

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)

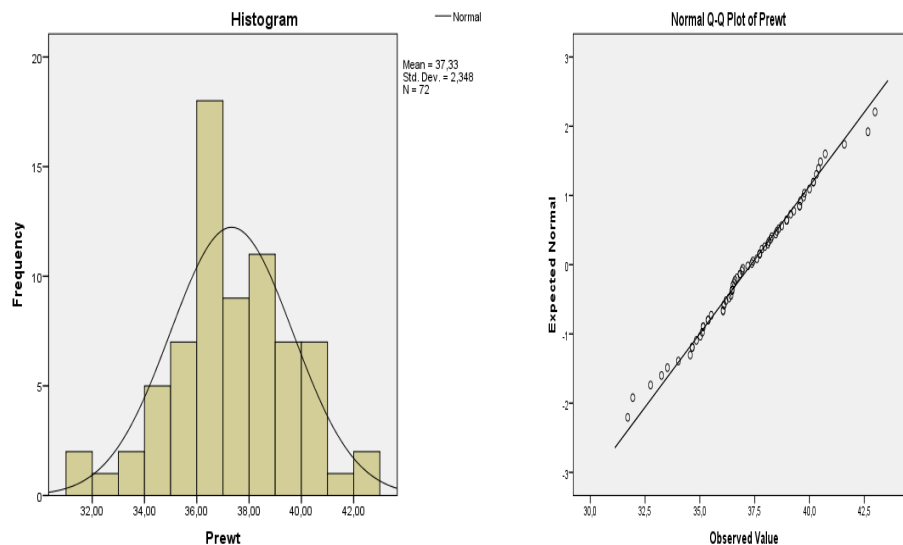
Succeskriteriet for behandlingerne er, at pigerne skal tage mindst 3 kg på i løbet af den tid, behandlingen varer, og opgaven er at undersøge, hvorvidt dette mål blev opnået, samt generelt at vurdere effekten af de 3 behandlinger og sammenligne disse.

I nedenstående besvarelse vil analyserne blive forklaret med menuer, første gang de optræder, men efterfølgende må man blade tilbage for at se vejledningen. Ellers bliver besvarelsen alt for tung at læse.

Vi indlæser først data i SPSS ved hjælp af menuen **File/Open/Data...**, som giver os de samme variabelbetegnelser som angivet ovenfor.

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 enkelt histogram, f.eks. ved hjælp af **Analyze/Descriptive Statistics/Frequencies**. Sæt **prewt** over i **Variable(s)** og vælg **Charts**, hvor man afkrydser **Histogram** og **Show normal curve on histogram**. For at undgå for meget udskrift, fjernes fluebenet i **Display frequency tables**. Herved får vi venstre side af figuren nedenfor:



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

For en ordens skyld (fordi vi lige om lidt skal udregne et referenceområde), vil vi dog lige checke hvordan fraktildiagrammet ser ud, idet et sådant er bedre til at vurdere eventuelle afvigelser fra normalfordelingen. Denne fås via **Analyze/Descriptive Statistics/Explore**: Sæt **prewt** over i **Dependent List**, og vælg **Plots**, hvor man afkrydser **Normality plots with tests**. For at undgå en ekstra figur, fjernes fluebenet i **Stem-and-leaf**. Dette fraktildiagram ses ovenfor og bekræfter en pæn normalfordelingstilpasning.

- (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.

Vi fik godt nok også sådanne ved brug af **Explore** ovenfor, men for at få lidt flere størrelser udregnet, benytter vi i stedet **Analyze/Descriptive Statistics/Frequencies** igen. Her går vi nu ind under **Statistics** og vælger de ønskede størrelser, her **Mean, Median, Quartiles, Std. deviation, Minimum** og **Maximum**. Ved af afkrydse **Percentiles** kan vi endvidere bede om 2.5%- og 97.5%-fraktilen (skriv f.eks. 2,5 og klik **Add**), som vi skal bruge nedenfor.

Vi finder så tabellen nedenfor:

Statistics		
Prewt		
N	Valid	72
	Missing	0
Mean		37,3310
Median		37,2800
Std. Deviation		2,34800
Minimum		31,71
Maximum		42,99
Percentiles	2,5	31,8998
	25	36,0600
	50	37,2800
	75	38,9600
	97,5	42,7260

Vi ser her, at gennemsnit og median er næsten sammenfaldende, samt at 25- og 75%-fraktilerne (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 sammenligne til de empiriske fraktiler, som også fremgår af outputtet ovenfor. Vi skal have fat i 2.5%-fraktilen (som ses at være 31.90) samt 97.5%-fraktilen (som ses at være 42.73).

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

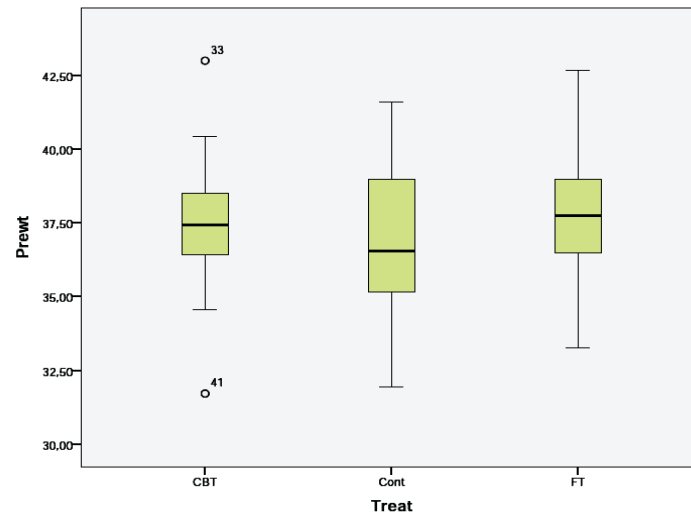
(b) *Ser det ud som om der er blevet randomiseret fornuftigt?*

