

Vejledende besvarelse af hjemmeopgave, efterår 2017

*Udleveret 3. oktober 2017, afleveres senest ved øvelserne i uge 44
(31. okt.-2. nov. 2017)*

På hjemmesiden

http://publicifsv.sund.ku.dk/~lts/basal17_2/hjemmeopgave/hjemmeopgave.txt

ligger data fra 72 unge piger med spiseforstyrrelsen anorexi.

De er inddelt i 3 grupper, alt efter hvad der er blevet gjort for at behandle denne spiseforstyrrelse:

Cont: Kontrolgruppen

CBT: Kognitiv behandling

FT: Familiebehandling

Der foreligger 4 variable for hver pige, og forslag til variabelnavne er angivet i 1. linie. Disse er:

idnr: Nummer på pigen (blot til brug for identifikation)

treat: Den anvendte behandling

prewt: Pigens vægt ved behandlingens start (i kg)

postwt: Pigens vægt ved afslutningen af behandlingen (i kg)

Succeskriteriet for behandlingerne er, at pigerne skal tage mindst 3 kg på i løbet af den tid, behandlingen varer, og opgaven er at undersøge, hvorvidt dette mål blev opnået, samt generelt at vurdere effekten af de 3 behandlinger og sammenligne disse.

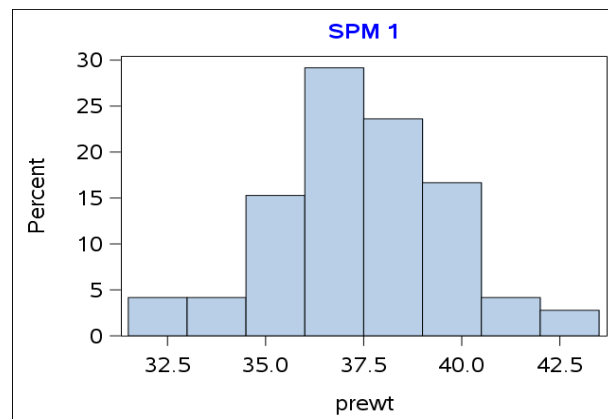
Først må vi jo indlæse vores data, og vi benytter de samme variabelbetegnelser som angivet ovenfor. Endvidere udregner vi for hver pige en vægtøgning $\text{diff} = \text{postwt} - \text{prewt}$, som i parentes bemærket kan være negativ, hvis der er tale om et vægttab:

```
data a1;
infile "http://staff.pubhealth.ku.dk/~lts/basal17_2/hjemmeopgave/hjemmeopgave.txt" URL firstobs=2;
input idnr treat $ prewt postwt;
diff=postwt-prewt;
run;
```

1. Lav en illustration af pigernes vægt ved starten på behandlingen (`prewt`).

Da der inden behandlingen burde være tale om en homogen gruppe, vælges blot et histogram:

```
title 'SPM 1';  
  
ods graphics / imagename="spm1";  
proc sgplot data=a1;  
    histogram prewt;  
run;
```



Her ser vi en pæn normalfordelingslignende fordeling, centreret omkring en vægt på ca 37 kg.

- (a) Angiv et 95% reference-interval for vægten inden behandling.
Er det usædvanligt at veje så lidt som 35 kg?

Vi skal udregne et reference-interval (normalområde), og da vi ovenfor så, at fordelingen ligner en normalfordeling, er det oplagt at benytte gennemsnit og spredning til dette.

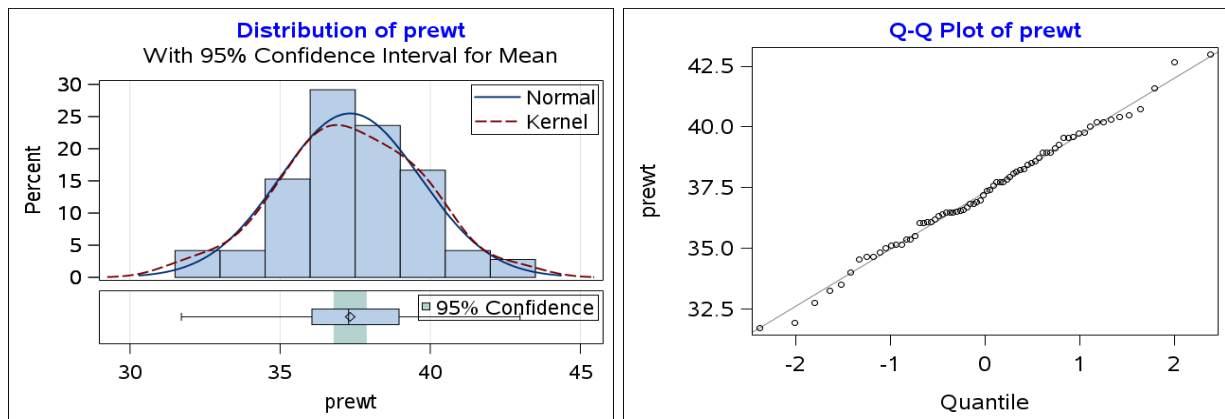
For en ordens skyld vil vi dog lige checke hvordan fraktildia-grammet ser ud, idet et sådant er bedre til at vurdere eventuelle afvigelser fra normalfordelingen, og desuden vil vi lægge en normalfordelingskurve ind på histogrammet ovenfor.

Sådanne to figurer fås nemt ved at bruge proceduren `ttest`, selv om vi slet ikke er interesseret i noget test for middelværdien af vægten:

```
title2 'SPM 1A';

proc ttest data=a1;
  var prewt;
run;
```

hvilket giver figurerne



Det ser faktisk rigtigt pænt ud, og vi fortsætter derfor med at udregne summary statistics for `prewt`, først default-værdier, men bagefter suppleret med median og kvartiler:

```
proc means data=a1;
  var prewt;
run;

proc means N mean median Q1 Q3 std data=a1;
  var prewt;
run;
```

hvorved vi får outputtet:

SPM 1A

The MEANS Procedure

Analysis Variable : prewt				
N	Mean	Std Dev	Minimum	Maximum
72	37.3309722	2.3480013	31.7100000	42.9900000

The MEANS Procedure

Analysis Variable : prewt					
N	Mean	Median	Lower Quartile	Upper Quartile	Std Dev
72	37.3309722	37.2800000	36.0600000	38.9600000	2.3480013

Vi ser her, at gennemsnit og median er næsten sammenfaldende, samt at Q1/Q3 (og faktisk også min/max) ligger ret pænt symmetrisk omkring gennemsnittet. Dette passer fint med vores vurdering af normalfordelingstilpasningen ovenfor.

Vi er derfor klar til at udregne et normalområde ud fra gennemsnit og spredning:

$$37.331 \pm 2 \times 2.348 = (32.64, 42.03)$$

Baseret på dette interval er det ikke særligt usædvanligt at veje så lidt som 35 kg, desværre.

Til trods for, at vi kun har 72 observationer, vil vi også forsøge at udregne et eksakt 95% normalområde, ud fra fraktilerne

```
proc univariate data=a1;
  var prewt;
  output out=spm1b pctlpts=2.5 97.5
           pctlpre=frak_ pctlname=lower upper;
run;

proc print data=spm1b;
run;
```

hvorved vi finder outputtet

	frak_ lower	frak_ upper
Obs		
1	31.94	42.67

Vi ser her en ret god overensstemmelse til det normalfordelings-baserede normalområde, selv om vi har så få observationer, men man har jo lov at være heldig.....

(b) *Ser det ud som om der er blevet randomiseret fornuftigt?*

Ovenfor har vi betragtet alle 72 piger under ét, men når de skal behandles, bliver de opdelt i tre grupper, som gerne skulle være ret ens, hvis randomiseringen har været fornuftig.

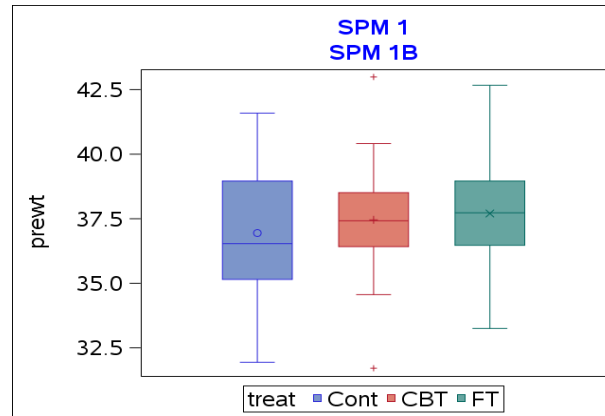
Vi kan se, om dette er tilfældet, dels ved at sammenligne grupperne grafisk (Boxplot) og dels ved at udregne simple summary statistics, opdelt på de tre randomiseringsgrupper:

```
title2 'SPM 1B';

ods graphics / imagename="spm1B";
proc sgplot data=a1;
vbox prewt / group=treat;
run;

proc means data=a1;
class treat;
var prewt;
run;
```

Herved finder vi figuren



samt outputtet

The MEANS Procedure

Analysis Variable : prewt						
treat	N		Mean	Std Dev	Minimum	Maximum
	Obs	N				
CBT	29	29	37.4575862	2.1957844	31.7100000	42.9900000
Cont	26	26	36.9465385	2.5858522	31.9400000	41.5900000
FT	17	17	37.7029412	2.2721734	33.2500000	42.6700000

Såvel figur som gennemsnit tyder på, at randomiseringen er forløbet tilfredsstillende.

