Phd-kursus i Basal Statistik, Opgaver til 1. uge

Opgave 1: Sundby

Vi betragter et lille uddrag af det såkaldte **Sundby95**-materiale, der er en stor undersøgelse af københavnernes sundhed.

Det totale datasæt ligger på hjemmesiden som præfabrikeret datasæt med et lille udpluk af informationerne i tekstfilen `Sundby_lille`, indeholdende variablene (i den nævnte rækkefølge)

kon: Personens køn (1: mand, 2:kvinde)

v75: Personens vægt, i *kg*

v76: Personens højde, i *cm*

v17: Fysisk aftivitet i fritid (kategorier 1-4, lave tal betyder mest aktiv)

v24af: Antal drukkede genstande sidste weekend

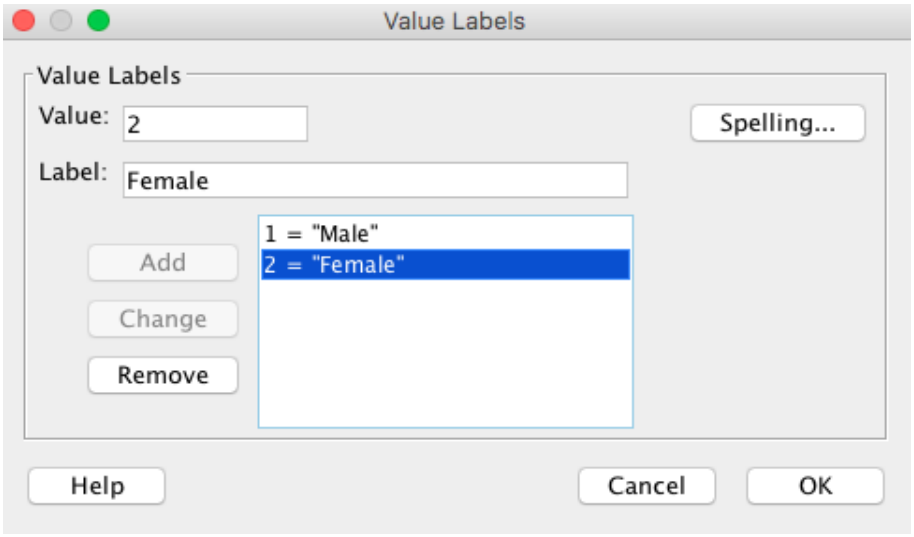
1. *Indlæs data og omdøb variablene, så de har forståelige navne.*

   Vi skal i det følgende kun bruge `kon`, `v75` og `v76`, så vi nøjes med at omdøbe de to sidstnævnte til hhv `vaegt` og `hoejde`. Samtidig definerer vi den ekstra variabel `bmi`, som kommenteres senere.

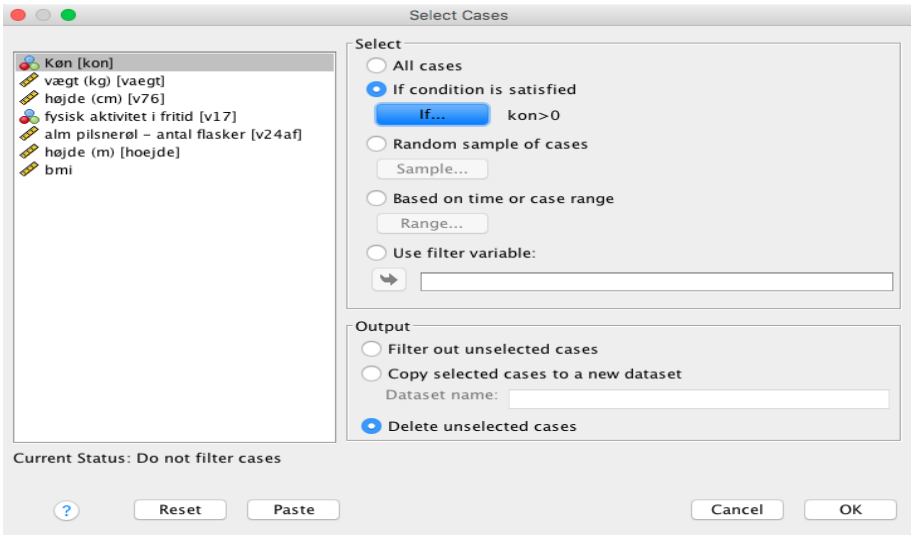
   Vi benytter `Transform/Compute variable`, hvor vi f.eks. sætter `hoejde` som `Target Variable` og skriver `v75/100` i definitionsfeltet. Herefter vil den nye variabel `hoejde` (som angiver højden i meter) være tilføjet datasættet. Tilsvarende med de to andre variabeldefinitioner:

   ```
   hoejde=v76/100
   vaegt=v75
   bmi=vaegt/hoejde**2
   ```

   Herefter kan man angive value labels til variablen `kon` ved at vælge `Variable View` i bunden af dataset-skærmbilledet og klikke på `Values`:



Nogle observationer indeholder ikke en angivelse for køn, og disse må vi slette. I menuen vælges `Data/Select Cases`:

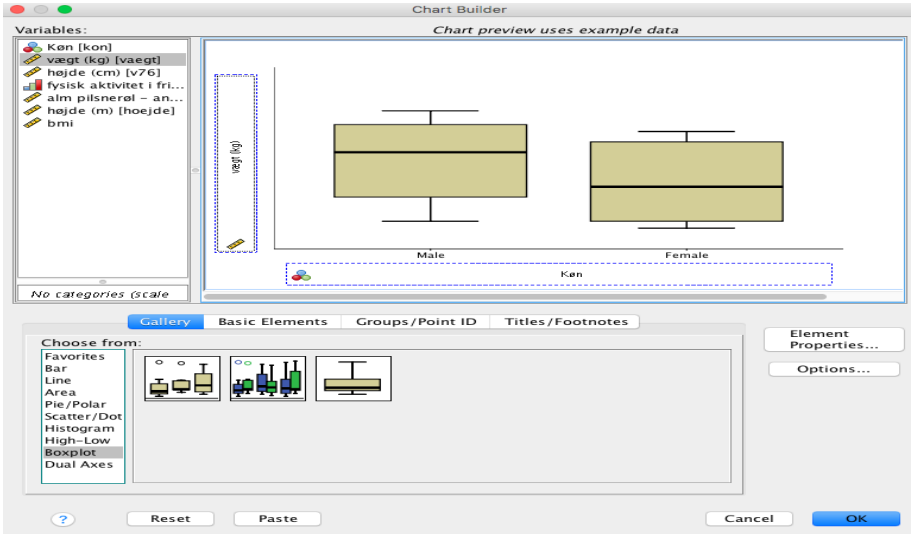


Nu har vi foretaget et par ændringer i vores datasæt `sundby`, nemlig:

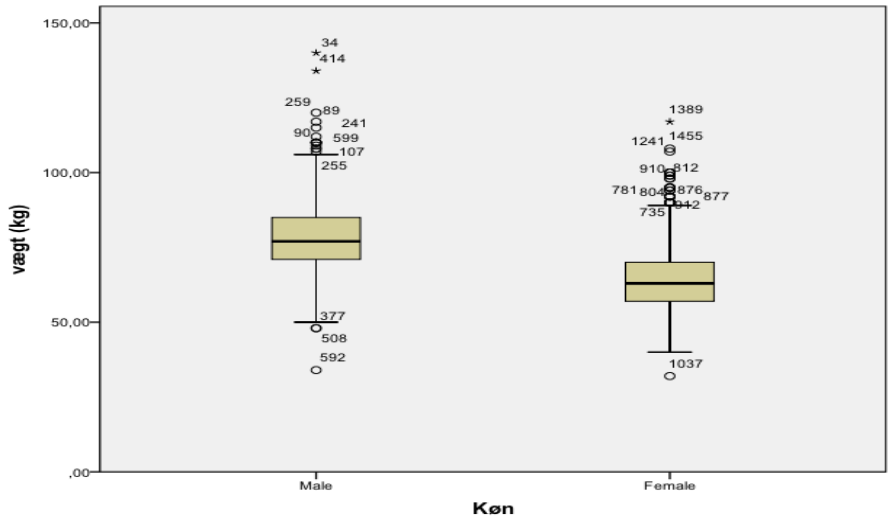
- Sletning af observationer med ukendt køn
- Definition af en mere forståelig angivelse af kønnet
- Mere intuitive betegnelser for højde- og vægt-variable, samt omkodning af højde fra *cm* til *m*
- Definition af en ny variabel, `bmi`, se nedenfor under spm. 6.

2. *Lav en illustration af vægtfordelingen (*`v75`*) for mænd hhv. kvinder (brug f.eks. Box plots eller histogrammer), og beskriv også fordelingen i tal, dvs. gennemsnit, median, spredning mv.*

   Box plottene kan laves med `Graphs/Chart Builder`, hvor man vælger `BoxPlot` (det simple yderst til venstre), sætter `vaegt` på Y-aksen og `kon` på X-aksen:



   Boxplottene kommer da til at se således ud:

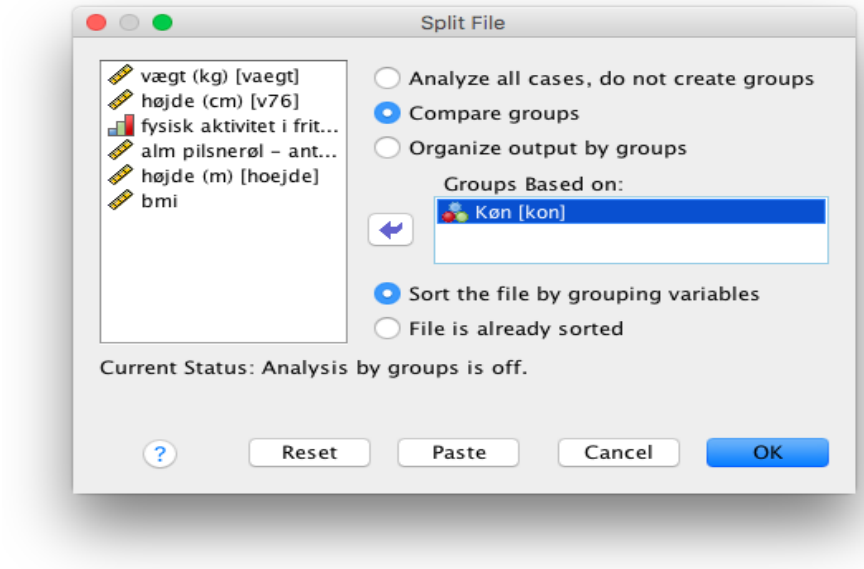


   Bemærk: Hvis Boxplottene ikke får fornuftige akser, kan man selv gå ind

og definere dem ved at klikke `Element properties`, markere `Y-Axis1` under `Edit Properties of` og sætte f.eks. `Minimum` under `Custom` til et bestemt tal.