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: Brug **Graph/Chart Builder** og vælg det simple Boxplot, hvor **prewt** sættes på Y-aksen og **Treat** på X-aksen) og dels ved at udregne simple summary statistics, opdelt på de tre randomiseringsgrupper (brug **Data/Split File**, sæt flueben i **Compare Groups** og sæt **Treat** over i **Groups based on:**, hvorefter der benyttes **Analyze/Descriptive Statistics/Descriptives**, hvor **prewt** sættes over i **Variable(s)**.

For at få pæne akser på Boxplottet (for at forhindre Y-aksen i at gå helt ned til 0), var jeg dog nødt til at gå ind i **Element Properties**, klikke på Y-aksen, fjerne fluebenet fra **Minimum:Automatic** og i stedet skrive 30 i **Custom**.

Vi får plottet



som synes at vise 3 rimeligt ens grupper.

De opdelte summary statistics ser således ud:

Descriptive Statistics						
Treat		N	Minimum	Maximum	Mean	Std. Deviation
CBT	Prewt	29	31,71	42,99	37,4576	2,19578
	Valid N (listwise)	29				
Cont	Prewt	26	31,94	41,59	36,9465	2,58585
	Valid N (listwise)	26				
FT	Prewt	17	33,25	42,67	37,7029	2,27217
	Valid N (listwise)	17				

og tyder ligeledes på, at randomiseringen er forløbet tilfredsstillende (ikke den store forskel hen over grupperne).

(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.

Inden vi går i gang med en ANOVA, må vi ind i **Data/Split File** og skifte fluebenet tilbage til **Analyze all cases**,... Herefter benytter vi **Analyze/General Linear Model/Univariate**, hvor vi sætter **prewt** i **Dependent Variable** og **Treat** i **Fixed Factor(s)**. Desuden går vi ind i **Options** og afkrydser **Parameter Estimates**, **Homogeneity tests** og evt. nogle plots (**Residual plot** er det eneste nogenlunde rimelige her, men det er ikke flot).

For at få Tukey-korrigerede differenser mellem grupperne parvist, går vi endvidere ind under **Post Hoc**, hvor vi sætter **Treat** over i **Post Hoc tests for** og afkrydser **Tukey**.

Herved finder vi outputtet nedenfor, men bemærk, at vi kun er interesserede i kvantificeringerne, ikke i test og P-værdier. Vi bemærker dog lige, at Levenes test ikke giver anledning til at betvivle rimeligt ens varianser i de 3 grupper ($P=0.32$).

Levene's Test of Equality of Error Variances^a

Dependent Variable: Prewt

F	df1	df2	Sig.
1,160	2	69	,320

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

a. Design: Intercept + Treat

Tests of Between-Subjects Effects

Dependent Variable: Prewt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6,660 ^a	2	3,330	,597	,553
Intercept	95379,840	1	95379,840	17104,211	,000
Treat	6,660	2	3,330	,597	,553
Error	384,771	69	5,576		
Total	100730,738	72			
Corrected Total	391,431	71			

a. R Squared = ,017 (Adjusted R Squared = -,011)

Parameter Estimates

Dependent Variable: Prewt

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	37,703	,573	65,830	,000	36,560	38,846
[Treat=CBT]	-,245	,721	-,340	,735	-1,684	1,194
[Treat=Cont]	-,756	,737	-1,027	,308	-2,226	,713
[Treat=FT]	0 ^a	.	.	.	.	.

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

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.

De Tukey-korrigerede parvise sammenligninger bliver:

Post Hoc Tests

Treat

Multiple Comparisons

Dependent Variable: Prewt

Tukey HSD

(I) Treat	(J) Treat	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
CBT	Cont	,5110	,63778	,703	-1,0166	2,0387
	FT	-,2454	,72133	,938	-1,9732	1,4824
Cont	CBT	-,5110	,63778	,703	-2,0387	1,0166
	FT	-,7564	,73655	,563	-2,5207	1,0079
FT	CBT	,2454	,72133	,938	-1,4824	1,9732
	Cont	,7564	,73655	,563	-1,0079	2,5207

Based on observed means.

The error term is Mean Square(Error) = 5,576.

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.

Her anvender vi `Data/Select Cases` og afkrydser `If condition is satisfied`, hvorefter der i det store felt skrives `Treat='FT' or Treat='Cont'`.

Herefter kan vi lave et T-test ved hjælp af

`Analyze/Compare Means/Independent-Samples T Test`, hvor vi vælger `prewt` som `Test variable(s)` og `Treat` som `grouping Variable`.

Man er også nødt til at klikke på `Define Groups` og sætte f.eks.

