Opgavebesvarelse, Basalkursus, uge 3

Opgave 1: Udskrivning af astma patienter (DGA s. 273)

I en randomiseret undersøgelse foretaget af Storr et. al. (Lancet, i, 1987) sammenlignes effekten af en enkelt dosis prednisolone med placebo for børn med akut astma.

Der var 73 børn i placebo gruppen og 67 i prednisolone gruppen.

Resultatafsnittets første sætning lyder: "2 patients in the placebo group (3%, 95% confidence interval -1 to 6%) and 20 in the prednisolone group (30%, 19 to 41%) were discharged at first examination ($P < 0.0001$)"

Metodeafsnittet forklarer, at ovenstående P-værdi er udregnet ved hjælp af Fishers eksakte test.

1. *Skriv 2 × 2-tabellen op.*

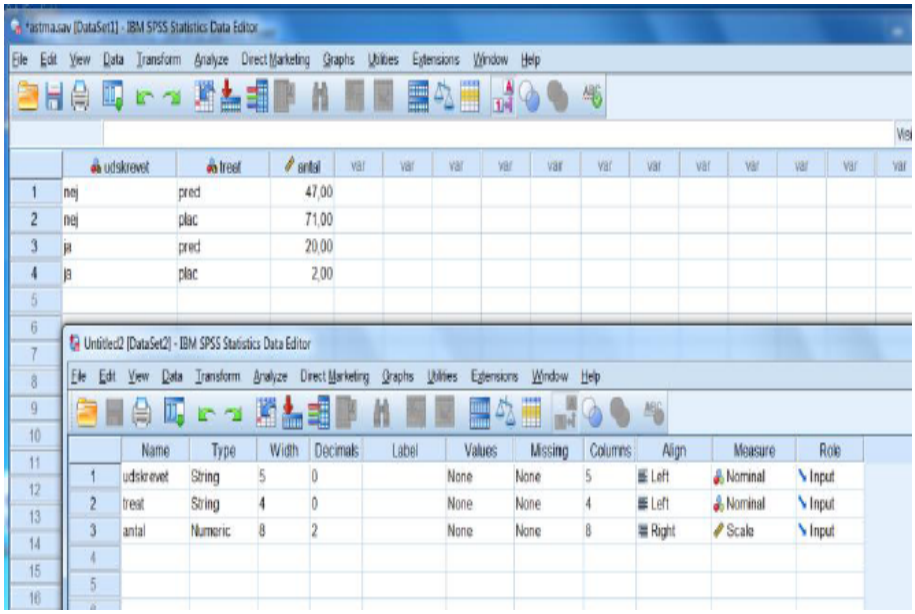
   Der var 2 ud af 73 placebo-randomiserede der blev udskrevet, dvs. 71 der ikke blev, og tilsvarende var der 20 af de 67 behandlede, der blev udskrevet, og 47, der ikke blev.

   To-gange-to tabellen ser derfor således ud:

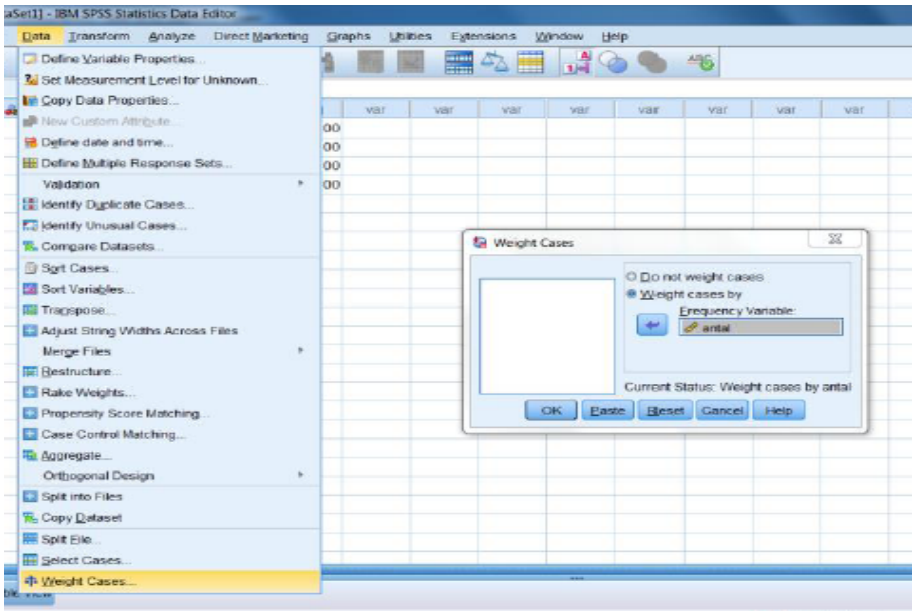
	Respons		
Gruppe	Hospitaliseret	Udskrevet	Total
Placebo	71	2	73
Prednisolone	47	20	67
	118	22	140