- (c) *Kvantificer forskellen på de to grupper, der er mest forskellige ved baseline, dels ud fra en variansanalyse inkluderende alle tre grupper og dels ud fra et T-test. Kommenter på forskellen mellem resultaterne.*

Baseret alene på gennemsnittene ses FT og Cont at adskille sig mest fra hinanden. Vi sammenligner derfor nu disse posthoc efter en ensidet ANOVA, og benytter Tukey-korrektion, fordi vi har udvalgt grupperne ud fra deres forskel, ikke fordi det stod i protokollen.

Da vi allerede har set Boxplots af alle 3 grupper ovenfor, inkluderes her ingen nye figurer.

```

title2 'SPM 1C';

proc glm data=a1;
    class treat(ref="Cont");
    model prewt = treat / solution clparm;
lsmeans treat / adjust=tukey cl;
means treat / hovtest;
run;
ods graphics off;

```

Herved finder vi outputtet nedenfor, men bemærk, at vi kun er interesserede i kvantificeringerne, ikke i test og P-værdier. Det er dog betryggende at se, at Levenes test for identitet af de 3 grupper ikke giver anledning til mistanke om forskelligheder.

SPM 1C
The GLM Procedure

Class Level Information

Class	Levels	Values
treat	3	CBT FT Cont

Number of Observations Read	72
Number of Observations Used	72

The GLM Procedure
Dependent Variable: prewt

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6.6595595	3.3297798	0.60	0.5532
Error	69	384.7712724	5.5763953		
Corrected Total	71	391.4308319			

R-Square	Coeff Var	Root MSE	prewt Mean
0.017013	6.325684	2.361439	37.33097

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	2	6.65955951	3.32977975	0.60	0.5532

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	36.94653846 B	0.46311634	79.78	<.0001
treat CBT	0.51104775 B	0.63778228	0.80	0.4257
treat FT	0.75640271 B	0.73654599	1.03	0.3080
treat Cont	0.00000000 B	.	.	.

Parameter	95% Confidence Limits	
Intercept	36.02264665	37.87043027
treat CBT	-0.76129308	1.78338857

treat	FT	-0.71296632	2.22577175
treat	Cont	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

treat	prewt LSMEAN	LSMEAN Number
CBT	37.4575862	1
FT	37.7029412	2
Cont	36.9465385	3

Least Squares Means for effect treat
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: prewt

i/j	1	2	3
1		0.9383	0.7035
2	0.9383		0.5625
3	0.7035	0.5625	

treat	prewt LSMEAN	95% Confidence Limits	
CBT	37.457586	36.582786	38.332386
FT	37.702941	36.560370	38.845513
Cont	36.946538	36.022647	37.870430

Least Squares Means for Effect treat

		Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
i	j			
1	2	-0.245355	-1.973101	1.482391
1	3	0.511048	-1.016589	2.038685
2	3	0.756403	-1.007796	2.520602

Levene's Test for Homogeneity of prewt Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
treat	2	48.3894	24.1947	0.44	0.6472
Error	69	3812.1	55.2485		

Level of treat	N	Mean	Std Dev
CBT	29	37.4575862	2.19578436
FT	17	37.7029412	2.27217342
Cont	26	36.9465385	2.58585219

Den ukorrigerede sammenligning af FT vs. Control giver en forskel i FT's favør på 0.756, med konfidensgrænser CI=(-0.713, 2.226). I denne sammenligning er der benyttet et spredningsestimat, baseret

på alle 3 grupper.

Den Tukey-korrigerede sammenligning af FT vs. Control giver ligeledes forskel i FT's favør på 0.756, nu med lidt bredere konfidensgrænser $CI=(-1.008, 2.521)$.

Laver vi i stedet et T-test kun på FT og Control, er CBT-gruppens ret lille spredning ikke med i sammenligningen, og konfidensgrænserne burde derfor blive bredere, men til gengæld korrigerer vi nu ikke mere for de øvrige 2 sammenligninger (til CBT-gruppen), og vi ender derfor alt i alt op med smallere grænser, nemlig $(-0.798, 2.311)$, som det ses nedenfor.

Koden til T-testet er:

```
title2 'SPM 1C';

proc ttest data=a1; where treat in ("Cont","FT");
  class treat;
  var prewt;
run;
```

og outputtet bliver

SPM 1C

The TTEST Procedure
Variable: prewt

treat	N	Mean	Std Dev	Std Err	Minimum	Maximum
Cont	26	36.9465	2.5859	0.5071	31.9400	41.5900
FT	17	37.7029	2.2722	0.5511	33.2500	42.6700
Diff (1-2)		-0.7564	2.4682	0.7698		

treat	Method	Mean	95% CL Mean	Std Dev
Cont		36.9465	35.9021 37.9910	2.5859
FT		37.7029	36.5347 38.8712	2.2722
Diff (1-2)	Pooled	-0.7564	-2.3111 0.7983	2.4682
Diff (1-2)	Satterthwaite	-0.7564	-2.2733 0.7605	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	41	-0.98	0.3316
Satterthwaite	Unequal	37.405	-1.01	0.3190

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
--------	--------	--------	---------	--------

Folded F 25 16 1.30 0.5997

Nedenfor sammenfatter vi de forskellige kvantificeringer af forskellene mellem de to grupper, **FT** og **Cont**, nu reduceret til to decimaler:

	Differens FT vs. Cont	P-værdi for identitet
Fra ANOVA ukorrigeret	0.76 (-0.71, 2.23)	0.31
Fra ANOVA Tukey korrigeret	0.76 (-1.07, 2.52)	0.56
Fra T-test ens spredninger	0.76 (-0.80, 2.31)	0.33
Fra T-test uens spredninger	0.76 (-0.76, 2.27)	0.32

Bemærk, at vi i alle tilfælde får samme estimat for forskellen (nemlig differensen mellem de to gennemsnit), men at konfidensintervallerne varierer, idet de er bredere, når man Tukey-korrigerer, og sædvanligvis også, når man ikke kan antage ens spredninger (i T-testet). Det sidste er dog ikke tilfældet her, fordi grupperne har uens størrelse og den største gruppe har den største varians.

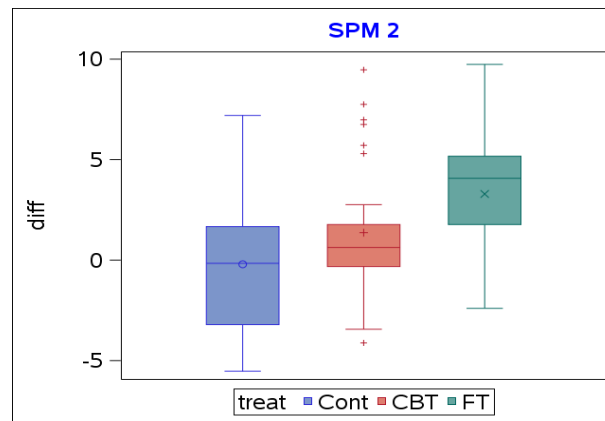
2. Kvantificer vægtændringen i hver af de 3 grupper.

I forbindelse med indlæsningen udregnede vi vægtændringen for hver pige som `diff=postwt-prewt`; og vi kvantificerer nu denne variabel med estimat og konfidensinterval for middelværdien. Først laver vi dog selvfølgelig en illustration, i form af et Boxplot af ændringerne.

```
title 'SPM 2';
title2;

ods graphics / imagename="spm2";
proc sgplot data=a1;
vbox diff / group=treat;
```

```
run;
```



```
proc means N mean clm data=a1;
    class treat;
    var diff;
run;
```

hvorved vi får

The MEANS Procedure

Analysis Variable : diff					
treat	N Obs	N	Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
CBT	29	29	1.3634483	0.1039684	2.6229281
Cont	26	26	-0.2050000	-1.6664766	1.2564766
FT	17	17	3.2911765	1.6240592	4.9582937

Vi ser, at gruppen **Cont** i gennemsnit har et ganske ubetydeligt vægttab, men med et konfidensinterval, der så absolut indeholder 0. Der er altså ikke evidens for nogen vægtændring i denne gruppe. For de to øvrige grupper ser vi derimod en gennemsnitlig vægtøgning, og af konfidensintervallerne kan vi se, at denne er signifikant i begge grupper, omend den nederste grænse er ret beskeden, specielt for gruppen CBT, hvor vægtøgningen måske kun beløber sig til ca. 100 gram i middel.