FT som `group 1` og Cont som `Group 2` (dette har kun betydning for fortegnene). Herved får vi:

Group Statistics

	Treat	N	Mean	Std. Deviation	Std. Error Mean
Prewt	FT	17	37,7029	2,27217	,55108
	Cont	26	36,9465	2,58585	,50713

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Prewt	Equal variances assumed	1,324	,257	,983	41
	Equal variances not assumed			1,010	37,405

Independent Samples Test

		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence ... Lower
Prewt	Equal variances assumed	,332	,75640	,76984	-,79832
	Equal variances not assumed	,319	,75640	,74891	-,76048

Independent Samples Test

		t-test for Equality of Means
		95% Confidence Interval of the ... Upper
Prewt	Equal variances assumed	2,31113
	Equal variances not assumed	2,27329

Nedenfor sammenfatter vi de forskellige kvantificeringer af forskellene mellem de to grupper, FT og Cont, nu reduceret til to deci-

maler:

	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

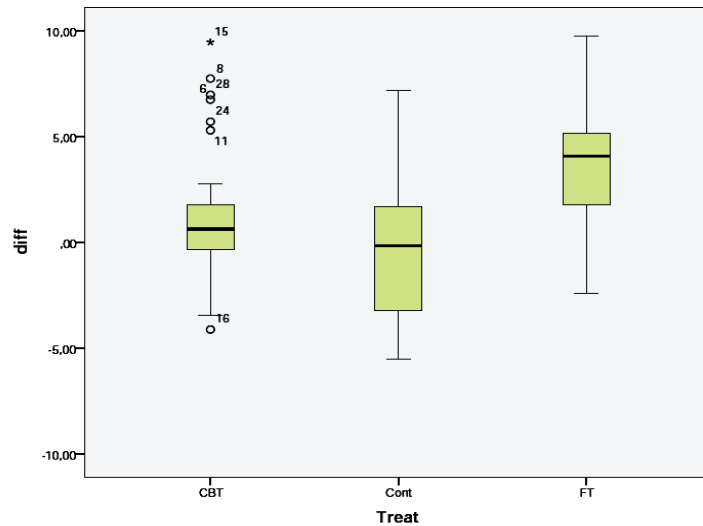
Bemærk, at vi i alle tilfælde får samme estimat for forskellen (nemlig differensen mellem de to gennemsnit), men at konfidensintervallerne varierer, idet de er bredere, når man Tukey-korrigerer, og sædvanligvis også, når man ikke kan antage ens spredninger (i T-testet). Det sidste er dog ikke tilfældet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

Husk at gå tilbage til alle data igen efterfølgende
(benyt `Data/Select Cases` og afkryds `All cases`).

2. *Kvantificer vægtændringen i hver af de 3 grupper.*

Vi udregner nu (med `Transform/Compute variable`) for hver pige en vægtøgning `diff=postwt-prewt`, som i parentes bemærket kan være negativ, hvis der er tale om et vægttab.

Først laver vi selvfølgelig en illustration, i form af et Boxplot af ændringerne.



Boxplottet tyder på en vis forskel på de 3 grupper, idet FT-gruppen ser ud til at klare sig bedst, og Cont-gruppen værst. Spredningen ser nogenlunde ens ud i de tre grupper.

Vi kvantificerer nu vægtændringen med estimat og konfidensinterval for middelværdien, for hver af de 3 grupper for sig. Dette gøres ved igen at benytte **Data/Split File** og dernæst benytte **Analyze/Compare Means/One-Sample T Test**, idet vi nu sætter **diff** som **Test Variable(s)**. Herved får vi samtidig et test for, om middelværdien er 0, dvs. om der er en signifikant vægtøgning (eller evt fald).

One-Sample Statistics

Treat		N	Mean	Std. Deviation	Std. Error Mean
CBT	diff	29	1,3634	3,31111	,61486
	diff_minus3	29	-1,6366	3,31111	,61486
Cont	diff	26	-,2050	3,61833	,70961
	diff_minus3	26	-3,2050	3,61833	,70961
FT	diff	17	3,2912	3,24246	,78641
	diff_minus3	17	,2912	3,24246	,78641

One-Sample Test

Test Value = 0						Test Value = 0	
Treat		t	df	Sig. (2-tailed)	Mean Difference	95% Confidence ... Lower	95% Confidence Interval of the ... Upper
CBT	diff	2,218	28	,035	1,36345	,1040	2,6229
	diff_minus3	-2,662	28	,013	-1,63655	-2,8960	-,3771
Cont	diff	-,289	25	,775	-,20500	-1,6665	1,2565
	diff_minus3	-4,517	25	,000	-3,20500	-4,6665	-1,7435
FT	diff	4,185	16	,001	3,29118	1,6241	4,9583
	diff_minus3	,370	16	,716	,29118	-1,3759	1,9583

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 er tilfældet for gruppen FT, men for at knytte P-værdier til (hvis man finder dette ønskværdigt), kan vi enten se på den nye variabel `diff_minus3=diff-3`, og så teste om middelværdien af denne er 0, som vi netop gjorde det ovenfor, eller man kan under `One-Sample T test` simpelthen ændre `Test Value` fra 0 til 3:

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 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 (`Analyze/General Linear Model/Univariate`), denne gang med `diff` som `Dependent Variable`, men igen med `Treat` som `Fixed Factor(s)`. For at få Tukey-korrigerede differenser mellem grupperne parvist, går vi endvidere ind under `Post Hoc`, hvor vi sætter `Treat` over i `Post Hoc tests for` og afkrydser `Tukey`. Husk endvidere at altid at afkrydse `Parameter estimates` og `Homogeneity tests` (Levenes test for varianshomogenitet) under `Options`.