2. *Udregn χ^2-teststørrelsen for test af uafhægighed.*

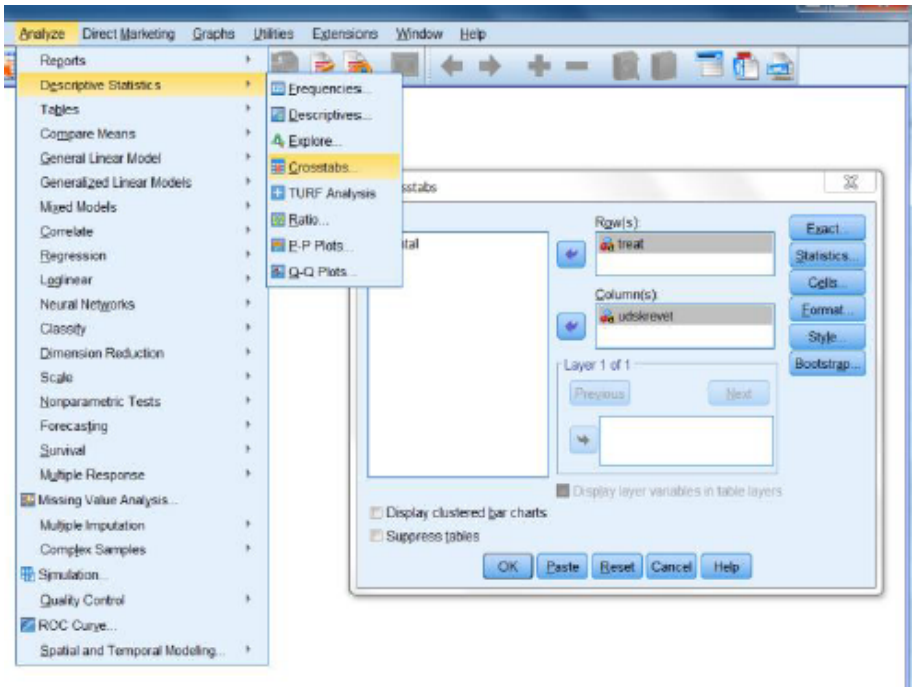
   Først skal vi indlæse tabellen svarende til ovenstående, i form af 4 observationer og 3 variable, som her kaldes `udskrevet` (ja/nej), `treat` (pred/plac) og `antal`. Nedenfor ses, hvordan datasættet samt variabeldefinitionerne tager sig ud



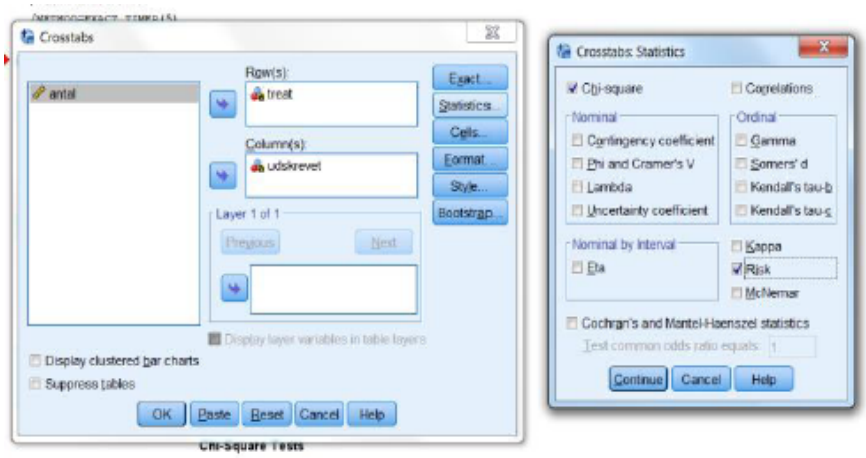
Før vi kan give os til at regne på disse data, skal vi lade programmet vide, at variablen `antal` angiver, hvor mange gange, den pågældende patienttype forekommer. Dette gøres ved at gå ind i `Data/Weight Cases`, afkrydse `Weight cases by` og sætte `antal` over i `Frequency Variable`:



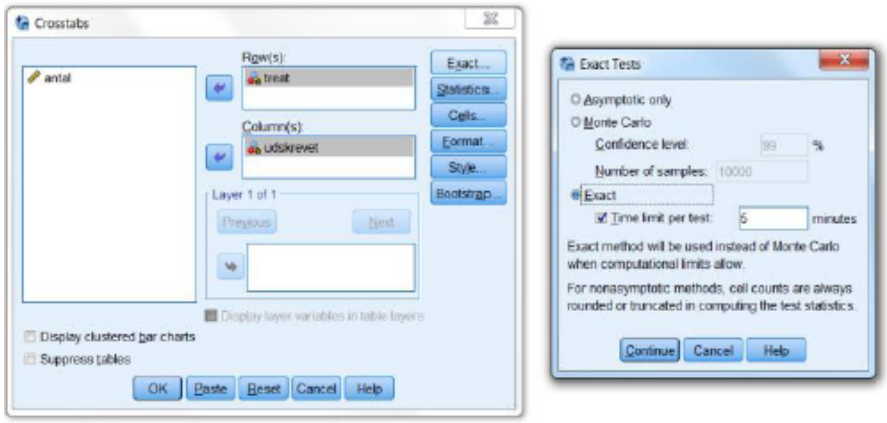
Herefter kan vi udføre χ^2-testet for uafhængighed ved at gå ind i `Analyze/Descriptive Statistics/Crosstabs/`, og sætte `treat` over i `Row(s)` og `udskrevet` i `Column(s)`



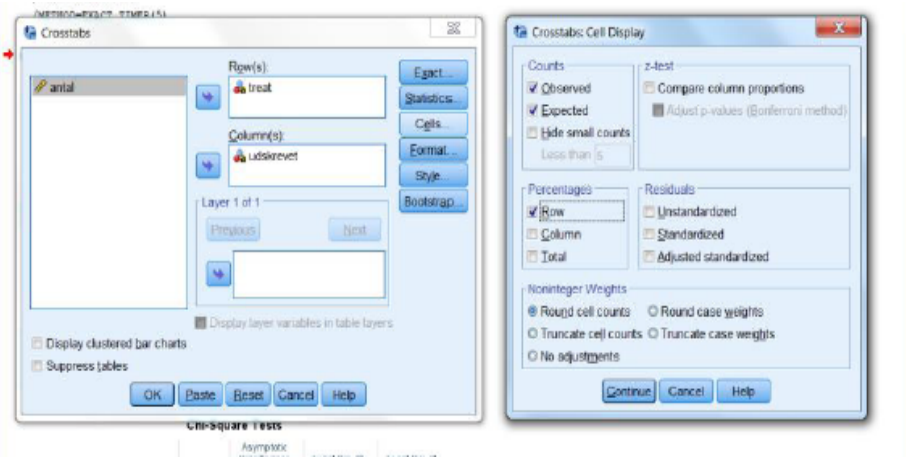
Vi går derefter ind i `Statistics` og afkrydser `Chi-square` (og vi afkrydser også `Risk`, fordi vi skal bruge det til et senere spørgsmål):



Desuden går vi ind i `Exact` og afkrydser `Exact`, så vi får Fishers eksakte test med:



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



Herved får vi (bl.a.) outputtet

treat * udskrevet Crosstabulation

			udskrevet		
			ja	nej	Total
treat	plac	Count	2	71	73
		Expected Count	11,5	61,5	73,0
		% within treat	2,7%	97,3%	100,0%
	pred	Count	20	47	67
		Expected Count	10,5	56,5	67,0
		% within treat	29,9%	70,1%	100,0%
Total		Count	22	118	140
		Expected Count	22,0	118,0	140,0
		% within treat	15,7%	84,3%	100,0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	19,387[a]	1	,000	,000	,000
Continuity Correction[b]	17,394	1	,000		
Likelihood Ratio	21,753	1	,000	,000	,000
Fisher's Exact Test				,000	,000
N of Valid Cases	140				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 10,53.

b. Computed only for a 2x2 table

Vi ser, at der er en stærkt signifikant forskel på de to behandlinger,

idet χ^2-størrelsen er 19.387, med $P = 0.000$.