Vi kan også få tilhørende P-værdier, men det tager vi i forbindelse med næste spørgsmål.

- (a) *Sker der en signifikant øgning i alle grupper?
Og kommer middelværdien for denne op på de ønskede 3 kg?*

Vi har allerede ovenfor konstateret, at vægtøgningen er signifikant i de to aktive behandlingsgrupper, omend vi ikke endnu har knyttet P-værdier til denne påstand. Det gør vi nu, men samtidig vil vi gerne svare på, om vægtøgningen i middel kommer op på de ønskede 3 kg, og vi vil derfor også teste, om middelværdierne i de enkelte grupper kunne være 3. Det kan vi dog allerede konstatere ud fra konfidensintervallerne ovenfor, at dette kun kan være et tilfældet for gruppen FT, men for at knytte P-værdier til (hvis man finder dette ønskværdigt), kan vi danne en ny variabel `diff_minus3=diff-3`;, og så teste om middelværdien af denne er 0:

```
title2 'SPM 2A';

proc means N mean t probt data=a1;
  class treat;
  var diff diff_minus3;
run;
```

Herved finder vi

SPM 2A

The MEANS Procedure

treat	N		N	Mean	t Value	Pr > t
	Obs	Variable				
CBT	29	diff	29	1.3634483	2.22	0.0349
		diff_minus3	29	-1.6365517	-2.66	0.0127
Cont	26	diff	26	-0.2050000	-0.29	0.7750
		diff_minus3	26	-3.2050000	-4.52	0.0001
FT	17	diff	17	3.2911765	4.19	0.0007
		diff_minus3	17	0.2911765	0.37	0.7160

Vi sammenfatter igen i en tabel:

Gruppe	Vægtændring	P-værdi for øgning 0 kg	P-værdi for øgning 3 kg
CBT	1.36 (0.10, 2.62)	0.035	0.013
Cont	-0.20 (-1.67, 1.26)	0.78	< 0.0001
FT	3.29 (1.62, 4.96)	0.0007	0.72

Baseret på dette, ser FT klart bedst ud, til trods for, at (eller måske delvis fordi?) denne gruppe startede en anelse højere end de andre.

(b) *Er der evidens for forskellig vægtændring i de 3 grupper?*

Her skal vi sammenligne vægtændringerne (**diff**) i de 3 grupper, og baseret på Boxplottet ovenfor ser dette tilladeligt ud, da grupperne synes at have nogenlunde sammenlignelig spredning og en rimelig symmetrisk fordeling hver især. Vi udfører derfor en ensidet variansanalyse (ANOVA), med kontrolgruppen som reference. Samtidig laver vi et test for varianshomogenitet, samt posthoc sammenligninger af grupperne parvist, korrigeret med Tukey-korrektion:

```
title2 'SPM 2B';

proc glm plots=all data=a1;
  class treat(ref="Cont");
  model diff = treat / solution clparm;
  lsmeans treat / adjust=tukey cl;
  means treat / hovtest;
run;
```

Vi får så outputtet

```
SPM 2B

The GLM Procedure

Class Level Information

Class      Levels  Values
treat      3      CBT FT Cont

Number of Observations Read      72
Number of Observations Used      72
```

Dependent Variable: diff

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	126.2443628	63.1221814	5.43	0.0065
Error	69	802.5018816	11.6304621		
Corrected Total	71	928.7462444			

R-Square	Coeff Var	Root MSE	diff Mean
0.135930	272.3435	3.410346	1.252222

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	2	126.2443628	63.1221814	5.43	0.0065

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-0.205000000	0.66882394	-0.31	0.7601
treat CBT	1.568448276	0.92107322	1.70	0.0931
treat FT	3.496176471	1.06370592	3.29	0.0016
treat Cont	0.000000000	.	.	.

Parameter	95% Confidence Limits
Intercept	-1.539267254 1.129267254
treat CBT	-0.269042524 3.405939076
treat FT	1.374141217 5.618211724
treat Cont	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Least Squares Means

Adjustment for Multiple Comparisons: Tukey-Kramer

treat	diff LSMEAN	LSMEAN Number
CBT	1.36344828	1
FT	3.29117647	2
Cont	-0.20500000	3

Least Squares Means for effect treat
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: diff

i/j	1	2	3
1		0.1611	0.2114
2	0.1611		0.0045
3	0.2114	0.0045	

treat	diff LSMEAN	95% Confidence Limits
CBT	1.363448	0.100078 2.626818
FT	3.291176	1.641096 4.941257
Cont	-0.205000	-1.539267 1.129267

Least Squares Means for Effect treat

i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)
1	2	-1.927728	-4.422907 0.567451
1	3	1.568448	-0.637737 3.774633
2	3	3.496176	0.948353 6.044000

Levene's Test for Homogeneity of diff Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
treat	2	89.8374	44.9187	0.22	0.8045
Error	69	14200.1	205.8		

Level of treat	N	Mean	Std Dev
CBT	29	1.36344828	3.31111254
FT	17	3.29117647	3.24245640
Cont	26	-0.20500000	3.61833359

Igen er der ingen tegn på afvigelser fra varianshomogenitet (Levenes test giver $P=0.80$).

Testet for identitet af de tre vægtøgninger ses at give en P-værdi på 0.0065, altså signifikant forskel. Selve størrelsen af vægtøgningerne har vi allerede kvantificeret i tabellen side 13??, og det ses, at det også her er gruppen FT, der adskiller sig mest fra Cont.

- (c) *Kvantificer forskellen i vægtændring på de to grupper, der adskiller sig mest fra hinanden i denne henseende.*

Vi tog allerede skridt til en sådan sammenligning ovenfor ved at lave en post hoc sammenligning af grupperne parvist, med Tukey korrektion.

Vi kunne dog også have valgt at sammenligne dem direkte med et T-test:

```
title2 'SPM 2C';

proc ttest data=a1; where treat in ("Cont","FT");
  class treat;
  var diff;
run;
```

som giver os outputtet

SPM 2C

The TTEST Procedure

Variable: diff

treat	N	Mean	Std Dev	Std Err	Minimum	Maximum
Cont	26	-0.2050	3.6183	0.7096	-5.5300	7.2000
FT	17	3.2912	3.2425	0.7864	-2.4000	9.7400
Diff (1-2)		-3.4962	3.4765	1.0843		

treat	Method	Mean	95% CL Mean		Std Dev
Cont		-0.2050	-1.6665	1.2565	3.6183
FT		3.2912	1.6241	4.9583	3.2425
Diff (1-2)	Pooled	-3.4962	-5.6860	-1.3063	3.4765
Diff (1-2)	Satterthwaite	-3.4962	-5.6425	-1.3499	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	41	-3.22	0.0025
Satterthwaite	Unequal	36.974	-3.30	0.0021

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	25	16	1.25	0.6593

Vi sammenfatter igen i en tabel, denne gang sammenligningerne af vægtændringerne i de 3 grupper, dels fra ANOVA'en og dels fra T-testet:

	Differens FT vs. Cont	P-værdi for identitet
Fra ANOVA ukorrigeret	3.50 (1.37, 5.62)	0.0016
Fra ANOVA Tukey korrigeret	3.50 (0.95, 6.04)	0.0045
Fra T-test ens spredninger	3.50 (1.31, 5.69)	0.0025
Fra T-test uens spredninger	3.50 (1.35, 5.64)	0.0021

Bemærk, at estimatet under alle omstændigheder blot er forskellen på gennemsnittene i de to grupper, men at konfidensintervallerne afhænger af metoden.

Tukey-korrektionen giver naturligvis det bredeste konfidensinterval, fordi vi skal korrigere for tre mulige sammenligninger. Endvidere giver T-testet en anelse bredere konfidensgrænser, fordi CBT (med en lidt mindre spredning) nu ikke er inkluderet i modellen og dermed heller ikke bidrager til spredningsestimatet.

- (d) *Finder vi nogenlunde samme konklusion, når vi blot sammenligner slutvægten i de 3 (hvv 2) grupper?*

Vi ville forvente at finde nogenlunde samme resultat for slutvægten som for vægtændringerne, fordi der er randomiseret fra starten, og grupperne burde derfor ikke adskille sig ret meget fra hinanden

(som vi da også så det i første spørgsmål).

Men da der naturligt forekommer små afvigelser allerede ved baseline (FT-gruppen lå højest her), ville vi forvente, at FT også adskilte sig mere fra de andre, når vi så på follow-up måling (det er simpel addition).