Vi finder så outputtet

Levene's Test of Equality of Error Variances^a

Dependent Variable: diff

F	df1	df2	Sig.
,314	2	69	,732

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

a. Design: Intercept + Treat

Tests of Between-Subjects Effects

Dependent Variable: diff

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	126,244 ^a	2	63,122	5,427	,006
Intercept	150,258	1	150,258	12,919	,001
Treat	126,244	2	63,122	5,427	,006
Error	802,502	69	11,630		
Total	1041,647	72			
Corrected Total	928,746	71			

a. R Squared = ,136 (Adjusted R Squared = ,111)

Parameter Estimates

Dependent Variable: diff

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	3,291	,827	3,979	,000	1,641	4,941
[Treat=CBT]	-1,928	1,042	-1,851	,069	-4,006	,150
[Treat=Cont]	-3,496	1,064	-3,287	,002	-5,618	-1,374
[Treat=FT]	0 ^a	.	.	.	.	.

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

Testet for identitet af de tre vægtøgninger ses at give en P-værdi på 0.006, 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, som vi dog først viser her:

Multiple Comparisons							
Dependent Variable: diff							
	(I) Treat	(J) Treat	Mean Difference (I-J)	Std. Error	Sig.	95% ... Lower Bound	Confidence . Upper Bound
Tukey HSD	CBT	Cont	1,5684	,92107	,211	-,6378	3,7747
		FT	-1,9277	1,04173	,161	-4,4230	,5675
	Cont	CBT	-1,5684	,92107	,211	-3,7747	,6378
		FT	-3,4962*	1,06371	,004	-6,0441	-,9483
	FT	CBT	1,9277	1,04173	,161	-,5675	4,4230
		Cont	3,4962*	1,06371	,004	,9483	6,0441
LSD	CBT	Cont	1,5684	,92107	,093	-,2690	3,4059
		FT	-1,9277	1,04173	,069	-4,0059	,1505
	Cont	CBT	-1,5684	,92107	,093	-3,4059	,2690
		FT	-3,4962*	1,06371	,002	-5,6182	-1,3741
	FT	CBT	1,9277	1,04173	,069	-,1505	4,0059
		Cont	3,4962*	1,06371	,002	1,3741	5,6182

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

Group Statistics					
	Treat	N	Mean	Std. Deviation	Std. Error Mean
diff	FT	17	3,2912	3,24246	,78641
	Cont	26	-,2050	3,61833	,70961

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
diff	Equal variances assumed	,304	,584	3,224	41
	Equal variances not assumed			3,301	36,974

Independent Samples Test

		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence ... Lower
diff	Equal variances assumed	,002	3,49618	1,08434	1,30632
	Equal variances not assumed	,002	3,49618	1,05924	1,34990

Independent Samples Test

		t-test for Equality of Means
		95% Confidence Interval of the ... Upper
diff	Equal variances assumed	5,68604
	Equal variances not assumed	5,64246

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.

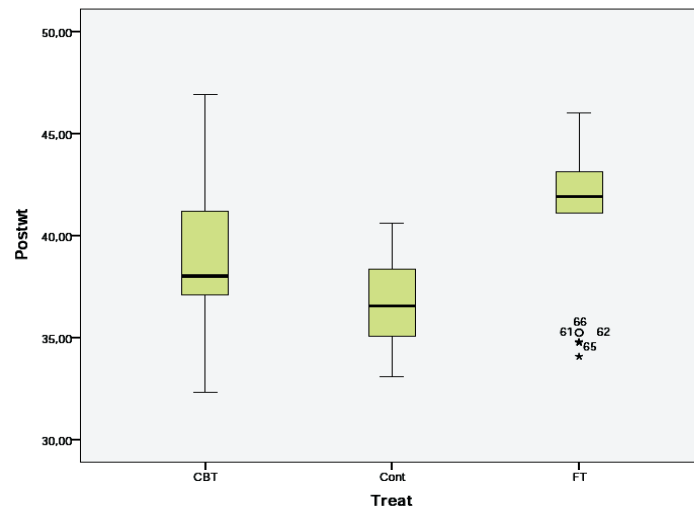
- (d) *Finder vi nogenslunde samme konklusion, når vi blot sammenligner slutvægten i de 3 (hvv 2) grupper?*

Vi ville forvente at finde nogenslunde 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:



Figuren synes klart 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 fra ANOVA:

Tests of Between-Subjects Effects

Dependent Variable: Postwt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	188,638 ^a	2	94,319	8,654	,000
Intercept	103101,512	1	103101,512	9459,322	,000
Treat	188,638	2	94,319	8,654	,000
Error	752,063	69	10,899		
Total	108124,429	72			
Corrected Total	940,701	71			

a. R Squared = ,201 (Adjusted R Squared = ,177)

Parameter Estimates

Dependent Variable: Postwt

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	40,994	,801	51,197	,000	39,397	42,592
[Treat=CBT]	-2,173	1,008	-2,155	,035	-4,185	-,161
[Treat=Cont]	-4,253	1,030	-4,130	,000	-6,307	-2,198
[Treat=FT]	0 ^a	.	.	.	.	.