3. *Var det nødvendigt at anvende Fishers eksakte test?*

   Når man skal vurdere hvorvidt det er nødvendigt at bruge Fisher's eksakte test, skal man se om de *forventede* værdier nogen steder er mindre end 5. De forventede værdier fremgår af tabellen ovenfor, og det ses, at de alle er en del større end 5 (den mindste er 10.5), så der intet til hinder for at bruge det sædvanlige χ^2-test. (Det, der har forvirret forfatterne, er selvfølgelig, at et af de *observerede* tal er mindre end 5).

4. *Kommenter de i resultatafsnittet angivne sikkerhedsintervaller.*

   Sikkerhedsintervallet for udskrivningssandsynligheden i placebo gruppen er lidt tosset, idet det er udregnet som estimatet plus minus 2 gange den spredning, der er baseret på en normalfordelingsapproksimation til Binomialfordelingen, uanset at de nedre grænser derved falder under 0. For placebo-gruppen kunne det være relevant at lave et eksakt konfidensinterval, men **det kan jeg ikke lokke SPSS til**. Det bliver (0.0033;0.0955), udregnet i SAS.

5. *Udregn/estimer (med 95% sikkerhedsgrænser):*

   - *Forskellen i udskrivningssandsynligheder.*

     Vi ser af rækkeprocenterne i tabellen ovenfor, at de to udskrivningssandsynligheder er hhv. 2.7% og 29.9%, dvs. med en forskel på 27.1%.

     For at få SPSS til at udregne denne, med tilhørende usikkerhed, går vi ind under `Statistics` og afkrydser `Somer's d`.

     Herved får vi output, der viser, at forskellen på ca. 27%, eller 0.27 har en usikkerhed på ca. 0.06, svarende til et konfidensinterval på (0.15, 0.39), altså en forskel på mellem 15% og 39%.

   - *Odds-ratio for udskrivning mellem prednisolon og placebo.*

I opsætningen af χ^2-testet ovenfor afkrydsede vi `Risk`, og vi får derfor outputtet

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for treat (plac / pred)	,066	,015	,297
For cohort udskrevet = ja	,092	,022	,378
For cohort udskrevet = nej	1,386	1,180	1,628
N of Valid Cases	140		

Bemærk, at alle de ovenfor angivne størrelser til sammenligning af de to grupper refererer til placebo vs. prednisolon (første vs. anden række i tabellen). For at svare på ordlyden af spørgsmålet, skal alle størrelser derfor inverteres.

Odds-ratio for udskrivning mellem placebo og prednisolone er 0.066 med et 95% konfidensinterval på (0.015;0.297). Den omvendte odds-ratio fås ved at tage den inverse, den bliver 15.2 (3.37;66.7).

- *Relativ risiko for udskrivning mellem prednisolon og placebo.*

  Den relative risiko for udskrivning er $\frac{1}{0.092} = 10.9$ med et 95% CI på $(\frac{1}{0.378}, \frac{1}{0.022}) = (2.65; 45.5)$.

- *Relativ risiko for hospitalisering mellem prednisolon og placebo.*

  Den relative risiko for hospitalisering for prednisolon vs. placebo er $\frac{1}{1.386} = 0.72$ med med et 95% konfidensinterval på $(\frac{1}{1.628}, \frac{1}{1.180}) = (0.61, 0.85)$.

  Bemærk at den relative risiko for hospitalisering har et konfidensinterval der er (relativt) noget snævrere end intervallerne for

OR og den relative risiko for udskrivning. Det skyldes i det væsentlige 2-tallet som for OR og RR for udskrivning er afgørende for nøjagtigheden af parameteren.

6. *Formuler en konklusion i ord ud fra hver af de 3 typer udregninger.*

   Sammenfattende kan man sige:

   - Andelen af patienter, der profiterer fra prednisolone er 27%. Et 95% konfidensintervallet for denne andel er (16%; 39%).
   - Forholdet mellem antallet af patienter der udskrives hhv. hospitaliseres er 15 gange større i prednisolone-gruppen end i placebogruppen. Et konfidensinterval for dette forhold er fra 3 til 68 gange.
   - Den relative risiko (chance) for at blive udskrevet for prednisolone-behandlede i forhold til placebo-behandlede er 11. Et 95% konfidensinterval for denne relative risiko er fra 2.6 til 45.
   - Den relative risiko for at blive hospitaliseret for placebo-behandlede i forhold til prednisolone-behandlede er 1.4. Et 95% konfidensinterval for denne relative risiko er fra 1.2 til 1.6.
   - Der er signifikant forskel på placebo- og prednisolone-behandling, $\chi^2 = 19.4$, p=0.000.