De synes at vise en vis skævhed i fordelingerne, så lad os se nærmere på histogrammer for at vurdere tilpasningen til normalfordelingen (som skal bruges i næste spørgsmål).

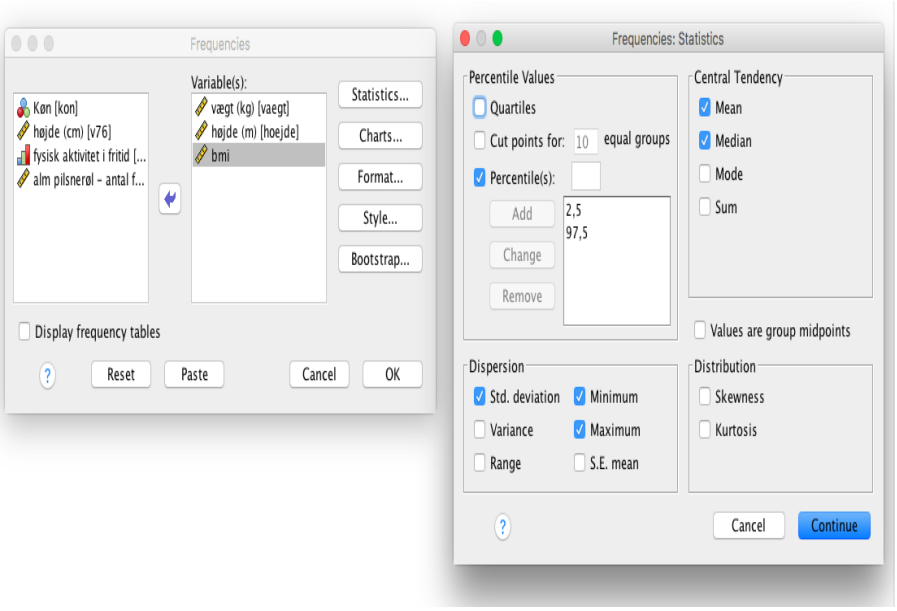
I forbindelse med histogrammerne udregner vi også 2.5% og 97.5% fraktilerne, idet vi benytter `Frequencies` under `Analyze/Descriptive Statistics`, men fordi vi vil have resultaterne opdelt på køn, skal vi først bruge `Data/Split file`, hvor vi vælger `Compare Groups` og sætter `kon` over i `Groups Based on`:

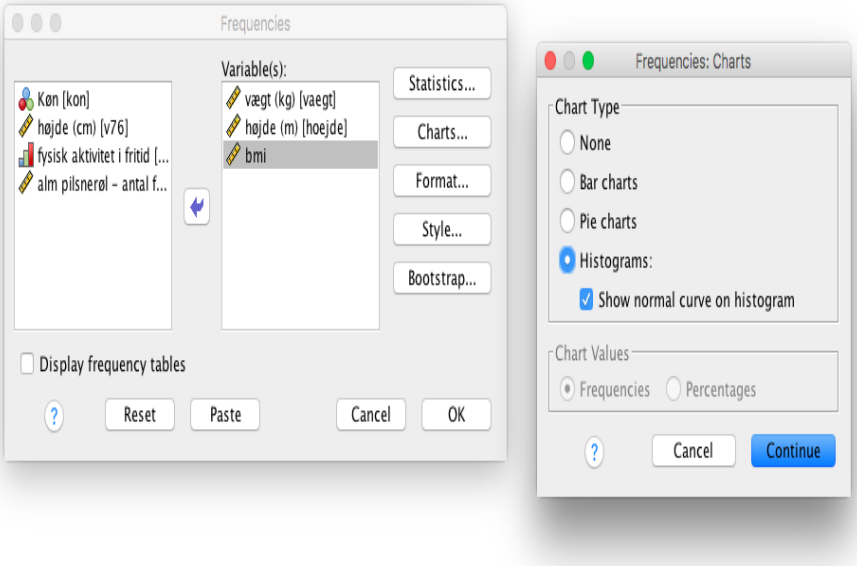


Bemærk: `Split File` *er slået til indtil man slår det fra igen! Vær desuden opmærksom på at selv om man vælger* `Compare Groups` *har dette kun effekt på opstillingen af data - Der foretages ingen statistiske sammenligninger!*

Nu kan vi køre `Frequencies`, og for nemhes skyld (for at kunne bruge det senere) gør vi det for alle de kvantitative variable på en gang. Vi sætter derfor `vaegt`, `hoejde` og `bmi` over i `Variables`, klikker `Statistics` og afkrydser det ønskede, herunder `Percentiles`, hvor man skriver `2,5/Add` og `97,5/Add/Continue`. Fjern også fluebenet ved `Display frequency tables`.

Under `Charts` `Histograms`, med flueben i `Show normal curve on histogram`.

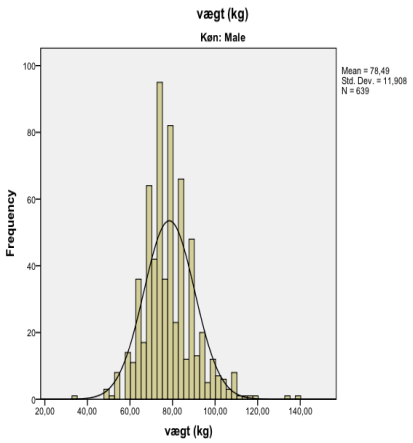


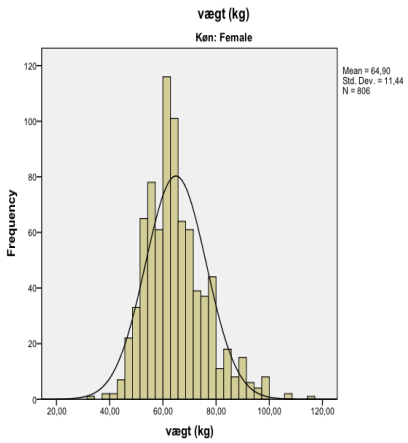


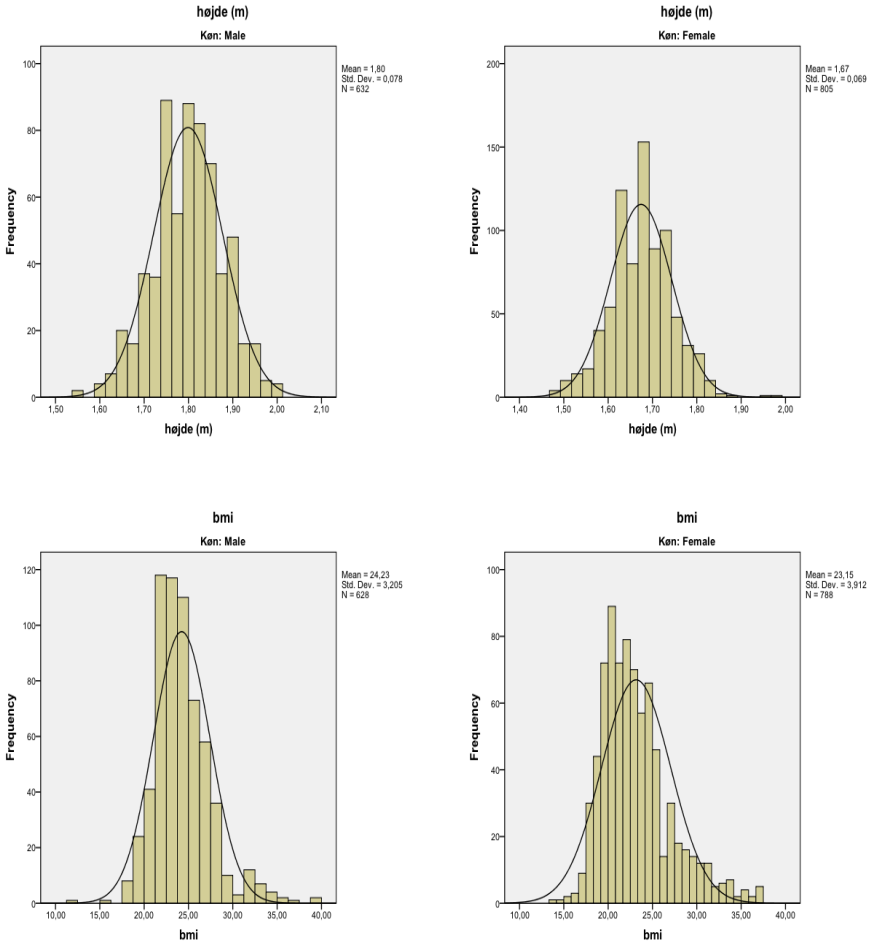
Herved får vi følgende output:

Statistics

Køn			vægt (kg)	højde (m)	bmi
Male	N	Valid	639	632	628
		Missing	8	15	19
	Mean		78,4883	1,7990	24,2284
	Median		77,0000	1,8000	23,8088
	Std. Deviation		11,90846	0,07793	3,20476
	Minimum		34,00	1,55	11,63
	Maximum		140,00	2,00	39,64
	Percentiles	2,5	58,0000	1,6400	19,0350
		97,5	106,0000	1,9500	32,7947
Female	N	Valid	806	805	788
		Missing	21	22	39
	Mean		64,8959	1,6742	23,1517
	Median		63,0000	1,6800	22,4191
	Std. Deviation		11,44037	0,06944	3,91164
	Minimum		32,00	1,48	13,67
	Maximum		117,00	1,98	37,37
	Percentiles	2,5	47,0000	1,5200	17,6124
		97,5	92,0000	1,8000	33,3502







Ikke overraskende kan vi konstatere, at mændene vejer mere end kvinderne. En del af årsagen hertil kunne jo være, at de også er højere (og det er de jo, hvilket også klart ses af såvel boxplot som summary statistics).