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

med Tukey-korrektioner:

Multiple Comparisons

Dependent Variable: Postwt

		Mean			95% ...	Confidence .	
	(I) Treat	(J) Treat	Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Tukey HSD	CBT	Cont	2,0795	,89166	,058	-,0563	4,2153
		FT	-2,1731	1,00846	,086	-4,5887	,2425
	Cont	CBT	-2,0795	,89166	,058	-4,2153	,0563
		FT	-4,2526 [*]	1,02974	,000	-6,7191	-1,7860
	FT	CBT	2,1731	1,00846	,086	-,2425	4,5887
		Cont	4,2526 [*]	1,02974	,000	1,7860	6,7191

der også klart dokumenterer en forskel på de tre grupper (ANOVA

giver $P=0.0004$; de ekstra decimaler kan ses ved at dobbeltklikke på 0.000), med et klart forspring til FT-gruppen.

Ser man på Levenes test for identiske varianser i de 3 grupper:

Test of Homogeneity of Variances

Postwt

Levene Statistic	df1	df2	Sig.
2,924	2	69	,060

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:

Robust Tests of Equality of Means

Postwt

	Statistic ^a	df1	df2	Sig.
Welch	9,912	2	36,235	,000

a. Asymptotically F distributed.

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 s. 20??, hvor vi sammenligner til det T-test, vi nu foretager:

Group Statistics

	Treat	N	Mean	Std. Deviation	Std. Error Mean
Postwt	FT	17	40,9941	3,83835	,93094
	Cont	26	36,7415	2,14847	,42135

Independent Samples Test

		Levene's Test for Equality of Variances	t-test for Equality of Means		
		F	Sig.	t	df
Postwt	Equal variances assumed	4,961	,031	4,659	41
	Equal variances not assumed			4,162	22,619

Independent Samples Test

		t-test for Equality of Means			95% Confidence Interval of the Difference
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower Upper
Postwt	Equal variances assumed	,000	4,25258	,91277	2,40920
	Equal variances not assumed	,000	4,25258	1,02185	2,13675

Independent Samples Test

		t-test for Equality of Means	95% Confidence Interval of the Difference
		Upper	
Postwt	Equal variances assumed	6,09596	
	Equal variances not assumed	6,36841	

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 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)` (under Transform/Compute Variable) og laver så en lille tabel, hvor andelen af piger med vægtstigning kvantificeres, opdelt på de 3 grupper. Dette gøres under menuen Analyze/Descriptive Statistics, hvor `Treat` sættes i Row(s) og `stigning` i Column(s). For at få relevante procenter på, går vi endvidere ind i `Cells` og afkrydser Row under `Percentages`, hvorved vi får tabellen:

Treat * stigning Crosstabulation

			stigning		
			,00	1,00	Total
Treat	CBT	Count	11	18	29
		% within Treat	37,9%	62,1%	100,0%
	Cont	Count	15	11	26
		% within Treat	57,7%	42,3%	100,0%
	FT	Count	4	13	17
		% within Treat	23,5%	76,5%	100,0%
Total	Count	30	42	72	
	% within Treat	41,7%	58,3%	100,0%	

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 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 går derfor tilbage i **Crosstabs**-menuen fra før og erstatter `stigning` med `fald`.

Endvidere benytter vi **Statistics**-knappen for at afkrydse **Chi-square**, og under **Exact** afkrydser vi endvidere **Exact** for også at få Fishers eksakte test.

Vi finder så outputtet

Treat * fald Crosstabulation

			fald		
			,00	1,00	Total
Treat	CBT	Count	18	11	29
		% within Treat	62,1%	37,9%	100,0%
	Cont	Count	12	14	26
		% within Treat	46,2%	53,8%	100,0%
	FT	Count	13	4	17
		% within Treat	76,5%	23,5%	100,0%
Total	Count	43	29	72	
	% within Treat	59,7%	40,3%	100,0%	

Chi-Square Tests

	Value	df	Asymptotic Significance (2- sided)	Exact Sig. (2- sided)
Pearson Chi-Square	4,039 ^a	2	,133	,133
Likelihood Ratio	4,138	2	,126	,133
Fisher's Exact Test	3,934			,133
N of Valid Cases	72			

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6,85.

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 nu igen den tidligere anvendte **Data/Select Cases** til at udvælge de to grupper, vi gerne vil sammenligne (afkryds

Use `filter variable` og sæt den tidligere dannede Filter-variabel ind).

Herefter går vi tilbage i **Crosstabs**-menuen fra før og går nu ind under **Statistics** for at afkrydse **Risk**, som vil give os den relative risiko for vægttab for den ene gruppe i forhold til den anden. Herved får vi outputtet:

Treat * fald Crosstabulation					
Treat	Cont		fald		Total
			,00	1,00	
		Count	12	14	26
		% within Treat	46,2%	53,8%	100,0%
	FT	Count	13	4	17
		% within Treat	76,5%	23,5%	100,0%
Total		Count	25	18	43
		% within Treat	58,1%	41,9%	100,0%

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3,882 ^a	1	,049	,064	,048
Continuity Correction ^b	2,736	1	,098		
Likelihood Ratio	4,026	1	,045	,064	,048
Fisher's Exact Test				,064	,048
N of Valid Cases	43				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 7,12.
b. Computed only for a 2x2 table

Risk Estimate			
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Treat (Cont / FT)	,264	,068	1,028
For cohort fald = ,00	,604	,369	,987
For cohort fald = 1,00	2,288	,905	5,788
N of Valid Cases	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 diskrepans i konklusionen på χ^2 og Fisher, selv om der i-flg. tommelfingerreglen om forventede værdier ikke er noget formelt galt med at benytte χ^2 -testet.

