as.factor(abstid)*as.factor(sas.ansat),data=oeko)
> summary(veksle)

Call:
lm(formula = lkconc ~ as.factor(abstid) * as.factor(sas.ansat),
    data = oeko)

Residuals:
    Min       1Q   Median       3Q      Max
-1.72950 -0.23766  0.01794  0.31855  1.03629

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.604562   0.090606  17.709   <2e-16 ***
as.factor(abstid)2    -0.036621   0.112143  -0.327    0.744
as.factor(abstid)3     0.121774   0.107085   1.137    0.257
as.factor(sas.ansat)nej  0.009058   0.159099   0.057    0.955
as.factor(abstid)2:as.factor(sas.ansat)nej  0.257597   0.206171   1.249    0.213
as.factor(abstid)3:as.factor(sas.ansat)nej  0.205230   0.191780   1.070    0.286
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.453 on 182 degrees of freedom
Multiple R-squared:  0.07245, Adjusted R-squared:  0.04697
F-statistic: 2.843 on 5 and 182 DF,  p-value: 0.01693

> confint(veksle)

                    2.5 %    97.5 %
(Intercept)      1.42578854 1.7833345
```

```

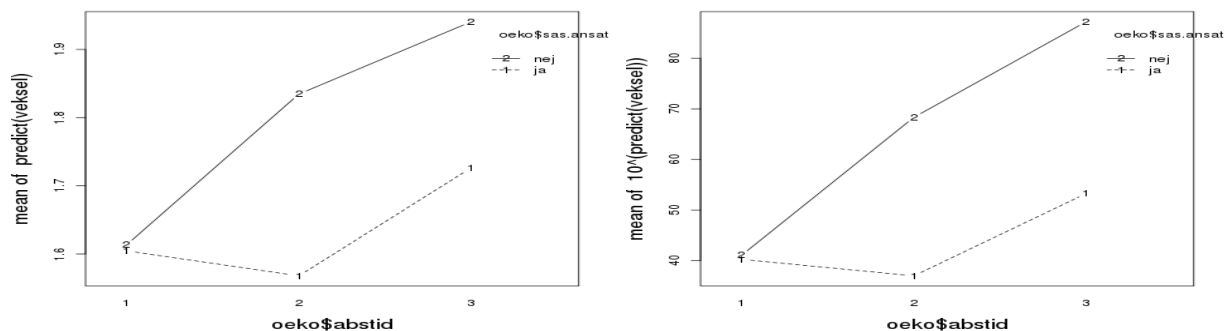
as.factor(abstid)2                -0.25788969 0.1846470
as.factor(abstid)3                -0.08951269 0.3330615
as.factor(sas.ansat)nej           -0.30485749 0.3229727
as.factor(abstid)2:as.factor(sas.ansat)nej -0.14919460 0.6643894
as.factor(abstid)3:as.factor(sas.ansat)nej -0.17316737 0.5836283
> anova(veksel)
Analysis of Variance Table

Response: lkonc

              Df Sum Sq Mean Sq F value Pr(>F)
as.factor(abstid)      2  1.246  0.62325   3.0367 0.05043 .
as.factor(sas.ansat)    1  1.325  1.32547   6.4583 0.01188 *
as.factor(abstid):as.factor(sas.ansat)  2  0.346  0.17290   0.8424 0.43233
Residuals             182 37.353  0.20524
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

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 abstinenstid, 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 data-materiale.

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). Tabellen nederst viser pulsen (slag pr. minut) før og efter flyvningen for hver af de 26 astronauter.