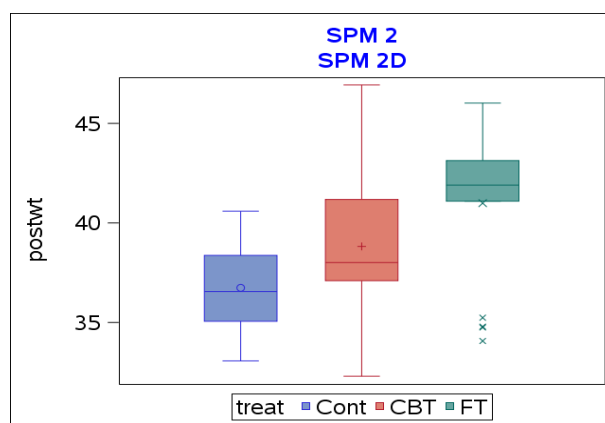
Vi foretager nu samme analyser som ovenfor, blot for `postwt`, først en ensidet variansanalyse (med tilhørende figur i form af et Boxplot):

```
title2 'SPM 2D';

proc sgplot data=a1;
vbox postwt / group=treat;
run;

proc glm data=a1;
class treat(ref="Cont");
model postwt = treat / solution clparm;
lsmeans treat / adjust=tukey cl;
means treat / hovtest welch;
run;
```

og får figuren



der klart synes at vise, at FT-gruppen ligger højest efter endt behandling, omend der forekommer nogle piger, som ligger meget

lavt i forhold til de øvrige i gruppen.

Vi får desuden outputtet

```
SPM 2D

The GLM Procedure

Class Level Information

Class          Levels    Values
treat          3         CBT FT Cont

Number of Observations Read      72
Number of Observations Used      72

Dependent Variable: postwt

Source          DF          Sum of Squares    Mean Square    F Value    Pr > F
Model           2      188.6379461      94.3189730      8.65    0.0004
Error           69      752.0628192      10.8994611
Corrected Total  71      940.7007653

R-Square      Coeff Var    Root MSE    postwt Mean
0.200529      8.556661      3.301433      38.58319

Source          DF    Type III SS    Mean Square    F Value    Pr > F
treat           2      188.6379461      94.3189730      8.65    0.0004

Parameter              Estimate      Standard Error    t Value    Pr > |t|
Intercept              36.74153846 B      0.64746432      56.75    <.0001
treat CBT              2.07949602 B      0.89165774      2.33    0.0226
treat FT              4.25257919 B      1.02973531      4.13    0.0001
treat Cont             0.00000000 B      .              .        .

Parameter              95% Confidence Limits
Intercept              35.44988249 38.03319444
treat CBT              0.30068749 3.85830455
treat FT              2.19831344 6.30684493
treat Cont             .          .

NOTE: The X'X matrix has been found to be singular, and a generalized
       inverse was used to solve the normal equations.  Terms whose
       estimates are followed by the letter 'B' are not uniquely estimable.

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

treat          postwt    LSMEAN
              LSMEAN    Number

CBT           38.8210345        1
FT            40.9941176        2
Cont          36.7415385        3

Least Squares Means for effect treat
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: postwt

i/j          1          2          3

1            0.0864      0.0580
2            0.0864      0.0003
3            0.0580      0.0003

treat          postwt    95% Confidence Limits
              LSMEAN
CBT           38.821034  37.598012  40.044057
FT            40.994118  39.396735  42.591501
Cont          36.741538  35.449882  38.033194

Least Squares Means for Effect treat

Difference      Simultaneous 95%
```

i	j	Between	Confidence Limits for	
		Means	LSMean(i) - LMSMean(j)	
1	2	-2.173083	-4.588576	0.242409
1	3	2.079496	-0.056232	4.215224
2	3	4.252579	1.786123	6.719035

Levene's Test for Homogeneity of postwt Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
treat	2	1468.4	734.2	3.82	0.0267
Error	69	13256.4	192.1		

Level of	-----postwt-----		
treat	N	Mean	Std Dev
CBT	29	38.8210345	3.78407409
FT	17	40.9941176	3.83834935
Cont	26	36.7415385	2.14846865

der også klart dokumenterer en forskel på de tre grupper (ANOVA giver $P=0.0004$), med et klart forspring til FT-gruppen.

Ser man på Levenes test for identiske varianser i de 3 grupper, ses en P-værdi på 0.027, så varianshomogeniteten er ikke helt i orden. Derfor supplerer vi med et Welch test for at sikre os, at signifikansen ikke hænger alt for meget på denne antagelse. Dette er gjort i sætningen `means treat / hovtest welch;` i koden ovenfor, og den giver outputtet

Welch's ANOVA for postwt

Source	DF	F Value	Pr > F
treat	2.0000	9.91	0.0004
Error	36.2348		

Vi må konstatere, at signifikansen holder fint til trods for en smule uens varianser.

Kvantificeringen af forskellen mellem FT og kontrolgruppen fremgår af outputtet ovenfor og er også indsat i tabellen nedenfor, hvor vi sammenligner til det T-test, vi nu foretager:

```
proc ttest data=a1; where treat in ("Cont","FT");
  class treat;
  var postwt;
run;
```

SPM 2D

The TTEST Procedure

Variable: postwt

treat	N	Mean	Std Dev	Std Err	Minimum	Maximum
Cont	26	36.7415	2.1485	0.4213	33.0700	40.5900
FT	17	40.9941	3.8383	0.9309	34.0700	46.0200
Diff (1-2)		-4.2526	2.9264	0.9128		

treat	Method	Mean	95% CL Mean	Std Dev
Cont		36.7415	35.8738 37.6093	2.1485
FT		40.9941	39.0206 42.9676	3.8383
Diff (1-2)	Pooled	-4.2526	-6.0960 -2.4092	2.9264
Diff (1-2)	Satterthwaite	-4.2526	-6.3684 -2.1368	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	41	-4.66	<.0001
Satterthwaite	Unequal	22.619	-4.16	0.0004

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	16	25	3.19	0.0092

Vi sammenfatter igen i en tabel:

	Differens FT vs. Cont	P-værdi for identitet
Fra ANOVA ukorrigeret	4.25 (2.20, 6.31)	< 0.0001
Fra ANOVA Tukey korrigeret	4.25 (1.79, 6.72)	0.0003
Fra T-test ens spredninger	4.25 (2.41, 6.10)	< 0.0001
Fra T-test uens spredninger	4.25 (2.14, 6.37)	0.0004

og når vi sammenligner denne tabel med den tilsvarende tabel s. 16 ser vi, at forskellen i slutvægt (som forventet) er større end forskellen i ændringerne, samt at P-værdierne er blevet mindre. Da dette skyldes den lette ubalance i baseline-værdier, er det resultaterne på s. 16, der er mest rimelige, sammen med de resultater, som vi finder senere i spm. 4.

3. Lav en tabel over, hvor mange piger, der stiger hhv falder i vægt, i hver

af de 3 grupper.

Giv i denne forbindelse en begrundelse for, hvordan man skal klassificere de piger, der holder præcis samme vægt igennem forløbet.

Da det drejer sig om at få pigerne til at tage på i vægt, er det her valgt at sige, at der skal være tale om en egentlig stigning, før det defineres som en sådan. Hvis man holder sin vægt uændret igennem behandlingen, defineres det ikke som en stigning (selv om det naturligvis heller ikke kan kaldes et fald).

Vi definerer altså `stigning=(diff>0)`; og laver så en lille tabel, hvor andelen af piger med vægtstigning kvantificeres, opdelt på de 3 grupper:

```
title 'SPM 3';
title2;

proc freq data=a1;
table treat*stigning / nopercnt nocol;
run;
```

hvorved vi finder

SPM 3
The FREQ Procedure

Table of treat by stigning

treat		stigning		
		0	1	Total
CBT	Frequency			
	Row Pct			
		11	18	29
		37.93	62.07	
Cont	Frequency	15	11	26
	Row Pct	57.69	42.31	
FT	Frequency	4	13	17
	Row Pct	23.53	76.47	
Total		30	42	72

Vi ser altså, at hele 76.5% af FT-pigerne og 62.1% af CBT-pigerne stiger i vægt, medens dette kun gælder for 42.3% af kontrolgruppens piger, altså *under halvdelen*.

Det kunne tyde på en reel forskel, hvilket undersøges i næste spørgsmål:

- (a) *Kan vi afvise, at der er samme risiko for at tabe sig i de 3 grupper?*

Her bliver vi bedt om at udtale os om sandsynligheder for at *tabe sig*, og da vi hidtil har defineret piger, der holdt deres vægt præcist som nogle, der ikke stiger, kan vi ikke bruge variablen `stigning` her, men definerer i stedet for variablen `fald=(diff<0)`. Vi kunne dog også argumentere for, at piger i voksenalderen i en vis forstand taber sig, hvis de ikke tager på over en periode....