Opgave 2: Postoperative komplikationer

En traditionelt anvendt operationsprocedure har en **kendt** risiko for postoperative komplikationer på 20%.

1. *Hvad er sandsynligheden for at operere 10 konsekutive patienter med den traditionelle metode, uden at se nogen tilfælde af postoperative komplikationer?*

   Da komplikationssandsynligheden for den traditionelle metode vides at være p_0=20%, må sandsynligheden for en komplikationsfri operation være 1-0.2=0.8. Da udfaldene af de enkelte operationer antages at

være uafhængige af hinanden, må sandsynligheden for 10 konsekutive operationer uden komplikationer være

$$0.8^{10} = 0.107$$

altså 10.7%. Det sker altså jævnligt.

En ny metode er nu blevet foreslået, og den er netop blevet afprøvet på de første 10 patienter, uden at dette har givet anledning til nogen postoperative komplikationer.

2. *Er der tilstrækkelig evidens for at den nye operationsteknik er bedre end den traditionelle?*

I lyset af svaret på spørgsmål 1 ovenfor, må vi sige, at det ikke er særligt spektakulært at observere 10 konsekutive komplikationsfri operationer, og der kan derfor ikke være evidens for, at den nye metode er bedre end den gamle.

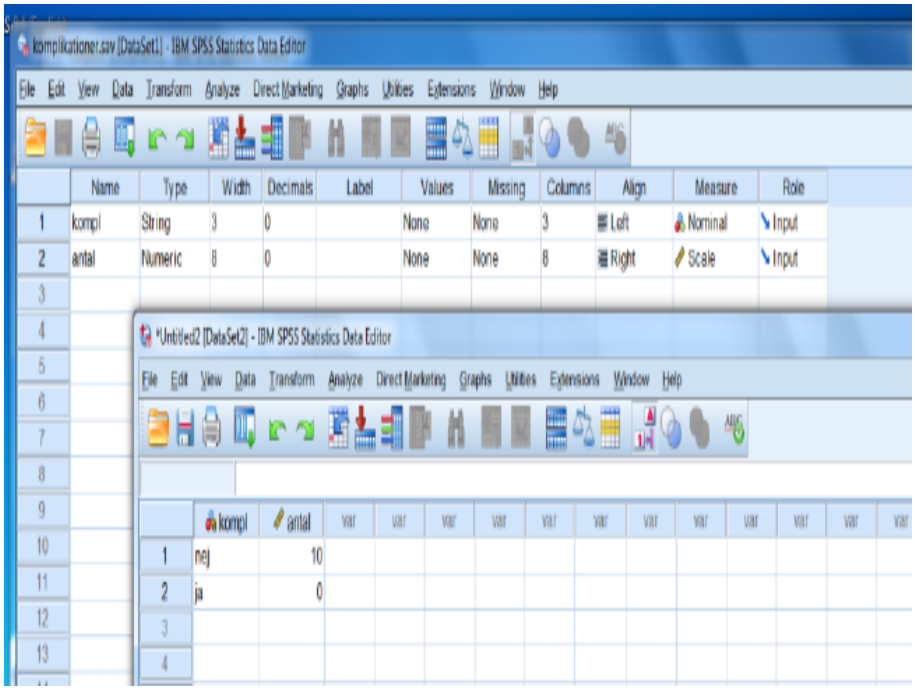
Faktisk vil de 10.7% netop være P-værdien for test af hypotesen

$$H_0 : p_1 = p_0$$

hvor p_1 betegner den ukendte sandsynlighed for komplikationer med den nye metode, **vel at mærke, hvis vi tester ensidigt**, dvs. hvis vi på forhånd er *sikre på*, at den nye metode i hvert fald ikke kan være *værre* end den gamle. P-værdien er jo netop *halesandsynligheden*, dvs. sandsynligheden for at observere *dette eller noget endnu mere ekstremt*, hvis hypotesen er sand. Og der *er* jo ikke noget, der er endnu mere ekstremt, hvis vi kun ser på den ene hale.