Countermeasure taken			Countermeasure not taken		
Pre	Post	Change	Pre	Post	Change
71	61	-10	61	61	0
65	59	-6	59	66	7
52	47	-5	52	61	9
68	65	-3	54	68	14
69	69	0	53	77	24
49	50	1	78	103	25
49	51	2	52	77	25
57	60	3	54	80	26
51	57	6	52	79	27
55	64	9			
58	67	9			
57	69	12			
59	72	13			
53	69	16			
53	72	19			
53	75	22			
48	77	29			
Mean	56.88	63.76	57.22	74.67	17.44
SD	7.30	8.86	8.44	13.01	10.11

Filen "space.txt" fra hjemmesiden ser således ud:

```
salt pre post
1 71 61
1 65 59
1 52 47
.....
.....
```

```

.....
0 52 77
0 54 80
0 52 79

```

og vi indlæser i data framen `sp`, og samtidig laver vi variables `salt` om til en faktor med niveauerne “S” og “C” for salt hhv. kontrol:

```

sp <-
read.table("http://publicifsv.sund.ku.dk/~lts/basal/data/space.txt",
header=T)

sp$salt <- factor(sp$salt, levels=c(0,1), labels=c("C", "S"))

```

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

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, så vi starter med at udregne disse samt gennemsnittene, da vi skal bruge dem til at lave Bland-Altman plot som check af antagelser i forbindelse med det parrede T-test:

```

sp$diff = sp$post - sp$pre
sp$mean = (sp$pre + sp$post)/2

```

Derefter skal vi opdele data, da vi i første omgang kun ser på salt-gruppen **S**. Vi gør det ved at opspalte vores data frame i to dele:

```

sp.S = sp[sp$salt=="salt",]
sp.C = sp[sp$salt=="control",]

```

Et histogram og et fraktildiagram for differenserne i salt-gruppen kan nu laves ved at skrive:

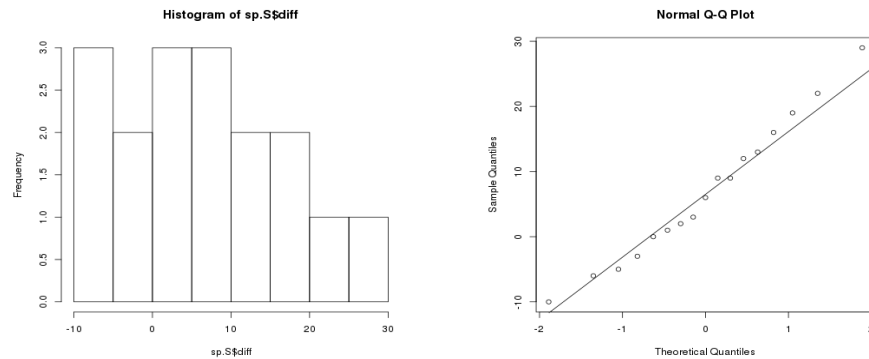
```

hist(sp.S$diff)

qqnorm(sp.S$diff)
qqline(sp.S$diff)

```

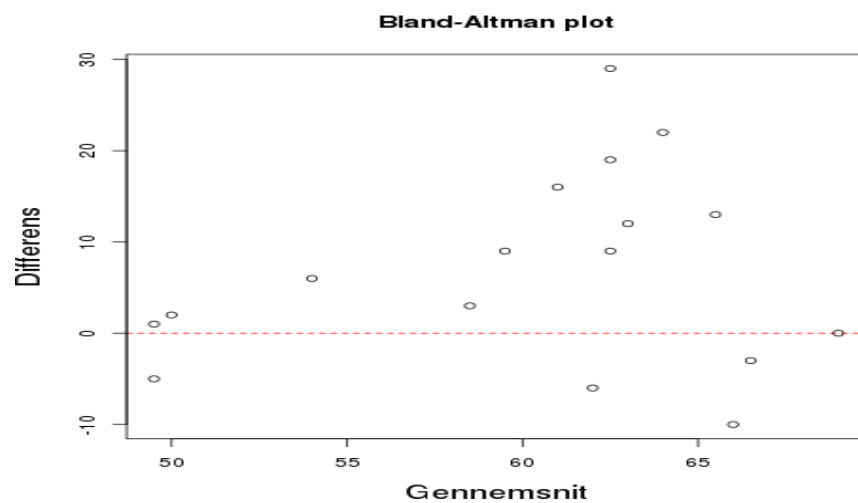

som giver figurene



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:

```
plot(sp.S$mean, sp.S$diff, abline(0,0,lty=2,col="red"),
     xlab="Gennemsnit", ylab="Differens", main="Bland-Altman plot")
```