3. *Kommenter fundene fra forrige spørgsmålet med henblik på at konstruere normalområder for vægten, for hvert køn for sig. Bestem et sådant normalområde, både med og uden brug af normalfordelingsantagelsen. Hvordan passer de sammen?*

   Udfra gennemsnit og spredning udregner vi normalområder under antagelse af, at vi har at gøre med normalfordelinger for hvert køn. Her er udregningerne foretaget i R, som (bl.a.) fungerer som regnemaskine:

```
> 64.896+c(-2,2)*11.44
[1] 42.016 87.776
> 78.488+c(-2,2)*11.908
[1]  54.672 102.304
```

Normalområderne for vægten er således:

Kvinder: (42.0, 87.8) kg

Mænd: (54.7, 102.3) kg

Sammenligner vi med $2\frac{1}{2}\%$ og $97\frac{1}{2}\%$ fraktilerne, som blev udregnet med `Frequencies`:

Statistics

Køn			vægt (kg)	højde (m)	bmi
Male	N	Valid	639	632	628
		Missing	8	15	19
	Mean		78,4883	1,7990	24,2284
	Median		77,0000	1,8000	23,8088
	Std. Deviation		11,90846	0,07793	3,20476
	Minimum		34,00	1,55	11,63
	Maximum		140,00	2,00	39,64
	Percentiles	2,5	58,0000	1,6400	19,0350
		97,5	106,0000	1,9500	32,7947
Female	N	Valid	806	805	788
		Missing	21	22	39
	Mean		64,8959	1,6742	23,1517
	Median		63,0000	1,6800	22,4191
	Std. Deviation		11,44037	0,06944	3,91164
	Minimum		32,00	1,48	13,67
	Maximum		117,00	1,98	37,37
	Percentiles	2,5	47,0000	1,5200	17,6124
		97,5	92,0000	1,8000	33,3502

ser vi, at de normalfordelingsbaserede intervaller skyder lidt for lavt, fordi de ikke tager hensyn til den lille skævhed, der er i fordelingerne (mere om dette senere).

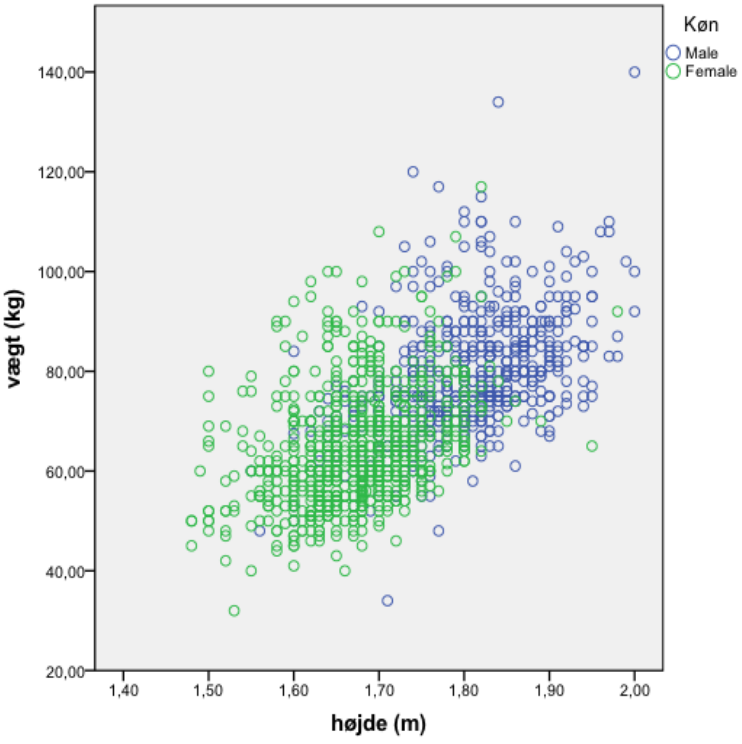
4. *Det ser klart ud til, at mændene vejer mere end kvinderne. Hvordan kunne man forklare det?*

Det har vi allerede berørt ovenfor: Det kan skyldes, at mændene er højere - men det kan selvfølgelig også skyldes kraftigere knoglebygning, større muskler osv.

5. *Lav et scatterplot af vægt overfor højde (*`v76`*), med forskellige symboler for mænd og kvinder. Ser det ud som om vægtforskellen kan forklares ved at mænd generelt er højere end kvinder?*

   Først må vi ind og aflyse `Data/Split File` ved i stedet at sætte fluebenet tilbage i `Analyze all cases, do not create groups`.

   Scatterplottet kræver farver for at man skal kunne se forskel på kønnene. Derfor laves et `Grouped Scatter` plot (`Scatter/Dot`, nr. 2 fra venstre i `Graphs/Chart Builder`, hvor kon trækkes over i group boksen øverst til højre, `hoejde` sættes på X-aksen og `vaegt` på Y-aksen. Resultatet bliver:



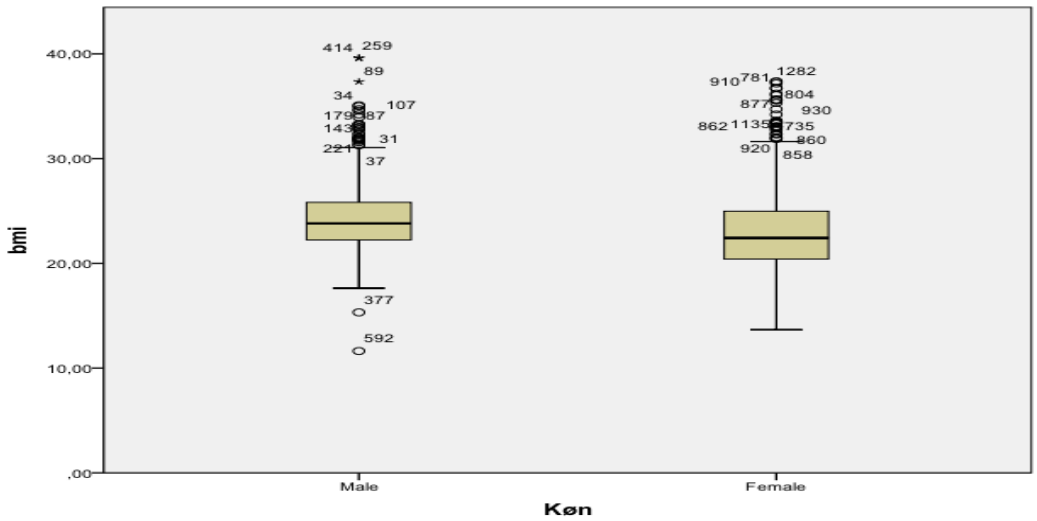
   Umiddelbart ser det på plottet ud som om højden kan forklare en god del af forskellen på vægten på mænd og kvinder.

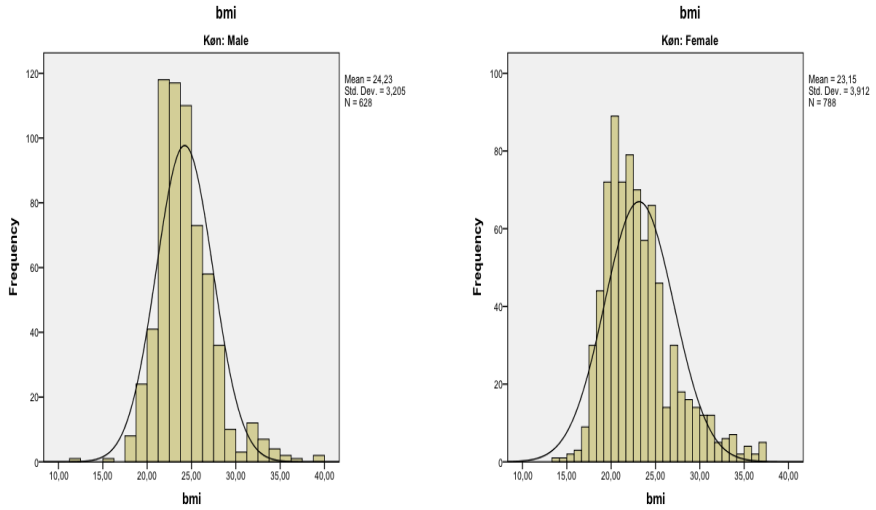
Et velkendt højde-korrigeret mål for vægt er *body mass index (BMI)*, der defineres som

$$\text{BMI} = \frac{\text{vægt i kg}}{\text{højde i meter, kvadreret}}$$

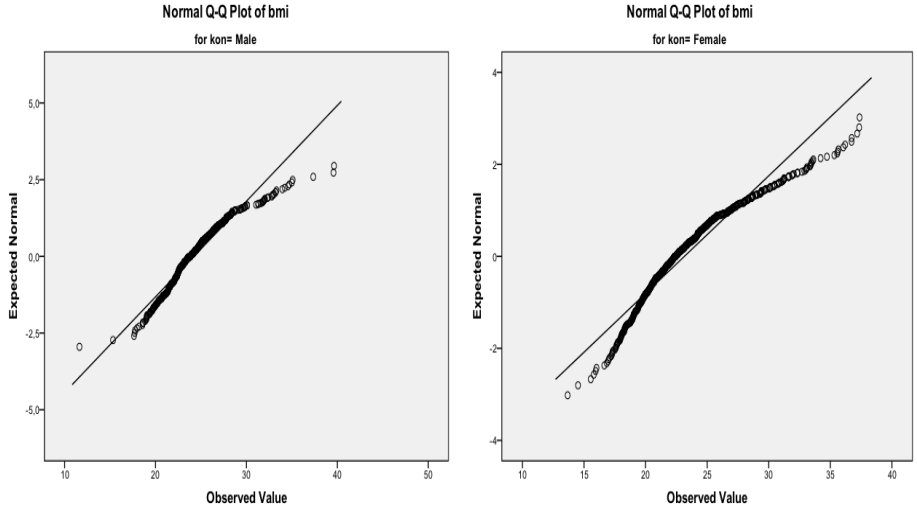
6. *Bestem nu et normalområde for* `bmi` *ved hjælp af en normalfordelingsantagelse.*