Vi skal herefter sammenligne 3 frekvenser og se, om disse kunne tænkes at fremkomme som udslag af 3 ens sandsynligheder. Det svarer til et test for uafhængighed i en 3-gange-2 tabel. Vi tilføjer derfor options for χ^2 -testet og Fishers eksakte test for identitet af sandsynligheden for vægtfald i de tre grupper:

```
title2 'SPM 3A';

proc freq data=a1;
table treat*fald / nopercnt nocol chisq;
exact fisher;
run;
```

og finder outputtet:

```
SPM 3A
The FREQ Procedure

Table of treat by fald

treat      fald

Frequency|
Row Pct  |          0|          1|  Total
-----+-----+-----+
CBT      |          18|          11|      29
          |    62.07 |    37.93 |
-----+-----+-----+
Cont     |          12|          14|      26
```

		46.15 53.85	
		-----+-----+-----+	
FT		13 4	17
		76.47 23.53	
		-----+-----+-----+	
Total		43 29	72

Statistics for Table of treat by fald

Statistic	DF	Value	Prob
Chi-Square	2	4.0387	0.1327
Likelihood Ratio Chi-Square	2	4.1376	0.1263

Fisher's Exact Test

Table Probability (P)	0.0069
Pr <= P	0.1327

Sample Size = 72

Det ses, at de to tests er ret enige om en P-værdi på godt 13%, altså *ikke* tilstrækkelig evidens for at sige, at de 3 grupper klarer sig forskelligt. Ærgerligt nok, da der jo vitterligt ser ud til at være forskelle, men disse kan altså godt tilskrives tilfældigheder.

Det er tit problemet med for vage hypoteser, at man også får for vage svar. Her har vi 3 grupper, vi sammenligner på en gang, og det giver mindre styrke end hvis vi kun så på to. Det ser vi på nedenfor.

- (b) *Kvantificer den relative risiko for at tabe sig, hvis man er i kontrolgruppen i forhold til FT-gruppen.
Kan der være tale om en femdoblet risiko?*

Vi benytter en **where**-sætning til at udvælge de to grupper, vi gerne vil sammenligne, og da vi skal udtale os om relativ risiko, tilføjer vi option **relrisk**. Vi har desuden tilføjet **riskdiff**, mest for at få konfidensgrænser for sandsynlighederne for vægttab i de enkelte grupper:

```
title2 'SPM 3B';
proc freq data=a1; where treat in ("Cont","FT");
table treat*fald / nopercnt nocol chisq riskdiff relrisk;
run;
```

Herved får vi outputtet

SPM 3B

The FREQ Procedure

Table of treat by fald

treat fald

Frequency				
Row Pct		0	1	Total
-----+-----+-----+				
Cont		12	14	26
		46.15	53.85	
-----+-----+-----+				
FT		13	4	17
		76.47	23.53	
-----+-----+-----+				
Total		25	18	43

Statistics for Table of treat by fald

Statistic	DF	Value	Prob

Chi-Square	1	3.8819	0.0488
Likelihood Ratio Chi-Square	1	4.0262	0.0448
Continuity Adj. Chi-Square	1	2.7361	0.0981

Fisher's Exact Test

Two-sided Pr <= P 0.0637

Statistics for Table of treat by fald

Column 2 Risk Estimates

	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.5385	0.0978	0.3468	0.7301	0.3337	0.7341
Row 2	0.2353	0.1029	0.0337	0.4369	0.0681	0.4990
Total	0.4186	0.0752	0.2712	0.5661	0.2701	0.5787
Difference	0.3032	0.1419	0.0250	0.5813		

Difference is (Row 1 - Row 2)

Odds Ratio and Relative Risks

Statistic	Value	95% Confidence Limits	

Odds Ratio	0.2637	0.0677	1.0279
Relative Risk (Column 1)	0.6036	0.3691	0.9870
Relative Risk (Column 2)	2.2885	0.9048	5.7881

Sample Size = 43

Vi ser, at den relative risiko for vægttab i kontrolgruppen i forhold til FT-gruppen er $\frac{53.85}{23.53} = 2.289$, hvilket vi også finder i den nederste del af outputtet. Her får vi tillige konfidensgrænser på, og da disse er (0.90, 5.79), må vi konkludere, at det ikke er helt usandsynligt, at der faktisk er 5 gange så stor risiko for yderligere vægttab i kontrolgruppen, sammenlignet med FT-gruppen. Det kan dog omvendt også godt tænkes, at der slet ingen forskel er, da den nedre grænse i konfidensintervallet er mindre end 1.

Alligevel giver χ^2 -testet en P-værdi på 0.0488, altså signifikans på 5%-niveauet, men ser vi på Fishers eksakte test, finder vi P-værdien 0.0637, som passer bedre med konklusionen baseret på relativ risiko. Dette er et eksempel på en lille diskrepans i konklusionen på χ^2 , Fisher, samt selve udregningen af CI for den relative risiko (da der benyttes forskellige approksimationer).

Bemærk konfidensgrænserne for risikoen for vægttab i hver af de to grupper: For kontrolgruppen med 26 piger og en ret stor risiko for vægttab stemmer de approksimative og de eksakte konfidensgrænser pænt overens, men for den noget mindre FT-gruppe, med lavere risiko for vægttab er der en klar forskel, idet de approksimative grænser er for lave.

4. *Lav en figur af vægtændringerne som funktion af udgangsværdien, med symboler svarende til behandlingen, og tilføj de tilhørende regressionsslinier.*

Her laves tegninger på to måder, dels hver gruppe for sig, og dels et fælles plot, hvor man bedre kan sammenligne linierne.

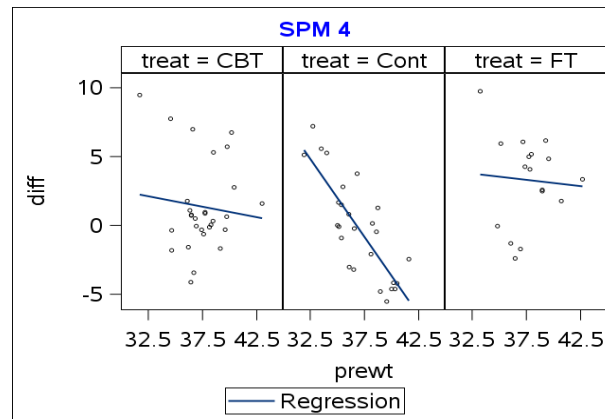
Først et såkaldt *panel*:

```
title 'SPM 4';
title2;

proc sgpanel data=a1;
  panelby treat / rows=1;
  reg Y=diff X=prewt;
```

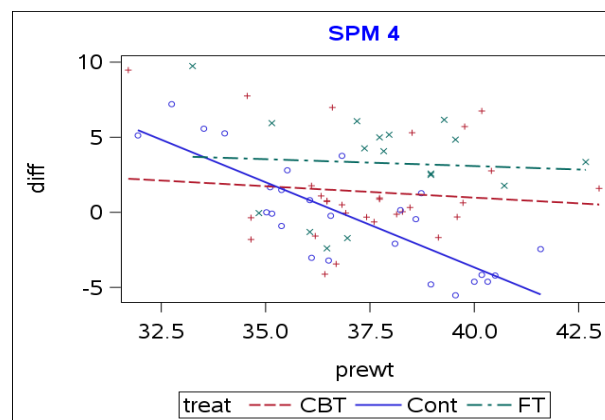
```
run;
```

som giver de opdelte figurer



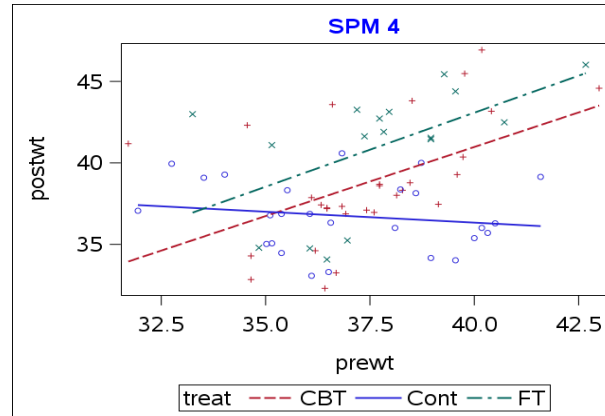
og så figuren med alle tre grupper på samme tegning:

```
proc sgplot data=a1;
    reg Y=diff X=prewt / group=treat;
run;
```



For kontrolgruppen ser det ud til, at pigerne, der vejer mindst fra starten, tager på, medens de, der vejer mest fra starten, taber sig. Hvis man tegner slutvægten op mod udgangsvægten ser man, at denne

i kontrolgruppen faktisk er næsten helt uafhængig af startvægten:



Omvendt finder vi for de to aktive behandlingsgrupper, at det er vægtændringen, der ser ud til kun at afhænge svagt af udgangsvægten, medens vægten ved afslutning af behandling derfor afhænger af startvægten.