Da dette heller ikke viser udprægede tegn på sammenhæng (eller er der måske lidt trompetfacon??), vil vi fortsætte med et parret t-test. Vi udfører t-testet ved at skrive

```
> t.test(sp.S$pre, sp.S$post, paired=T)
```

Paired t-test

```
data: sp.S$pre and sp.S$post
t = -2.6521, df = 16, p-value = 0.0174
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.383672 -1.381034
sample estimates:
mean of the differences
      -6.882353
```

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), se kode og output nedenfor. Herved finder vi en P-værdi på 0.028, som understøtter konklusionen fra t-testet:

```
> wilcox.test(sp.S$pre, sp.S$post, paired=T)
```

Wilcoxon signed rank test with continuity correction

```
data: sp.S$pre and sp.S$post
V = 25, p-value = 0.0279
alternative hypothesis: true location shift is not equal to 0
```

Warning messages:

```
1: In wilcox.test.default(sp.S$pre, sp.S$post, paired = T) :
  cannot compute exact p-value with ties
2: In wilcox.test.default(sp.S$pre, sp.S$post, paired = T) :
  cannot compute exact p-value with zeroes
```

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

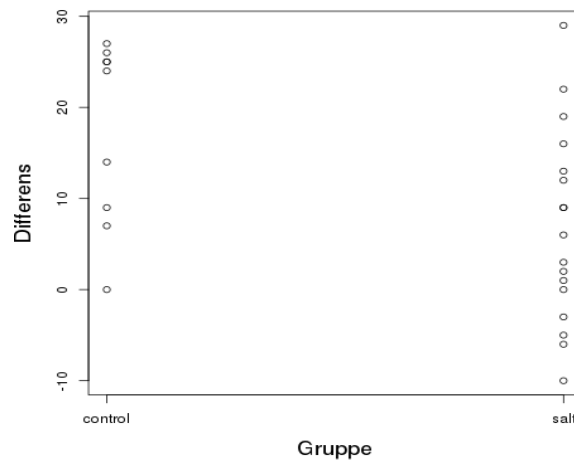
Udregning af et 95% konfidensinterval for middelværdien af pulsstigningen fås automatisk ved at udføre t-testet. Det ses at være (1.38,12.38), altså ganske bredt. Testet gav signifikans på et 5% niveau, svarende til, at 0 ikke er inkluderet i konfidensintervallet. Vi er altså noget usikre på, hvor stor 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:

```
stripchart(sp$diff ~ sp$salt, vertical=TRUE, pch=1,  
          ylab="Differens", xlab="Gruppe", cex.lab=1.5)
```



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:

```
> t.test(sp.S$diff, sp.C$diff, paired=F)
```

Welch Two Sample t-test

```
data: sp.S$diff and sp.C$diff
t = -2.4827, df = 17.26, p-value = 0.0236
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -19.527470 -1.596713
sample estimates:
mean of x mean of y
 6.882353 17.444444
```

Vi ser, at P-værdien for sammenligning af middelværdierne for differenserne i de to grupper er 0.024, 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 (1.60, 19.53). Ikke særligt overbevisende, men alligevel ...

Det tilsvarende non-parametriske test giver

```
> wilcox.test(sp.S$diff, sp.C$diff, paired=F)
```

Wilcoxon rank sum test with continuity correction

data: sp.S\$diff and sp.C\$diff

W = 36.5, p-value = 0.03308

alternative hypothesis: true location shift is not equal to 0

Warning message:

```
In wilcox.test.default(sp.S$diff, sp.C$diff, paired = F) :  
cannot compute exact p-value with ties
```

Vi finder altså også en signifikans i det non-parametriske test (P-værdien er 0.033 mod 0.024 i T-testet).

4. *To astronauter deltog i to forskellige flyvninger og optræder altså i data-materialet 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).

Sådanne problemer med manglende uafhængighed vil blive nærmere omtalt i kursets sidste uge.

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.