   Vi tager lige et kig på figurerne svarende til vægten ovenfor, nu blot for body mass index, `bmi`:





Her supplerer vi med fraktildiagrammer, som kan dannes på flere måder. Vi skal have `Data/Split File` tilbage på banen og kan derefter enten bruge `Analyze/Descriptive Statistics/Q-Q Plots` med `bmi` i `Variables` eller via `Analyze/Descriptive Statistics/Explore/Plots`, hvor man sætter flueben i `Normality Plots with tests` og fjerner fluebenet ved `Stem-and-leaf`. Vi får figurerne:



Igen ser vi, at normalfordelingen næppe er helt rimelig, da der ses en skævhed i fordelingen. Vi fortsætter alligevel, men sørger for at

huske, at vi ikke kan stole helt på udregninger, der baserer sig på normalfordelingsantagelsen. Det ville nok være en bedre ide at foretage udregningerne på de logaritmetransformerede værdier og så tilbagetransformere endepunkterne....

Statistics

bmi

Male	N	Valid	628
		Missing	19
	Mean		24,2284
	Median		23,8088
	Std. Deviation		3,20476
	Minimum		11,63
	Maximum		39,64
	Percentiles	2,5	19,0350
		97,5	32,7947
Female	N	Valid	788
		Missing	39
	Mean		23,1517
	Median		22,4191
	Std. Deviation		3,91164
	Minimum		13,67
	Maximum		37,37
	Percentiles	2,5	17,6124
		97,5	33,3502

Vi benytter de ovenfor udregnede størrelser fra `Frequencies` (husk `Split File...` og udregner (her gjort ved hjælp af R):

```
23.1517+c(-2,2)*3.9116 = 15.3285 30.9749
24.2284+c(-2,2)*3.2048 = 17.8188 30.6380
```

Normalområderne for body mass index er således:

Kvinder: (15.3,31.0)

Mænd: (17.8,30.6)

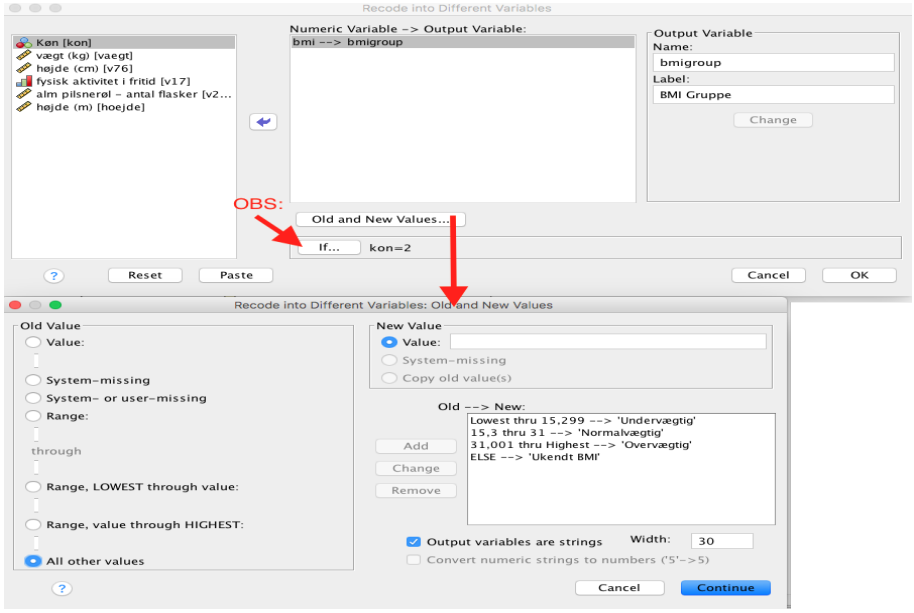
- *Hvordan passer det med den faktiske fordeling (fraktiler=percentiler)?*

  De direkte udregnede fraktiler for body mass index ses i outputtet fra `Frequencies` og vi ser, at de normalfordelingsbaserede grænser her ligger lidt lavere end de, der er baseret på fraktilerne.

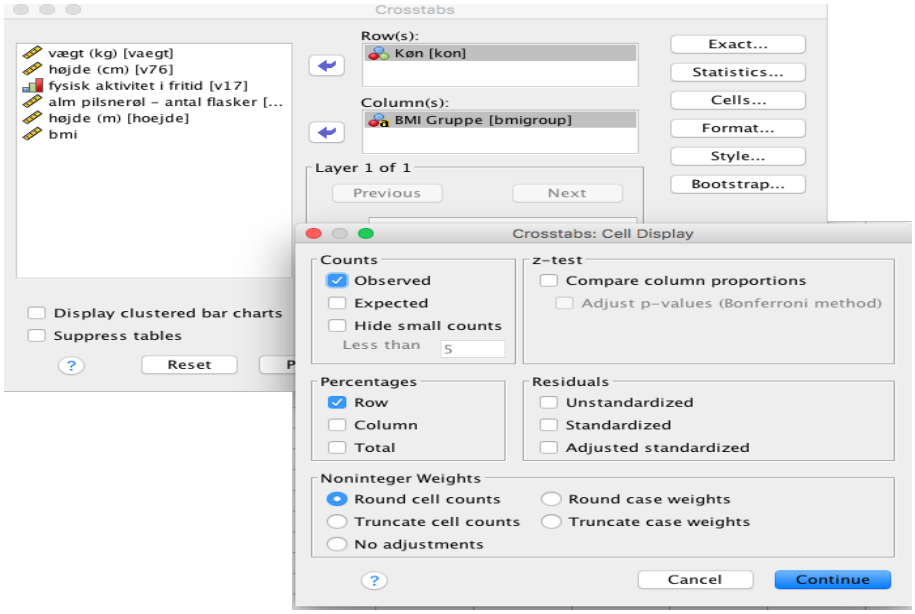
- *Hvor mange procent falder udenfor det normalfordelingsbaserede område?*

  Her skal man lige lave lidt ekstra omkodning, så det er et lidt svært spørgsmål. V i laver en ny variabel, `bmigroup` ved hjælp af `Transform/Recode Into Different Variables` (gøres for hvert køn ved at benytte `If`-knappen nederst i boxen, hvor man afkrydser `Include if case satisfies condition` og skriver enten `kon=1` eller `kon=2`): Sæt nu `bmi` over i `Numeric Variable -- Group Variable` og skriv `bmigroup` i `Output Variable/Name` (og evt. `BMI Gruppe` i `Label`). Herefter defineres værdierne for `bmigroup` således:

  - Sæt flueben i `Output Variables are strings`
  - Klik `Range/LOWEST through Value`, sæt value til nedre grænse for normalvægt for det pågældende køn og klik `Add`
  - Klik `Range/LOWEST through Value`, sæt value til nedre grænse for normalvægt for det pågældende køn og klik `Add`
  - Klik `Range/Value through HIGHEST`, sæt value til øvre grænse for normalvægt for det pågældende køn og klik `Add`



Hernæst kan vi opgøre fordelingen som procent med `Crosstabs` i `Analyze/Descriptive Statistics` (som vi kommer til at se nærmere på senere på kurset): Sæt `kon` i `Rows`, `bmigroup` i `Column(s)` og klik `Cells`, hvor man afkrydser `Counts: Observed` og i `Percentages: Row`:



Vi får så outputtet

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Køn * BMI Gruppe	1474	100,0%	0	0,0%	1474	100,0%

Køn * BMI Gruppe Crosstabulation

			BMI Gruppe			
			Normalvægtig	Overvægtig	Ukendt BMI	Undervægtig
Køn	Male	Count	594	29	19	5
		% within Køn	91,8%	4,5%	2,9%	0,8%
	Female	Count	744	42	39	2
		% within Køn	90,0%	5,1%	4,7%	0,2%
Total		Count	1338	71	58	7
		% within Køn	90,8%	4,8%	3,9%	0,5%

Køn * BMI Gruppe Crosstabulation

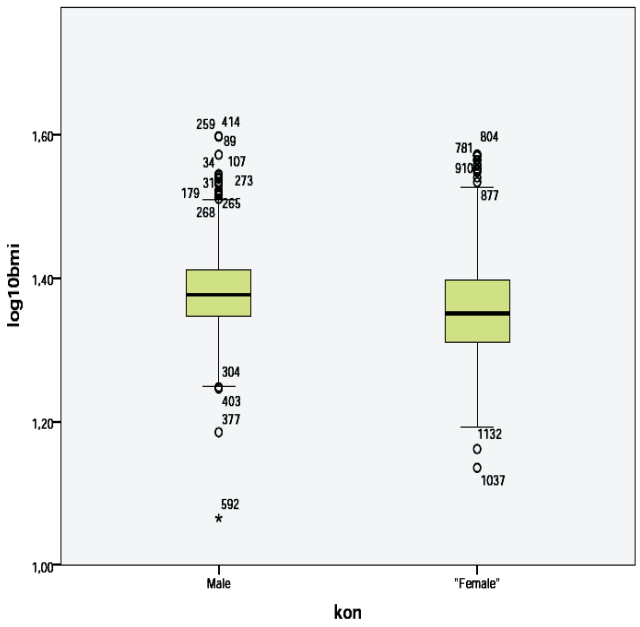
			Total
Køn	Male	Count	647
		% within Køn	100,0%
	Female	Count	827
		% within Køn	100,0%
Total		Count	1474
		% within Køn	100,0%