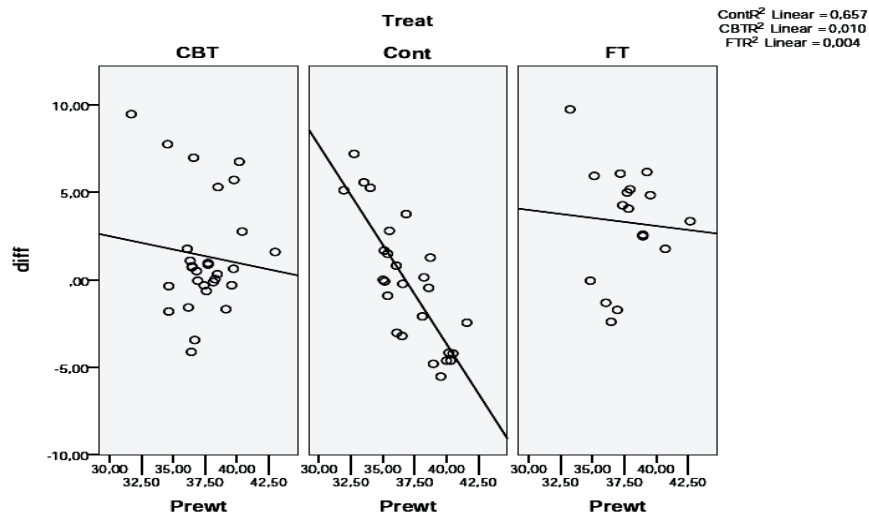
4. *Lav en figur af vægtændringerne som funktion af udgangsværdien, med symboler svarende til behandlingen, og tilføj de tilhørende regressionslinier.*

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*, altså de opdelte figurer, som dannes ved at gå ind i **Graphs/Chart Builder** og vælge det simple **Scatter/Dot**. Herefter vælges **Groups/Point ID** og der afkrydses i **Columns panel variable**. Herefter sættes **Treat** i **Panel**, **prewt** på X-aksen og **diff** på Y-aksen.

Herved fremkommer tre grafer, dog uden tilhørende regressionslinier. Disse tilføjes ved at dobbeltklikke på grafen og klikke på **Add Fit Line at Total**. Fjern evt også fluebenet ved **Attach Label to Line**, før der klikkes **Apply/Close**.

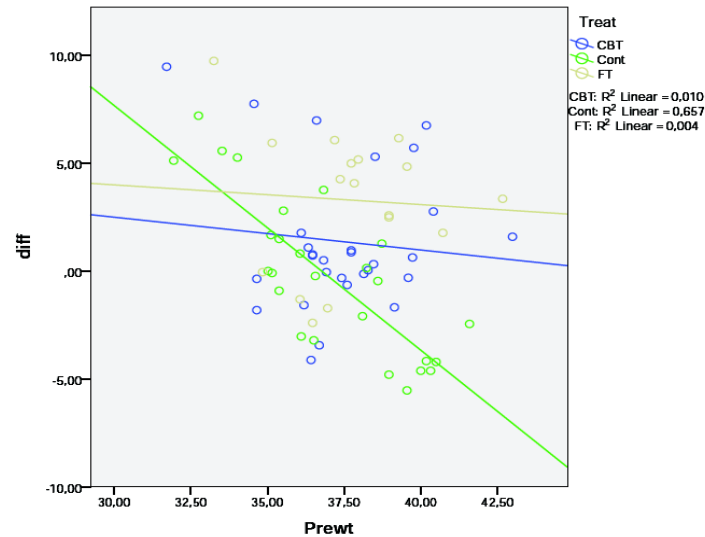
Herved fremkommer figuren:



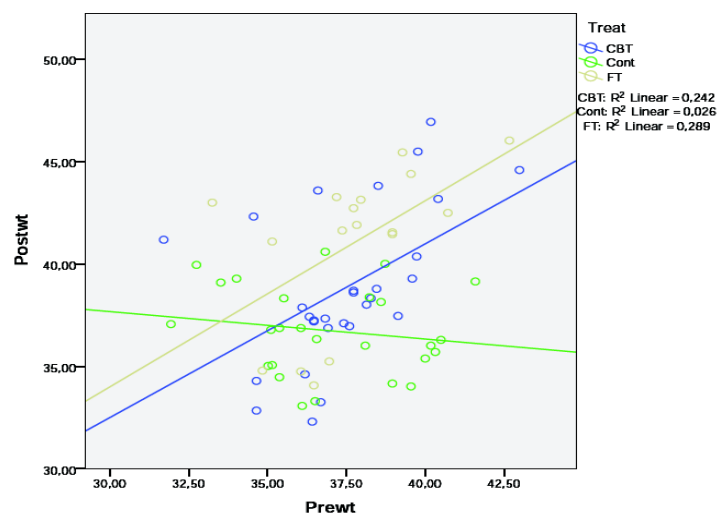
For at få alle tre grupper på samme tegning, anvendes igen **Graphs/Chart Builder**, men nu vælges det opdelte **Scatter/Dot**. Herefter sættes **Treat** i **Set Color**, **prewt** på X-aksen og **diff** på Y-aksen.

Igen skal vi bagefter dobbeltklikke på grafen og derefter klikke på **Add Fit Line at Subgroups**. Fjern igen fluebenet ved **Attach Label to Line**, før der klikkes **Apply/Close**.