- (a) *Ser vægtændringen ud til at være afhængig af udgangspunktet? I så fald i hvilken retning? Og kan det forstås?*

Vi kommenterede allerede på dette ovenfor, så her skal vi “bare” diskutere, om man kan forstå mønsteret.....

Man kunne forestille sig, at kontrolgruppen dækkede over “sædvanlig” behandling, som måske er mest intensiv for de svageste.

- (b) *Er der grund til at tro, at forskellen på effekten af behandlingerne afhænger af udgangsværdien?*

Hvis vi tager effekten af behandlingerne til at betyde vægtændringer, ser det ud til, at forskellen på FT og CBT er ret konstant, medens kontrolbehandlingen kun ser ud til at kunne fungere for de allermest undervægtige piger.

Denne forskel i effekt ses af de ikke-parallele linier i plottet ovenfor. Vi kan teste, om denne interaktion er signifikant ved hjælp af nedenstående model, der består af 3 vilkårlige (ikke-parallele) linier:

```

title2 'SPM 4B';

proc glm plots=ALL data=a1;
  class treat(ref="Cont");
model diff=treat prewt treat*prewt / solution clparm;
run;

```

Dette producerer outputtet

```

SPM 4B

The GLM Procedure

Class Level Information

Class          Levels   Values
treat           3      CBT FT Cont

Number of Observations Read      72
Number of Observations Used      72

Dependent Variable: diff

Source          DF          Sum of Squares    Mean Square  F Value  Pr > F
Model            5      344.9575854      68.9915171    7.80    <.0001
Error            66      583.7886591       8.8452827
Corrected Total  71      928.7462444

R-Square      Coeff Var    Root MSE    diff Mean
0.371423      237.5059      2.974102      1.252222

Source          DF      Type I SS    Mean Square  F Value  Pr > F
treat            2      126.2443628      63.1221814    7.14    0.0016
prewt            1      123.0653437      123.0653437   13.91    0.0004
prewt*treat      2       95.6478789      47.8239394    5.41    0.0067

Source          DF      Type III SS    Mean Square  F Value  Pr > F
treat            2      82.11969808      41.05984904    4.64    0.0130
prewt            1      74.39271321      74.39271321    8.41    0.0051
prewt*treat      2      95.64787887      47.82393943    5.41    0.0067

Parameter              Estimate      Standard Error    t Value    Pr > |t|
Intercept              41.68621837 B      8.51875817      4.89    <.0001
treat CBT              -34.62924843 B      12.83758297     -2.70    0.0089
treat FT               -34.95039844 B      15.01016923     -2.33    0.0230
treat Cont              0.00000000 B      .              .        .
prewt                  -1.13383337 B      0.23002877     -4.93    <.0001
prewt*treat CBT        0.98183421 B      0.34414139      2.85    0.0058
prewt*treat FT         1.04247065 B      0.39999167      2.61    0.0113
prewt*treat Cont       0.00000000 B      .              .        .

Parameter              95% Confidence Limits
Intercept              24.67796892  58.69446783
treat CBT              -60.26030981 -8.99818705
treat FT               -64.91916824 -4.98162864
treat Cont              .              .
prewt                  -1.59310063 -0.67456612
prewt*treat CBT        0.29473375  1.66893468
prewt*treat FT         0.24386150  1.84107979
prewt*treat Cont       .              .

NOTE: The X'X matrix has been found to be singular, and a generalized
      inverse was used to solve the normal equations. Terms whose
      estimates are followed by the letter 'B' are not uniquely estimable.

```

hvoraf vi tydeligt ser, at der er signifikant interaktion ($P=0.0067$), samt at denne kommer til udtryk ved hældninger, der er hhv 0.98

og 1.04 større for de to aktive behandlingsgrupper i forhold til kontrolgruppen (hvis hældning er så tilpas negativ, at det svarer til, at de to andre grupper har en hældning nær 0).

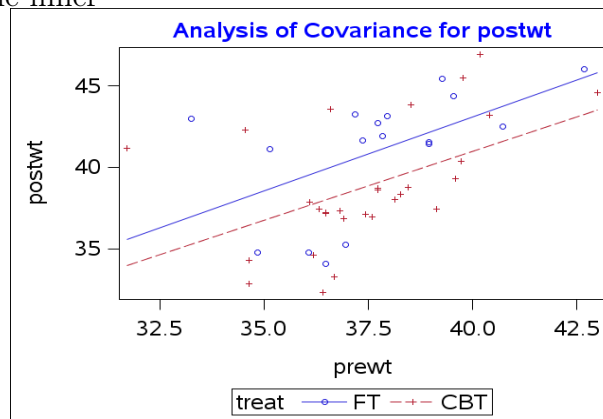
- (c) *Sammenlign de to aktive behandlingsgrupper, justeret for forskelle i udgangsværdi, og brug modellen til at svare på følgende:
Hvis 2 piger på 37 kg starter på hver sin behandling, hvilken forventet vægt vil de så hver især have efter endt behandling?
Og hvor stor vil forskellen være? (Husk konfidensinterval).*

Her er det vægten efter endt behandling (**postwt**), vi skal fokusere på, og når vi kun ser på de to aktive behandlingsgrupper, forventer vi at se parallelle linier, altså ingen interaktion. Vi checker lige og laver samtidig de automatiske plots i form af predikterede værdier og diagnostics plot:

```
title2 'SPM 4C';

proc glm plots=ALL data=a1; where treat in ("CBT","FT");
  class treat(ref="CBT");
  model postwt=treat prewt treat*prewt / solution clparm;
run;
```

På tegningen af de predikterede værdier nedenfor ses meget parallelle linier



og outputtet bliver:

SPM 4C

The GLM Procedure

Class Level Information

Class	Levels	Values
treat	2	FT CBT

Number of Observations Read	46
Number of Observations Used	46

Dependent Variable: postwt

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	215.8906260	71.9635420	6.41	0.0011
Error	42	471.3848892	11.2234497		
Corrected Total	45	687.2755152			

R-Square	Coeff Var	Root MSE	postwt Mean
0.314125	8.454802	3.350142	39.62413

Source	DF	Type III SS	Mean Square	F Value	Pr > F
treat	1	0.0037240	0.0037240	0.00	0.9856
prewt	1	158.1375869	158.1375869	14.09	0.0005
prewt*treat	1	0.1884247	0.1884247	0.02	0.8975

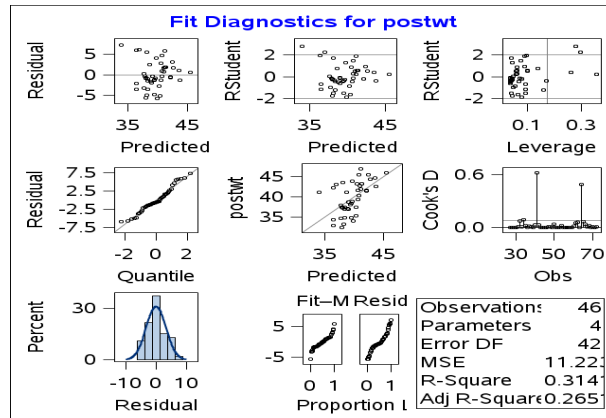
Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	7.056969944 B	10.81816352	0.65	0.5177
treat FT	-0.321150008 B	17.63047379	-0.02	0.9856
treat CBT	0.000000000 B	.	.	.
prewt	0.848000839 B	0.28833309	2.94	0.0053
prewt*treat FT	0.060636435 B	0.46798069	0.13	0.8975
prewt*treat CBT	0.000000000 B	.	.	.

Parameter	95% Confidence Limits	
Intercept	-14.77496792	28.888907810
treat FT	-35.90088658	35.258586562
treat CBT	.	.
prewt	0.266121115	1.429880564
prewt*treat FT	-0.883786827	1.005059697
prewt*treat CBT	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Vi ser, at testet for interaktion giver en P-værdi på 0.90, og dermed ingen indikation af ikke-parallellitet.

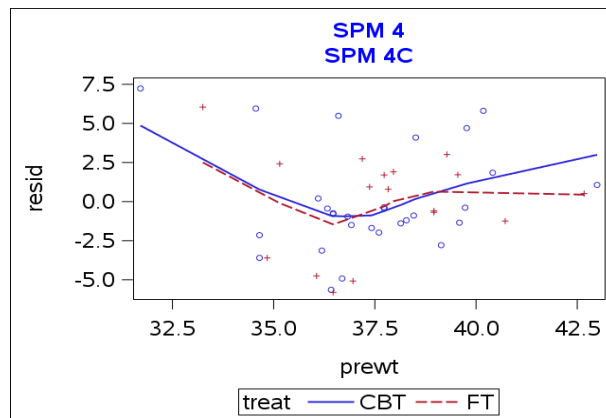
Modelkontrollen ses på figureren nedenfor, og det ser i det store hele tilforladeligt ud. Bemærk dog de to stærkt indflydelsesrige observationer (Cooks-figuren), som vi skal vende tilbage til senere.