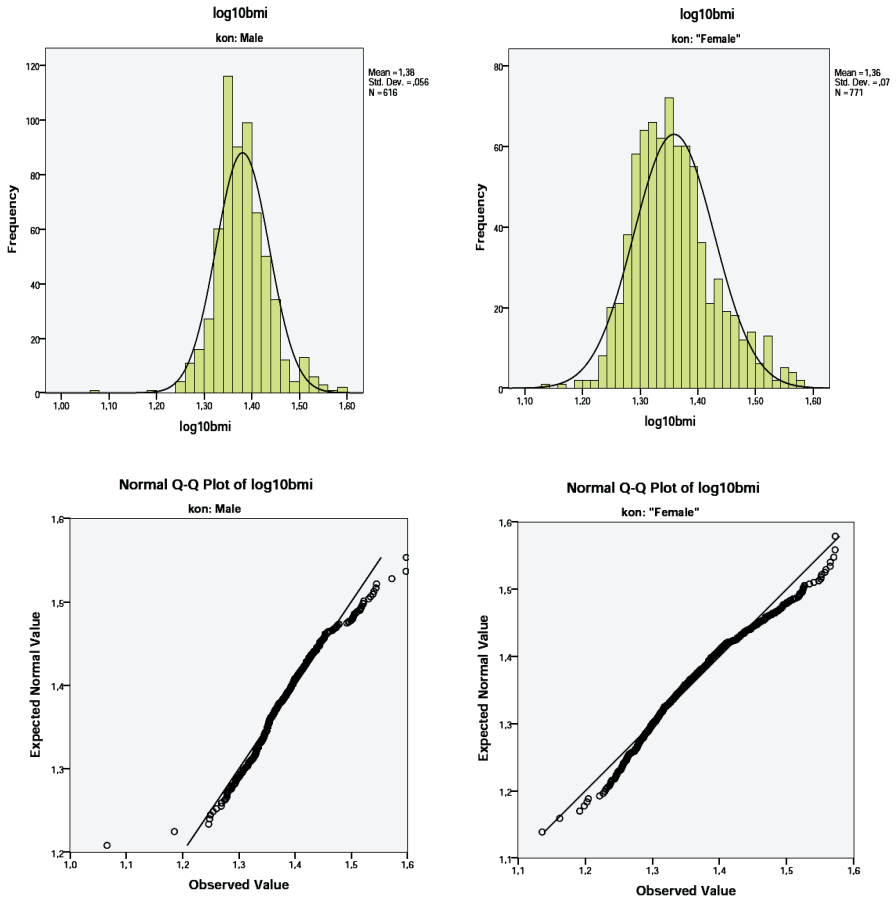
Vi finder altså ret få (under 1%), der ligger under normalområdet (og altså ville blive vurderet til at være for tynde), men lidt for mange, der ligger over (4-5%). Det afspejler (igen) den skæve fordeling af `bmi`.

- *Hvordan ser normalområderne ud, hvis de er baseret på en normalfordelingsantagelse for logaritmen til* `bmi`*?*

  Vi bemærkede ovenfor en vis skævhed i (bl.a) fordelingen af `bmi`, og en deraf følgende diskrepans mellem normalområder udregnet dels fra gennemsnit og spredning og dels fra de empiriske fraktiler. Denne skævhed forsøger vi nu at transformere os ud af, idet vi definerer `log10bmi=Lg10(bmi)` under `Transform/Compute`.

  Vi kan nu gentage figurerne fra tidligere, blot med denne transformerede variabel, og finder:





Disse figurer synes at vise en noget bedre normalfordelingstilpasning end den utransformerede variabel, og vi fortsætter derfor med at udregne summary statistics for denne:

Statistics

log10bmi

Male	N	Valid	616
		Missing	31
	Mean		1,3807
	Median		1,3766
	Std. Deviation		,05590
	Minimum		1,07
	Maximum		1,60
	Percentiles	2,5	1,2795
		97,5	1,5159
"Female"	N	Valid	771
		Missing	56
	Mean		1,3588
	Median		1,3509
	Std. Deviation		,06978
	Minimum		1,14
	Maximum		1,57
	Percentiles	2,5	1,2453
		97,5	1,5234

Igen regner vi videre i R, idet vi i et hug udregner normalområder for den logaritmerede variabel, samt tilbagetransformerer til den oprindelige skala. Dette kan selvfølgelig også gøres i flere trin og med en lommeregner:

```
> 10^(1.3588+c(-2,0,2)*0.0698)
[1] 16.56533 22.84546 31.50649

> 10^(1.3807+c(-2,0,2)*0.0557)
[1] 18.59088 24.02702 31.05275
```

Normalområderne for body mass index er således:

Kvinder: (16.6,31.5)

Mænd: (18.6,31.1)

hvilket er en smule bedre i overensstemmelse med områderne, som udregnes fra de empiriske fraktiler.

Man kan evt. gå videre med denne opgave og se på forskellige forklarende variable for BMI.

Måske er BMI højt for

- fysisk inaktive personer (`v17`)
- personer, der drikker meget (`v24af`)

men så skal man bruge analyseteknikker, vi endnu ikke har set på.

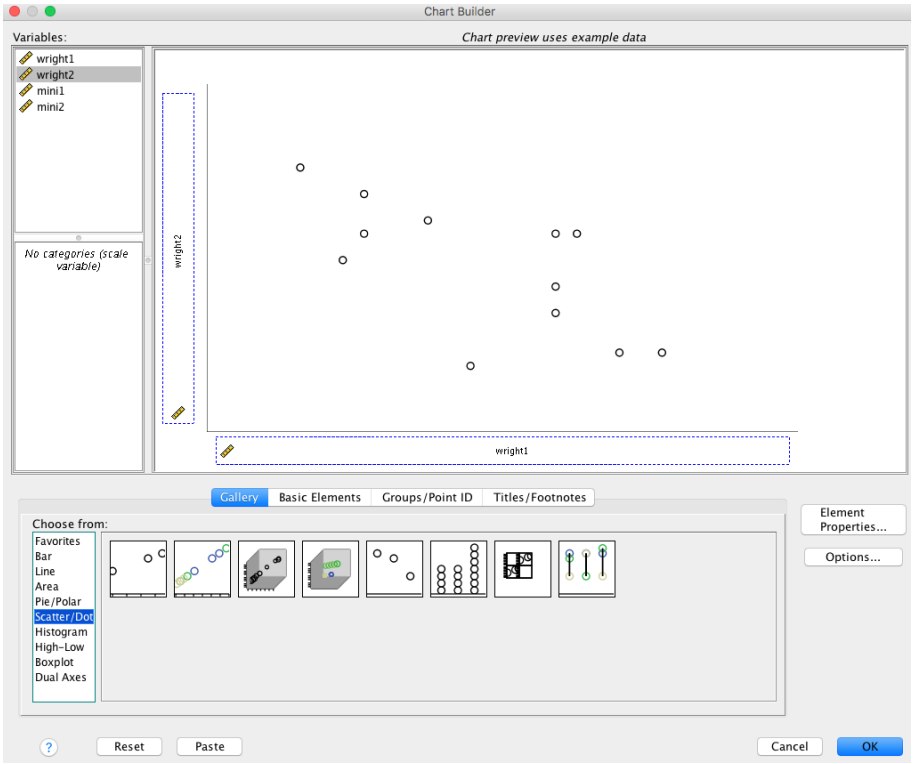
Opgave 2: Sammenligning af målemetoder

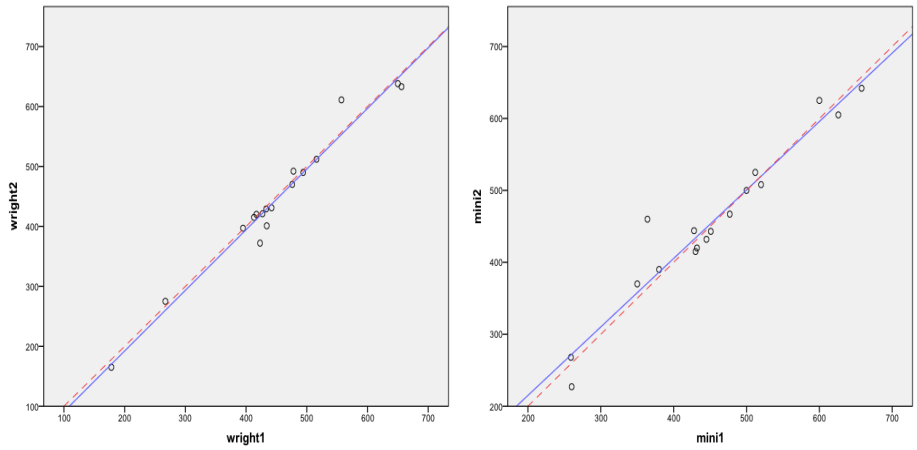
For 17 patienter er der målt **peak expiratory flow rate** (maksimal udåndingshastighed, i l/min) på to forskellige måder, dels ved at anvende det **traditionelle** *Wright peak flow meter*, og dels med det **nye** såkaldte *mini Wright flow meter* (Bland and Altman, 1986).

Med begge apparater er der foretaget dobbeltbestemmelser, således at der i alt foreligger 4 observationer for hver person.