Herved fremkommer figuren:



På begge disse figurer, kan vi se, at det i 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 at gå ind i `Analyse/General Linear Model/Univariate`, hvor vi sætter `diff` i `Dependent Variable`, `Treat` i `Fixed Factor(s)` og `prewt` under `Covariate(s)`. Under `Model` sættes både `Treat` og `prewt` over i `Model`, sammen med interaktionsleddet `Treat*prewt`, som her er i fokus. Desuden går man ind i `Options` og afkrydser det ønskede, hvilket altid vil være `Parameter Estimates`, og måske også `Residual Plot`.

Vi har nu opbygget en model, der består af 3 vilkårlige (ikke-parallele) linier, og får outputtet:

Tests of Between-Subjects Effects

Dependent Variable: diff

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	344,958 ^a	5	68,992	7,800	,000
Intercept	85,738	1	85,738	9,693	,003
Treat	82,120	2	41,060	4,642	,013
Prewt	74,393	1	74,393	8,410	,005
Treat * Prewt	95,648	2	47,824	5,407	,007
Error	583,789	66	8,845		
Total	1041,647	72			
Corrected Total	928,746	71			

a. R Squared = ,371 (Adjusted R Squared = ,324)

Parameter Estimates

Dependent Variable: diff

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	6,736	12,359	,545	,588	-17,939	31,411
[Treat=CBT]	,321	15,652	,021	,984	-30,928	31,570
[Treat=Cont]	34,950	15,010	2,328	,023	4,982	64,919
[Treat=FT]	0 ^a	.	.	.	.	.
Prewt	-,091	,327	-,279	,781	-,745	,562
[Treat=CBT] * Prewt	-,061	,415	-,146	,884	-,890	,769
[Treat=Cont] * Prewt	-1,042	,400	-2,606	,011	-1,841	-,244
[Treat=FT] * Prewt	0 ^a	.	.	.	.	.

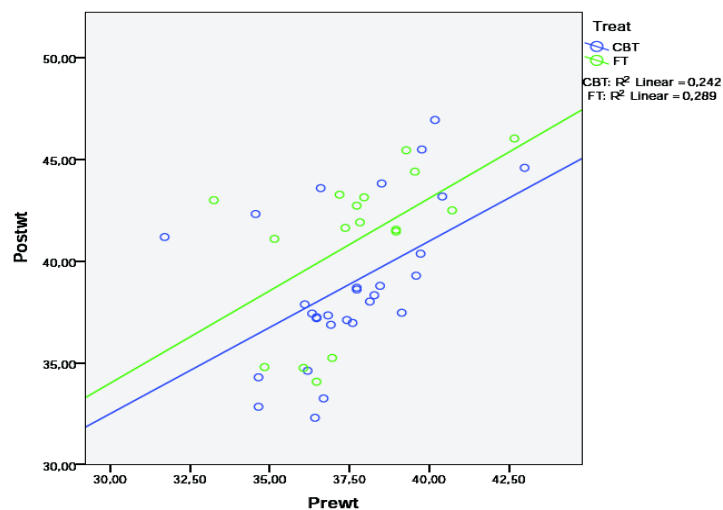
a. This parameter is set to zero because it is redundant.

hvoraf vi tydeligt ser, at der er signifikant interaktion ($P=0.007$).

- (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 med

et plot efter opskriften beskrevet ovenfor:



her ses meget parallelle linier, og for disse to grupper alene (benyt den tidligere beskrevne metode til at se på kun nogle bestemte udvalgte grupper) får vi analyseresultaterne nedenfor.

Bemærk, at vi her benytter tricket med at lægge Y-aksen hen i 37, nemlig ved at definere $\text{prewt37} = \text{prewt} - 37$ og benytte denne som kovariat i stedet for blot prewt , idet vi bliver bedt om at udtale os om resultatet for piger, der starter ud med en vægt på 37 kg.

Tests of Between-Subjects Effects

Dependent Variable: Postwt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	215,891 ^a	3	71,964	6,412	,001
Intercept	61559,562	1	61559,562	5484,906	,000
Treat	36,649	1	36,649	3,265	,078
prewt37	158,138	1	158,138	14,090	,001
Treat * prewt37	,188	1	,188	,017	,898
Error	471,385	42	11,223		
Total	72910,574	46			
Corrected Total	687,276	45			

a. R Squared = ,314 (Adjusted R Squared = ,265)

Parameter Estimates

Dependent Variable: Postwt

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
Intercept	40,355	,853	47,319	,000	38,634	42,077
[Treat=CBT]	-1,922	1,064	-1,807	,078	-4,069	,225
[Treat=FT]	0 ^a	.	.	.	.	.
prewt37	,909	,369	2,465	,018	,165	1,653
[Treat=CBT] * prewt37	-,061	,468	-,130	,898	-1,005	,884
[Treat=FT] * prewt37	0 ^a	.	.	.	.	.

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

Vi ser, at testet for interaktion giver en P-værdi på 0.90, og dermed ingen indikation af ikke-parallellitet. Det kan dog bemærkes, at hvis man ser på Cooks afstand til vurdering af indflydelsesrige observationer, så finder man to af sådanne. Vi skal dog ikke gå videre med dette her.

Da der ikke spores nogen interaktion, fortsætter vi med en model, der antager parallelle linier, altså det, der går under betegnelsen kovariansanalyse (gå tilbage til den tidligere analyse, og inde under Model fjernes leddet **Treat*prewt**).

I en sådan model uden interaktion er forskellen på de to grupper konstant (uanset hvilken startvægt, de har), og denne kan derfor

direkte aflæses af outputtet.