Vi vil også lige checke lineariteten i `prewt`, og dette gøres ud fra de gemte residualer i datasættet `ny`:

```
proc sgplot data=ny;
    loess Y=resid X=prewt / group=treat;
run;
```

som giver os nedenstående figur:

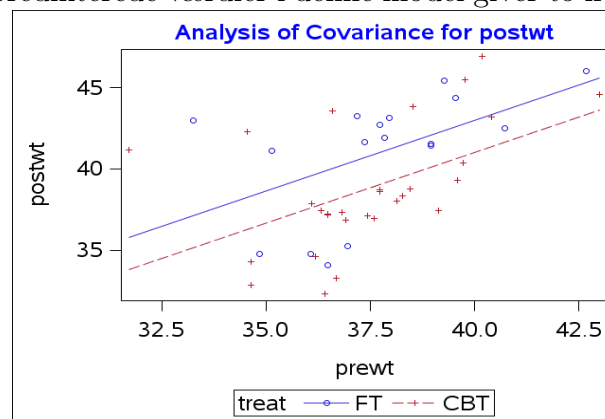


Denne giver anledning til lidt panderynken, da der optræder en bueform. Denne skyldes dog hovedsagelig de samme to observationer, der ovenfor nævnes som indflydelsesrige, så det vil blive kommenteret på nedenfor.

Da der ikke spores nogen interaktion, fortsætter vi med en model, der antager parallelle linier, altså det, der går under betegnelsen kovariansanalyse. I en sådan model er forskellen på de to grupper konstant, men for at estimere slutvægten for piger, der starter på 37 kg, inkluderer vi to `estimate`-sætninger nedenfor:

```
proc glm plots=ALL data=a1; where treat in ("CBT","FT");
  class treat(ref="CBT");
model postwt=treat prewt / solution clparm;
estimate "FT 37 kg" intercept 1 treat 1 0 prewt 37;
estimate "CBT 37 kg" intercept 1 treat 0 1 prewt 37;
output out=ny p=pred r=resid cookd=cook;
run;
```

De predikterede værdier i denne model giver to helt parallel linier:



og outputtet bliver

SPM 4C

The GLM Procedure

Class Level Information

Class	Levels	Values
treat	2	FT CBT

Number of Observations Read	46
Number of Observations Used	46

Dependent Variable: postwt

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	215.7022013	107.8511006	9.83	0.0003
Error	43	471.5733139	10.9668213		

```

Corrected Total          45    687.2755152

R-Square      Coeff Var      Root MSE    postwt Mean
0.313851      8.357582      3.311619      39.62413

Source                DF    Type III SS    Mean Square    F Value    Pr > F
treat                  1     41.0240189     41.0240189      3.74    0.0597
prewt                  1    165.0915668    165.0915668     15.05    0.0004

Parameter              Estimate      Standard      t Value    Pr > |t|
Intercept              6.194772865 B     8.43147320      0.73      0.4665
treat      FT          1.959374374 B     1.01306825      1.93      0.0597
treat      CBT          0.000000000 B      .            .            .
prewt              0.871018795      0.22449438      3.88      0.0004

Parameter              95% Confidence Limits
Intercept              -10.80891336  23.198459086
treat      FT          -0.083672462  4.002421211
treat      CBT          .            .
prewt              0.418282740  1.323754851

```

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Af dette output kan vi direkte aflæse forskellen på de to behandlingstyper til 1.959 i FT-gruppens favør, med CI=(-0.084, 4.002) og P=0.0597, altså ikke helt overbevisende, at der overhovedet er en forskel, men at vi heller ikke kan udelukke en forskel på op til ca. 4 kg!

Fra de to `estimate`-sætninger får vi resultatet

```

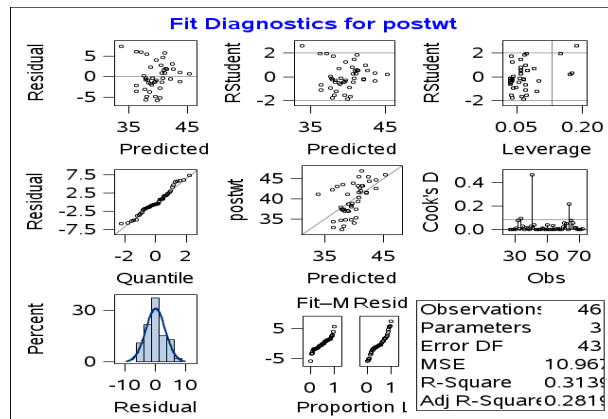
Parameter              Estimate      Standard      t Value    Pr > |t|
FT 37 kg              40.3818427     0.81854137     49.33      <.0001
CBT 37 kg              38.4224683     0.62347316     61.63      <.0001

Parameter              95% Confidence Limits
FT 37 kg              38.7310967  42.0325887
CBT 37 kg              37.1651148  39.6798218

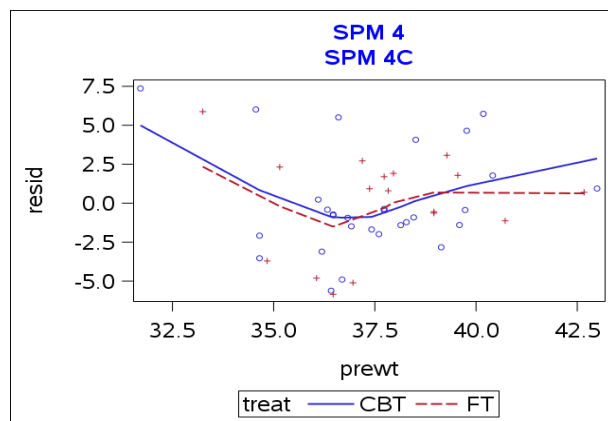
```

så hvis en pige starter med at veje 37 kg, forventer vi altså, at hun efter CBT-behandling vil veje omkring 38.4 kg, men hele 40.4 kg, hvis hun havde været i FT-behandling.

Modelkontroltegningerne er



og



Vi ser stadig, at der er to observationer, der har en rigtig stor værdi af Cooks afstand og for at finde ud af, hvad det er for nogle, har vi i koden ovenfor inkluderet en `output`-sætning, som skriver disse Cooks afstande ud i et datasæt `ny`, som vi herefter benytter således:

```
proc print data=ny;
  where cook>0.2;
  var idnr treat prewt postwt diff cook;
run;
```

og får udskriften

SPM 4C

Obs	idnr	treat	prewt	postwt	diff	cook
-----	------	-------	-------	--------	------	------

15	41	CBT	31.71	41.18	9.47	0.46391
38	64	FT	33.25	42.99	9.74	0.21766

Heraf kan vi se, at det er de to piger, der fra starten er allermest undervægtige, der også har den største indflydelse på estimationen. Man kunne med rette argumentere for, at disse piger er så ekstreme, at de burde udelukkes fra analysen, da de dels har stor indflydelse og da vi endvidere ikke med særlig stor overbevisning kan udtale os om så undervægtige piger, når der er så få af dem. Vi skal dog ikke forlænge opgavebesvarelsen med at gøre dette.

- (d) *Samme spørgsmål som ovenfor, men denne gang for de to udvalgte grupper fra spørgsmål 1c.*

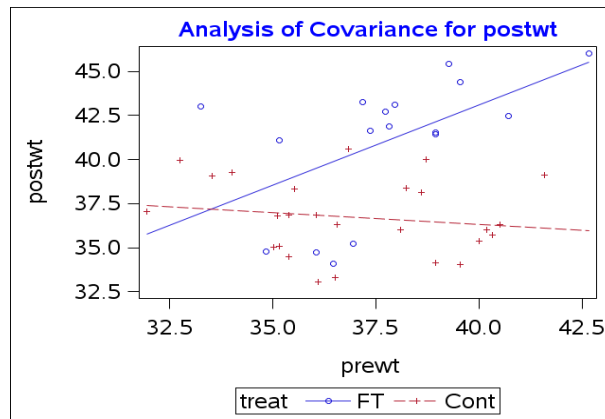
Først checker vi, om der er interaktion mellem `treat` og `prewt`, når vi sammenligner FT og Cont (hvilket vi ganske bestemt forventer ud fra den tidligere figur med 3 regressionslinier.