1. **Indlæs data i SPSS (menuen `File/Open/Data...`).**

   Til en start kan vi se på et plot af dobbeltbestemmelser mod hinanden, for hver af de to målemetoder med et scatterplot (`Graphs/Chart builder...`) hvor de to målinger trækkes ind på hhv. x- og y-aksen:





Referencelinjer kan tilføjes i graf-editoren som åbnes ved at dobbeltklikke på grafen. Det ses, at observationerne fordeler sig rimeligt omkring en linie (den blå regressionslinie, fuldt optrukket), og hvis man kigger nærmere efter, er denne linie næsten identitetslinien (den røde skraverede), og dermed vil vi umiddelbart sige, at dobbeltbestemmelserne stemmer rimeligt godt overens.

De efterfølgende spørgsmål skal lede igennem forskellige betragtninger vedrørende vurdering af hver af målemetoderne samt sammenligning af de to målemetoder. Det endelige formål er at kvantificere overensstemmelsen mellem de to målemetoder (hhv. Wright og Mini Wright).

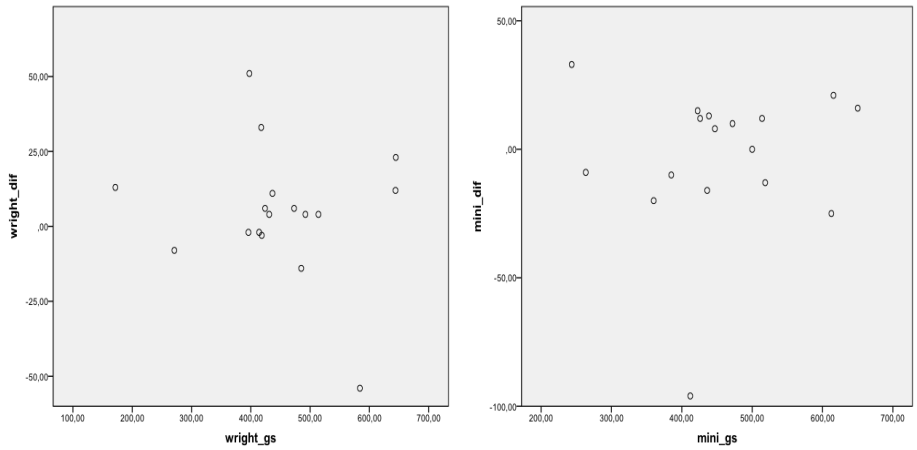
2. *Vurder (for hver af de to målemetoder for sig) om differensen mellem dobbeltmålinger afhænger af niveauet af lungefunktionen. En god metode til dette er det såkaldte Bland-Altman plot (scatter plot af differenser mod gennemsnit).*

   Vi har nu brug for at ændre i datasættet `wright`, fordi vi skal danne nogle nye variable. Nedenfor udregner vi differenser mellem dobbeltbestemmelser for hver af metoderne, gennemsnit af selvsamme dobbeltbestemmelser, samt (til brug for spm. 6) differenser mellem gennemsnit af de to metoder (`dif`) samt gennemsnittet af disse gennemsnit (`gnsnit`, som altså blot er gennemsnittet af alle fire målinger).

   De nye variable laves med `Transform/Compute variable` (en gang for hver ny variabel) eller ved at skrive følgende kode i en `SYNTAX`-fil og køre den:

```
COMPUTE wright_dif=wright1-wright2.
COMPUTE wright_gs=(wright1+wright2)/2.
COMPUTE mini_dif=mini1-mini2.
COMPUTE mini_gs=(mini1+mini2)/2.
COMPUTE dif=wright_gs-mini_gs.
COMPUTE gnsnit=(wright_gs+mini_gs)/2.
EXECUTE.
```

Vi laver herefter (for hver af målemetoderne for sig) et scatter-plot af differenserne mod gennemsnittet, hvorved vi får figurerne



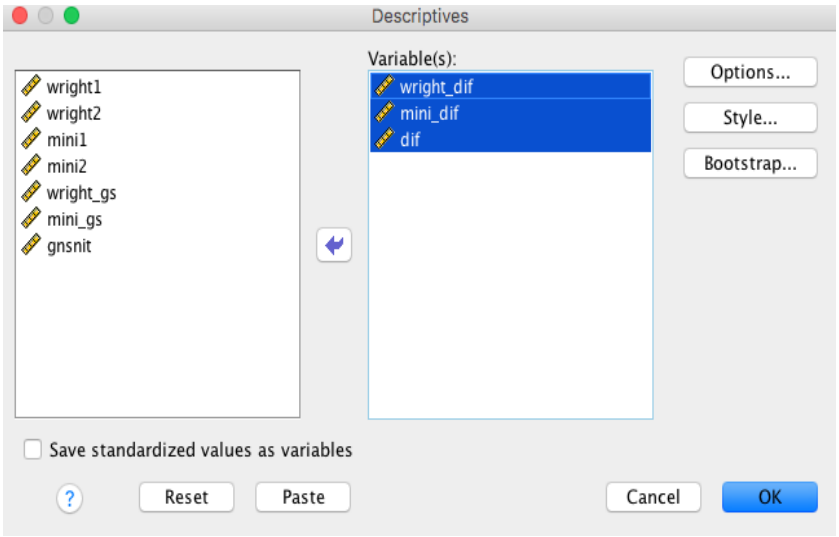
Disse figurer går under betegnelsen 'Bland-Altman plots', efter artiklen Bland&Altman(1986). Vi ser af disse plots, at differenserne generelt ligger i et bånd omkring 0 af nogenlunde lige stor bredde hele vejen, omend det lille antal observationer ikke tillader alt for kategoriske konklusioner.

Der synes at være en enkelt person, hvor der er meget stor forskel mellem de to observationer af `mini`. Vi vil komme tilbage til dette senere.

3. **Reproducerbarhed:** *Udregn og fortolk limits of agreement (normalområde for differenser), igen separat for hver af metoderne, uden at transformere. Vurder rimeligheden af de nødvendige antagelser.*

   Limits of agreement er normalområder for differenserne, så vi skal finde gennemsnit og spredning for disse, ved hjælp af `Descriptives` i menuen

`Analyze/Descriptive Statistics`. Vi gør dette for alle tre differenser på en gang ved at tilføje dem på en gang:



hvorved vi får

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
wright_dif	17	-54,00	51,00	4,9412	21,72404
mini_dif	17	-96,00	33,00	-2,8824	28,87231
dif	17	-92,00	51,50	-6,0294	33,20414
Valid N (listwise)	17				

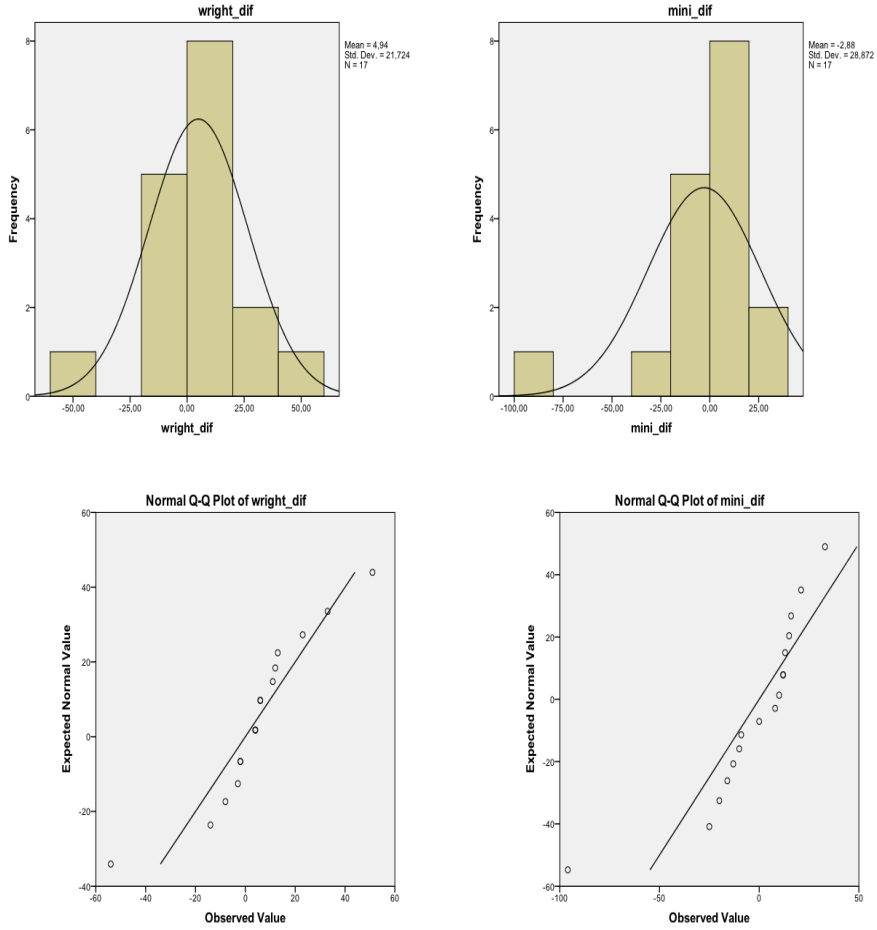
Vi går ud fra, at de 17 personer ikke er familiemæssigt relateret, og at de 17 differenser derfor er uafhængige. For at anvende ovenstående spredninger til at udregne normalområder, skal vi yderligere sikre os, at differenserne er rimeligt normalfordelte og nogenlunde af samme størrelsesorden uanset niveau. Det sidste var netop hvad vi vurderede i spørgsmålet ovenfor, så tilbage står antagelsen om normalitet. Nedenfor ses histogrammer for hhv. `wright_dif` og `mini_dif` og vi ser, at der er nogen afvigelse fra en normalfordeling. Usikkerheden i vurderingen er imidlertid stor med så få observationer.