Tests of Between-Subjects Effects

Dependent Variable: Postwt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	215,702 ^a	2	107,851	9,834	,000
Intercept	62416,033	1	62416,033	5691,351	,000
Treat	41,024	1	41,024	3,741	,060
prewt37	165,092	1	165,092	15,054	,000
Error	471,573	43	10,967		
Total	72910,574	46			
Corrected Total	687,276	45			

a. R Squared = ,314 (Adjusted R Squared = ,282)

Parameter Estimates

Dependent Variable: Postwt

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
Intercept	40,382	,819	49,334	,000	38,731	42,033
[Treat=CBT]	-1,959	1,013	-1,934	,060	-4,002	,084
[Treat=FT]	0 ^a	.	.	.	.	.
prewt37	,871	,224	3,880	,000	,418	1,324

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

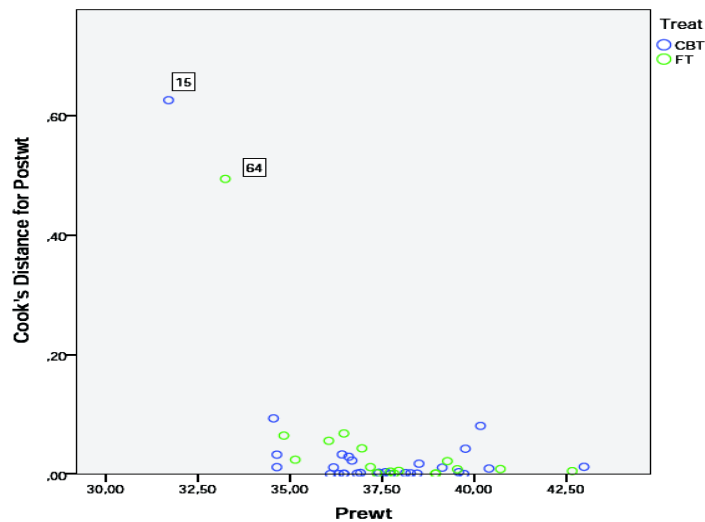
Af dette output kan vi direkte aflæse forskellen på de to behandlingstyper til 1.959 i FT-gruppens favør, med CI=(-0.084, 4.002) og P=0.060, altså ikke helt overbevisende, at der overhovedet er en forskel, men at vi heller ikke kan udelukke en forskel på op til ca. 4 kg!

Vi kan desuden aflæse, at piger, der har en startvægt på 37 kg og kommer i behandlingsgruppen FT igennemsnit ender på en slutvægt på 40.382 kg, CI=(38.731, 42.033).

Den tilsvarende værdi for CBT-piger er ikke direkte at aflæse fra output, men kan udregnes som $40.38 - 1.96 = 38.42$, men konfidensgrænserne for denne værdi lader sig ikke så let udregne, idet man ikke bare kan lægge sådanne sammen. Her må man i stedet sno sig for at få en anden referencegruppe.

Vi ser nu på et plot af Cooks afstand for indflydelsesrige observa-

tioner (benyt **Save**-knappen og afkryds **Cook's distance** under **Diagnostics**, hvorefter der i datasættet bliver lagt en variabel, kaldet **COO_1**, der kan benyttes til tegninger).



og bemærker to observationer, der har en rigtig stor værdi af Cooks afstand, nemlig de to piger, der fra starten er allermest undervægtige, der også har den største indflydelse på estimationen.

- (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 viste figur).

Vi finder outputtet:

Tests of Between-Subjects Effects

Dependent Variable: Postwt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	257,085 ^a	3	85,695	11,939	,000
Intercept	57538,768	1	57538,768	8016,314	,000
Treat	126,949	1	126,949	17,686	,000
prewt37	33,189	1	33,189	4,624	,038
Treat * prewt37	60,081	1	60,081	8,371	,006
Error	279,931	39	7,178		
Total	64018,382	43			
Corrected Total	537,016	42			

a. R Squared = ,479 (Adjusted R Squared = ,439)

Parameter Estimates

Dependent Variable: Postwt

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
Intercept	40,355	,682	59,170	,000	38,976	41,735
[Treat=Cont]	-3,621	,861	-4,206	,000	-5,363	-1,879
[Treat=FT]	0 ^a	.	.	.	.	.
prewt37	,909	,295	3,082	,004	,312	1,505
[Treat=Cont] * prewt37	-1,042	,360	-2,893	,006	-1,771	-,314
[Treat=FT] * prewt37	0 ^a	.	.	.	.	.

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

og ser her som forventet en klar indikation af interaktion, dvs. at linierne *ikke* er parallelle ($P=0.006$).

Her er der også en enkelt meget indflydelsesrig pige, denne gang en pige i FT-gruppen, der tager rigtig meget på, næsten 10 kg. Man bør nok checke, om dette kan have sin rigtighed....

Ud fra dette output kan vi direkte aflæse forskellen for piger med en startvægt på 37 kg til 3.62 i FT-gruppens favør, med et $CI=(1.88, 5.36)$.

Vi kan desuden se, at piger fra FT-gruppen i gennemsnit vil slutte på 40.36 kg, $CI=(38.98, 41.74)$, og at den tilsvarende slutvægt i Cont-gruppen tilsvarende vil være $40.36 - 3.62 = 36.74$ kg. Her har vi dog problemer med at angive konfidensintervallet, da man ikke bare kan addere sådanne. Her er man nødt til at skifte referencegruppe.

- (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 $CI=(1.31, 5.69)$.