Vi inkluderer desuden nogle ganske indviklede `estimate`-sætninger, som vil blive kommenteret nedenfor.

```
title2 'SPM 4D';

proc glm plots=ALL data=a1; where treat in ("Cont","FT");
  class treat(ref="Cont");
model postwt=treat prewt treat*prewt / solution clparm;
run;
```

De predikterede værdier ser således ud:



og viser klart interaktionen, og outputtet bekræfter dette:

```
SPM 4C

The GLM Procedure

Class Level Information

Class      Levels      Values
treat      2          FT Cont

Number of Observations Read      43
Number of Observations Used      43

Dependent Variable: postwt

Source      DF      Sum of Squares      Mean Square      F Value      Pr > F
Model        3      257.0852295      85.6950765      11.94      <.0001
Error       39      279.9306356      7.1777086
Corrected Total      42      537.0158651

R-Square      Coeff Var      Root MSE      postwt Mean
0.478729      6.972749      2.679125      38.42279

Source      DF      Type III SS      Mean Square      F Value      Pr > F
treat        1      47.95620450      47.95620450      6.68      0.0136
prewt        1      33.18894225      33.18894225      4.62      0.0378
prewt*treat  1      60.08104629      60.08104629      8.37      0.0062

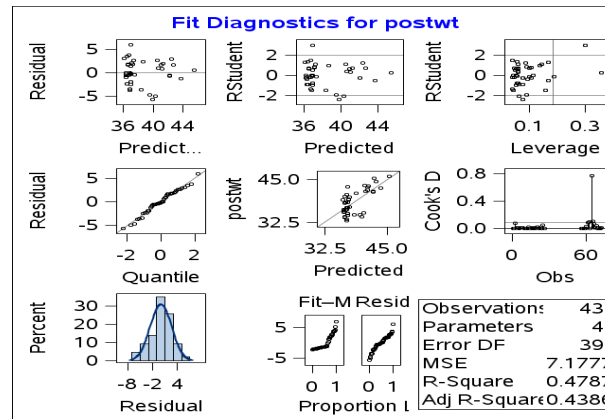
Parameter      Estimate      Standard Error      t Value      Pr > |t|
Intercept      41.68621837 B      7.67385064      5.43      <.0001
treat          FT      -34.95039844 B      13.52143052      -2.58      0.0136
treat          Cont      0.00000000 B      .      .      .
prewt          FT      -0.13383337 B      0.20721406      -0.65      0.5221
prewt*treat    FT      1.04247065 B      0.36031970      2.89      0.0062
prewt*treat    Cont      0.00000000 B      .      .      .

Parameter      95% Confidence Limits
Intercept      26.16439036      57.20804638
treat          FT      -62.30007318      -7.60072369
treat          Cont      .      .
prewt          FT      -0.55296336      0.28529661
prewt*treat    FT      0.31365527      1.77128603
prewt*treat    Cont      .      .

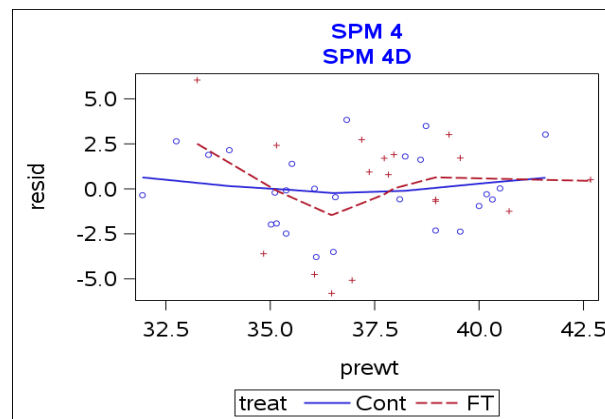
NOTE: The X'X matrix has been found to be singular, and a generalized
inverse was used to solve the normal equations. Terms whose
estimates are followed by the letter 'B' are not uniquely estimable.
```

Her ser vi som forventet en klar indikation af interaktion, dvs. at linierne *ikke* er parallelle ($P=0.0062$).

Modelkontroltegningerne er:



og



Vi bemærker også en enkelt meget indflydelsesrig pige, og ved at benytte samme metode som ovenfor, finder vi udskriften

SPM 4D

Obs	idnr	treat	prewt	postwt	diff	cook
35	64	FT	33.25	42.99	9.74	0.77302

Det ses, at den indflydelsesrige person nu er en pige i FT-gruppen, der tager rigtigt meget på, næsten 10 kg. Man bør nok checke, om dette kan have sin rigtighed....

Når vi har at gøre med en model, der inkluderer et interaktionsled, er det noget mere besværligt at kvantificere forskellen for piger på 37 kg (og også slutresultatet for hver af de to behandlingsgrupper), men vi kan nemt benytte tricket med at lægge Y-aksen hen i 37, nemlig ved at definere `prewt37=prewt-37`; og benytte denne som kovariat i stedet for blot `prewt`. For at estimere begge niveauer, inkluderes tillige et par `estimate`-sætninger (hvoraf den ene faktisk er overflødig, da den helt gratis kommer ud som interceptet for reference-gruppen):

```
title2 'SPM 4D';

proc glm plots=ALL data=a1; where treat in ("Cont","FT");
  class treat(ref="Cont");
model postwt=treat prewt37 treat*prewt37 / solution clparm;
estimate "FT 37 kg" intercept 1 treat 1 0;
estimate "Cont 37 kg" intercept 1 treat 0 1;
run;
```

Bemærk, at vi nu i `estimate`-sætningerne *ikke* har noget interaktionsled, idet vi nu har lagt nulpunktet hen i `prewt=37`. Vi finder outputtet:

The GLM Procedure

Class Level Information

Class	Levels	Values
treat	2	FT Cont

Number of Observations Read	43
Number of Observations Used	43

Dependent Variable: postwt

Parameter	Estimate	Standard Error	t Value	Pr > t
FT 37 kg	40.3553991	0.68202205	59.17	<.0001
Cont 37 kg	36.7343835	0.52553633	69.90	<.0001

Parameter	95% Confidence Limits	
FT 37 kg	38.9758793	41.7349189
Cont 37 kg	35.6713860	37.7973811

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	36.73438352 B	0.52553633	69.90	<.0001

treat	FT	3.62101557	B	0.86101250	4.21	0.0001
treat	Cont	0.00000000	B	.	.	.
prewt37		-0.13383337	B	0.20721406	-0.65	0.5221
prewt37*treat	FT	1.04247065	B	0.36031970	2.89	0.0062
prewt37*treat	Cont	0.00000000	B	.	.	.

Parameter		95% Confidence Limits	
Intercept		35.67138595	37.79738109
treat	FT	1.87945341	5.36257773
treat	Cont	.	.
prewt37		-0.55296336	0.28529661
prewt37*treat	FT	0.31365527	1.77128603
prewt37*treat	Cont	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Ud fra dette output kan vi direkte aflæse forskellen til 3.62 i FT-gruppens favør, med et CI=(1.88, 5.36), og fra `estimate`-sætningerne aflæses de to opnåede vægte som 36.7 kg, CI=(35.7, 37.8) resp. 40.4 kg, CI=(39.0, 41.7).

Hvis vi skulle undgå tricket med at flytte nulpunktet og benytte `estimate`-sætninger i den oprindelige modelformulering i stedet for, skulle de se således ud:

```
proc glm plots=ALL data=a1; where treat in ("Cont","FT");
  class treat(ref="Cont");
model postwt=treat prewt treat*prewt / solution clparm;
estimate "FT 37 kg" intercept 1 treat 1 0 prewt 37 treat*prewt 37 0;
estimate "Cont 37 kg" intercept 1 treat 0 1 prewt 37 treat*prewt 0 37;
estimate "FT-Cont, 37 kg" treat 1 -1 treat*prewt 37 -37;
run;
```

og vi ville så få outputtet

Parameter	Estimate	Standard Error	t Value	Pr > t
FT 37 kg	40.3553991	0.68202205	59.17	<.0001
Cont 37 kg	36.7343835	0.52553633	69.90	<.0001
FT-Cont, 37 kg	3.6210156	0.86101250	4.21	0.0001

Parameter	95% Confidence Limits	
FT 37 kg	38.9758793	41.7349189
Cont 37 kg	35.6713860	37.7973811
FT-Cont, 37 kg	1.8794534	5.3625777

- (e) *Sammenlign resultaterne i spørgsmål 1c, 2c og 4d og formuler en konklusion om effekten af behandlingerne.*

Hvis vi går ud fra, at fokus fra protokollens side er på sammenligningen af FT og Control, så vil de mest fornuftige resultater være at finde fra spørgsmål 4d og til en vis grad 2c.

Vi må altså konkludere, at FT ser ud til at virke bedre end Control, med en øget vægtforøgelse på 3.62 kg, med CI=(1.88, 5.36).

I 2c, hvor vi ser direkte på sammenligning af vægtforøgelserne, uden at korrigere for baseline vægten, får vi noget meget tæt på, nemlig en øget vægtforøgelse for FT vs. Control på 3.50 kg, med CI=(1.31, 5.69).