Endvidere er vist et fraktildiagram, selv om vi må erkende, at vi ikke

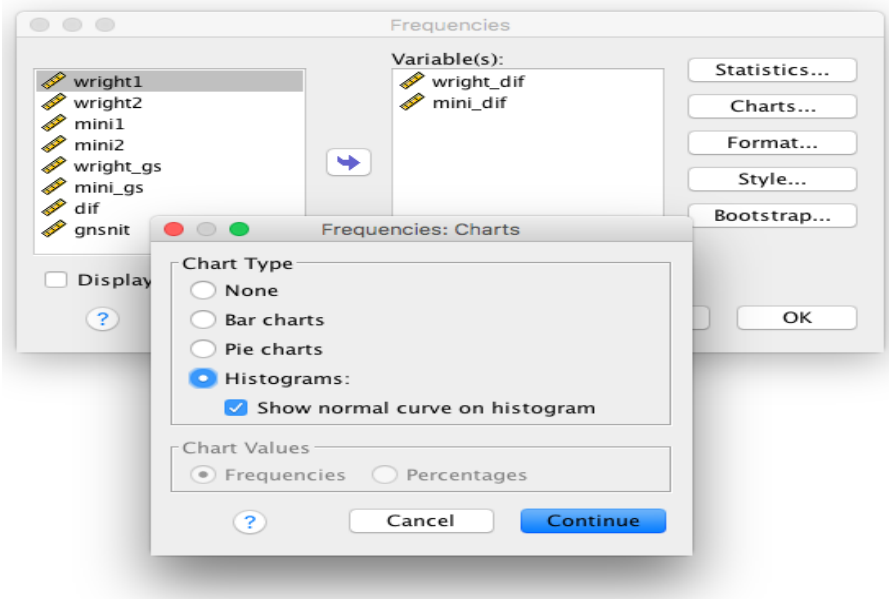
med nogen rimelighed kan vurdere normalfordelingstilpasningen med så få observationer.

Derfor bliver limits-of-agreement også usikre, hvilket vi vil komme tilbage til nedenfor i en teknisk note.

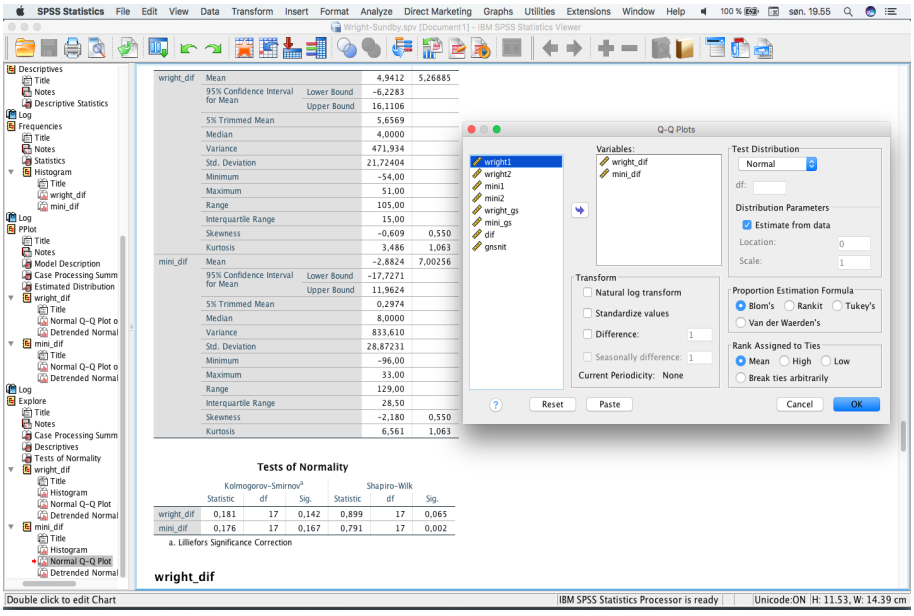
Figurerne kan dannes på flere måder, som alle findes i menuen `Analyze/Descriptive Statistics`. Undermenuen `Frequencies` giver histogrammer med en overlagt normalfordelingskurve, `Q-Q Plots...` laver udelukkende Q-Q plots mens `Explore...` kan begge dele, blot uden en standardkurve for normalfordeling på histogrammet.



Menuvalgene og syntaxen til de tre forskellige måder at lave plots på ser således ud:

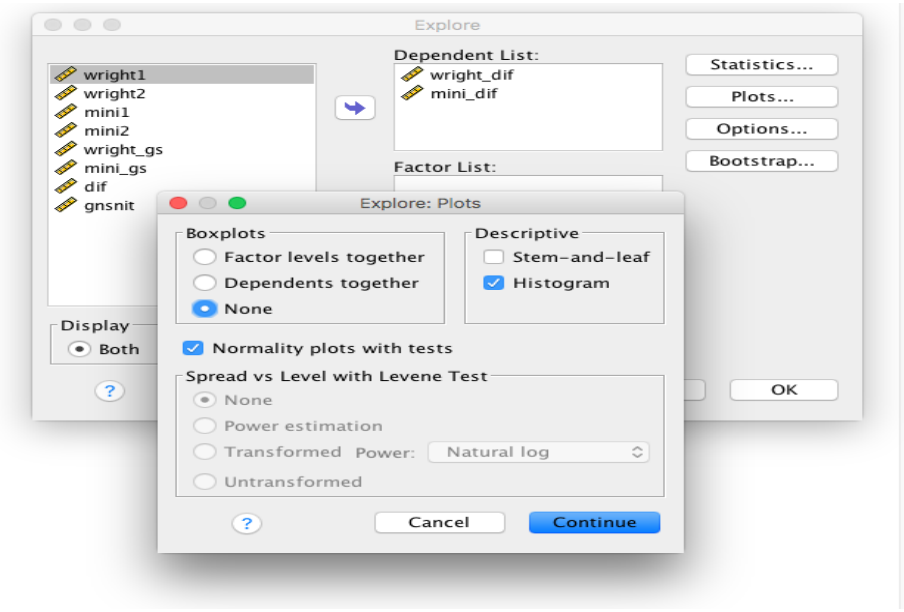



```
FREQUENCIES VARIABLES=wright_dif mini_dif
  /FORMAT=NOTABLE
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.
```




```
PPLOT
  /VARIABLES=wright_dif mini_dif
```

```
  /NOLOG
  /NOSTANDARDIZE
  /TYPE=Q-Q
  /FRACTION=BLOM
  /TIES=MEAN
  /DIST=NORMAL.
```




```
EXAMINE VARIABLES=wright_dif mini_dif
  /PLOT HISTOGRAM NPPLOT
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Med outputtet fra `Explore` får man også tests for normalitet (som man normalt ikke får meget relevant information fra, fordi de plejer at blive godkendt, når man har få observationer og forkastet, når man har mange...) giver faktisk her en afvisning for Mini Wright, på trods af det sparsomme materiale. Det skyldes nok hovedsagelig den (numerisk) meget store differens på -96.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
wright_dif	0,181	17	0,142	0,899	17	0,065
mini_dif	0,176	17	0,167	0,791	17	0,002

a. Lilliefors Significance Correction

Denne afvisning - ligesom i øvrigt det lave antal observationer - gør, at vi skal tage de nedenfor udregnede grænser med et stort forbehold. Vi finder limits of agreement til

Wright: $4.94 \pm 2 \times 21.72 = (-38.50, 48.38)$
Mini Wright: $-2.88 \pm 2 \times 28.87 = (-60.62, 54.86)$

Betydningen af limits of agreement er, at differenserne mellem dobbeltbestemmelser med 95% sandsynlighed vil ligge indenfor disse grænser, dvs. de udtrykker troværdigheden af en enkelt måling med hver af apparaterne.

Da datamaterialet er så lille, kunne vi også have valgt at bruge en passende t-fraktil til at udregne disse normalområder, det ville i så fald være med 16 frihedsgrader, altså 2.12.

Teknisk note 1:
Man kunne ligeledes overveje, om man skulle kræve, at differenserne havde middelværdi 0 og dermed estimere spredningen ved $\frac{1}{17}\sum_{p=1}^{17} \text{dif}_p^2$ i stedet for $\frac{1}{16}\sum_{p=1}^{17}(\text{dif}_p - \bar{\text{dif}})^2$

Herved ville vi få symmetriske normalområder (limits of agreement):

Wright: $0 \pm 2 \times 21.65 = \pm 43.30$
Mini Wright: $0 \pm 2 \times 28.16 = \pm 56.32$

Teknisk note 2: Usikkerhed på grænserne:
Paa grund af det lille materiale, vil selve grænserne være ret usikkert bestemt, nemlig med en standard error på ca. $\sqrt{\frac{3}{n}} \times s$, hvilket her giver

- Wright:

- nedre grænse: $-48.38 \pm 2 \times \sqrt{\frac{3}{17}} \times 21.72 = (-66.6, -30.1)$
- øvre grænse: $38.50 \pm 2 \times \sqrt{\frac{3}{17}} \times 21.72 = (20.3, 56.7)$

- Mini Wright:
  - nedre grænse: $-54.86 \pm 2 \times \sqrt{\frac{3}{17}} \times 28.87 = (-79.1, -30.6)$
  - øvre graense: $60.62 \pm 2 \times \sqrt{\frac{3}{17}} \times 28.87 = (36.4, 84.9)$

Det ses af ovenstående, at grænserne er frygteligt dårligt bestemt ud fra så få observationer, så når man vil udtale sig om limits-of-agreement (eller normalområder generelt) bør man have langt flere observationer.

4. *Hvilken af metoderne har den bedste reproducerbarhed?*

   Baseret på de udregnede limits of agreement ovenfor, ser det ud som om Wright metoden har en noget bedre reproducerbarhed end Mini Wright, idet dens limits of agreement er smallest.

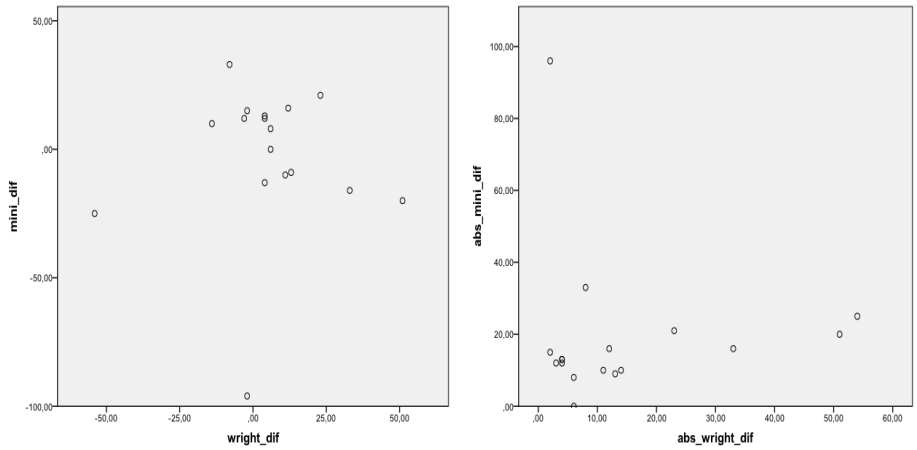
   Man kunne godt lave et egentligt test for dette, men det fører alt for vidt her...specielt naar vi lige har indset, *hvor* usikre, disse graenser er.

5. *Tegn et scatter plot af de numeriske differenser mellem dobbeltbestemmelser, med de to metoder paa hver sin akse, og vurder på baggrund af dette, om der er nogle personer, der ser ud til at være mere ustabile at måle på end andre.*

   Den venstre af figurerne nedenfor viser de to sæt differenser (med fortegn) plottet mod hinanden, medens den højre figur plotter de tilsvarende numeriske (absolutte) differenser, dannet ved f.eks.

```
COMPUTE abs_wright_dif=abs(wright_dif).
COMPUTE abs_mini_dif=abs(mini_dif).
EXECUTE.
```

   Hvis fortegnet på differensen skønnes at være vigtigt (hvis der f.eks. ses en generel stigning fra første til anden måling) bør venstre figur benyttes, ellers er højre lettere at se på.

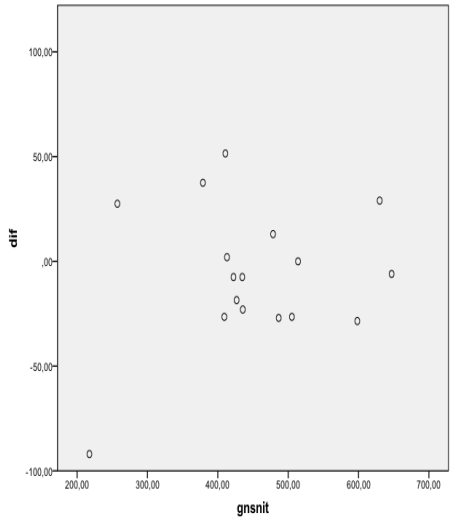


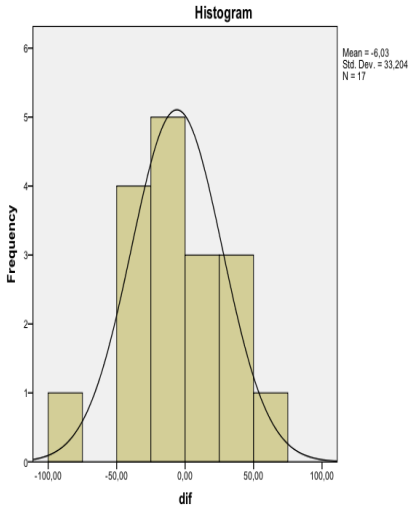
Vi skal vurdere om der er enkelte personer, der har store differenser mellem dobbeltbestemmelserne for *begge* målemetoder, og dette ses ikke umiddelbart at være tilfældet. Vi har (som tidligere bemærket) en enkelt person med en stor diskrepans mellem de to målinger for Mini Wright, men denne person har pænt overensstemmende målinger for Wright apparaturet, så man kan ikke sige, at vedkommende er svær at måle på. Sådanne personer, der er 'svære at måle på' ses i andre sammenhænge, såsom vurdering af leverstørrelse, hvor overvægtige personer er sværere at vurdere.

6. **Overensstemmelse:** *Sammenlign nu gennemsnittene af dobbeltbestemmelserne for de to metoder, dvs. tegn igen Bland-Altman plot og udregn limits of agreement, denne gang for sammenligning af de to målemetoder. Kommenter den kliniske anvendelighed af disse grænser.*

   Vi arbejder nu videre med de to gennemsnit, ovenfor kaldet `wright_gs` hhv. `mini_gs`. Igen skal vi se på et plot af differenser (`dif`) mod gennemsnit (`gnsnit`) samt vurdere rimeligheden af normalfordelingsantagelsen, inden vi går over til at udregne normalområder for differenserne.

   De relevante tegninger er





og ved hjælp af `Descriptives` finder vi de størrelser, vi skal bruge til at udregne normalområder

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
dif	17	-92,00	51,50	-6,0294	33,20414
Valid N (listwise)	17				

Selv om det (igen) er lidt vovet på så få observationer, udregner vi nu limits of agreement til

Wright vs. Mini Wright: $-6.03 \pm 2 \times 33.20 = (-72.43, 60.37)$

Når vi anvender disse grænser i praksis, skal vi huske på, at de er udregnet på baggrund af gennemsnit af to dobbeltbestemmelser. Hvis dette ikke er sædvanlig klinisk praksis, dvs. hvis man i praksis kun foretager en enkelt måling, så vil disse grænser være *for snævre*!

7. *Er der systematisk forskel på de to målemetoder? Kvantificer!*

Vi interesserer os her for middelværdierne af de to målemetoder, nærmere betegnet om disse afviger signifikant fra hinanden. Igen er der tale om parrede observationer (gennemsnittene `wright_gs` hhv `mini_gs`), så vi ser enten på differenserne `dif` og tester om disse har middelværdi 0 eller foretager et parret t-test, som findes i menuen
`Analyze/Compare Means/Paired Samples T Test`.
Forudsætningen for dette er rimelig normalitet for differenserne, hvilket vi allerede checkede ovenfor.

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	wright_gs	447,8824	17	117,47321	28,49144
	mini_gs	453,9118	17	111,29118	26,99208

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	wright_gs & mini_gs	17	0,959	0,000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	wright_gs - mini_gs	-6,02941	33,20414	8,05319	-23,10140	11,04258	-0,749	16	0,465

Vi ser altså, at T-testet giver teststørrelsen t=-0.75, svarende til P=0.47, og altså ingen signifikant forskel på de to målemetoder. En tilsvarende konklusion opnås fra et nonparametrisk test.

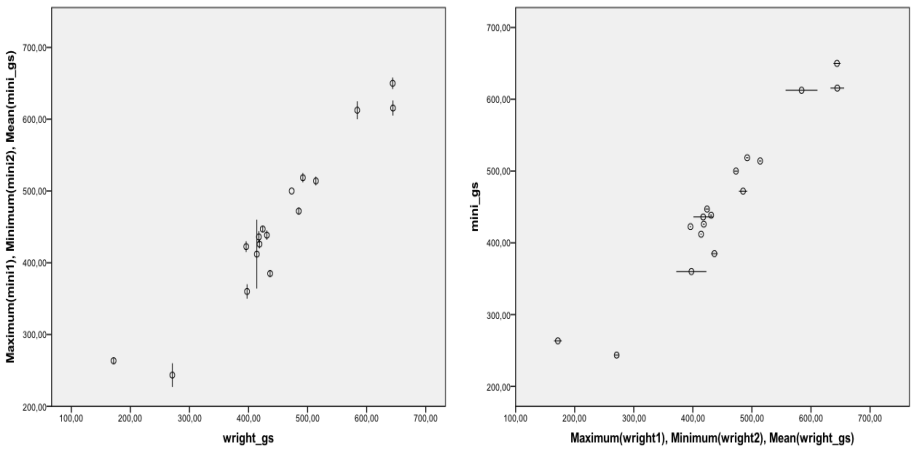
Hermed kan vi imidlertid ikke være sikre på, at der ingen forskel er, så vi kvantificerer den sandsynlige forskel ved et konfidensinterval for forskellen mellem middelværdier, som her aflæses fra outputtet til at være $(-23.10, 11.04)$

Vi kan altså ikke udelukke at forskellen på middelværdierne kan være op til ca. 10 'den ene vej' eller lidt over 20 'den anden vej'.

8. *Hvis en forskel på 75 l/min skønnes at have klinisk betydning, kan vi så erstatte Wright med det nye mini Wright?*

   Her skal vi vurdere om der hyppigt forekommer forskelle på 75 l/min, når man måler to gange på samme person med hvert af de to forskellige apparater. Ud fra limits of agreement ser vi, at 75 l/min ligger udenfor det, der 'normalt' forekommer, dvs. det, der forekommer i 95% af tilfældene. Det vil således være relativt sjældent, at vi blot ved et tilfælde ser klinisk betydelige afvigelser mellem de to målemetoder, igen **forudsat at vi til daglig virkelig benytter gennemsnit af dobbeltbestemmelser!**

Sluttelig skal vi se to figurer (High-Low plots), der forsøger at medtage alle observationer på en gang:



For hver person råder vi over 4 observationer, 2 med hver målemetode. Disse 4 er opsat som linjer, idet dobbeltbestemmelser foretaget med samme målemetode er forbundet med et liniestykke.

Figurerne er illustrative, fordi de både viser overensstemmelsen mellem de to typer af måleappartur (ligger krydsene cirka på identitetslinien?) samt reproducerbarheden for hver af metoderne (hhv. længden af stregerne). Det, vi så i spørgsmål 3, var, at det traditionelle apparatur (`wright`) var en anelse bedre

end det nye (`mini`), hvilket her svarer til, stregerne til højre er en anelse kortere end dem til venstre. Vi kan af tegningen se, at dette hovedsagelig skyldes en enkelt person, hvor der var stor uoverensstemmelse mellem de to `mini`-målinger.

Reference:

Bland, J.M. and Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* **i**, 307-310.