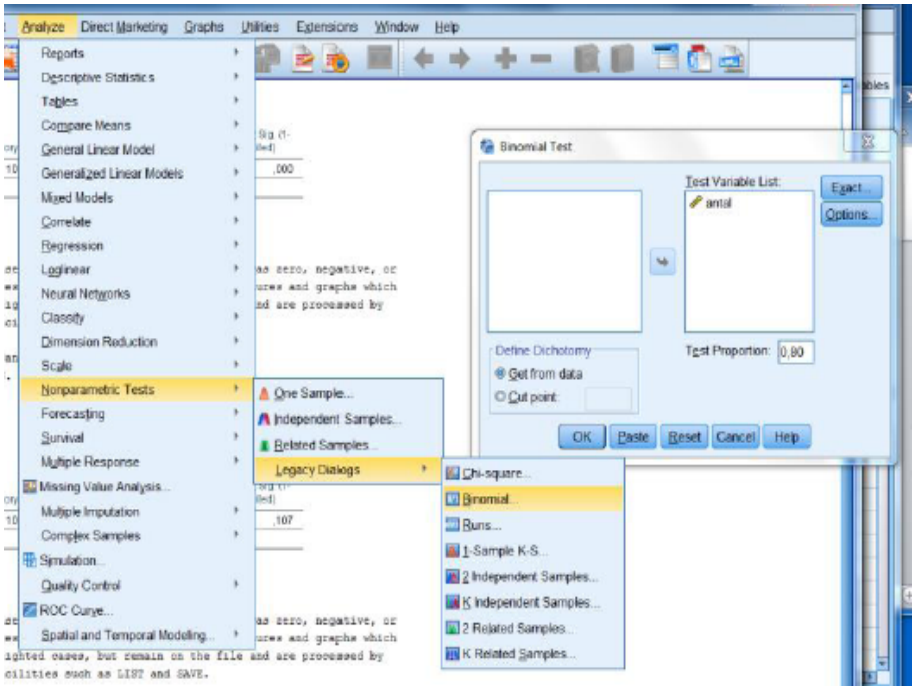
Det er sædvanligvis ikke rimeligt at teste ensidigt, så derfor er det mere rimeligt at sige, at P-værdi er det dobbelte, altså 0.214, eller 21.4%.

Lad os prøve at få SPSS til at svare på dette spørgsmål. Vi skal så have nogle data, og i stil med den forrige opgave, drejer det sig om 2 observationer, nemlig 0 komplikationer (`kompl=ja`) og 10 ikke-komplikationer (`kompl=nej`). Dette skriver vi således i datasættet:



Vi skal nu lave et test for om sandsynligheden for ikke-komplikationer kunne antages at være 0.8 (ligesom den gamle metode). Bemærk, at vi er nødt til at fokusere på ikke-komplikationer, da det er det eneste, der forekommer (når antallet er 0 for komplikationer, kan SPSS ikke "se" dem).

For at teste en sandsynlighed i Binomialfordelinegn, går vi ind i `Analyze/Nonparametric Tests/Legacy Dialogs/Binomial/`, sætter `antal` over i `Test Variable List` og skriver 0.8 i `Test proportion`:



hvorved vi får

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
antal	Group 1	10	10	1,0	,8	,107
	Total		10	1,0		

Desværre giver SPSS i dette tilfælde *også* et ensidigt test, så vi skal stadig selv gange med 2, så så får vi netop 21.5%, og vi kan derfor ikke forkaste vores hypotese, og der er dermed *ikke* evidens for, at den nye metode er bedre end den gamle.

Den tilhørende kode er ganske simpel

```
NPAR TESTS
/BINOMIAL (0.80)=antal
/MISSING ANALYSIS.
```

Vi benytter naturligvis det eksakte test til sådanne små datamaterialer, og vi burde også udregne et eksakt sikkerhedsinterval, men **det kan jeg ikke få SPSS til at udregne**.

Ved brug af SAS, udregnes det til (0.69,1.00), medens det tilsvarende approksimative sikkerhedsinterval, baseret på en approksimation af Binomialfordelingen til Normalfordelingen, er (1.00,1.00), hvilket naturligvis er helt vanvittigt og skyldes, at standard error her estimeres til 0.

3. *Hvor mange patienter skal man operere uden postoperative komplikationer før man med rimelig sikkerhed kan sige, at den nye metode er bedre end den traditionelle?*

Vi så ovenfor, at P-værdien for hypotesen om identitet af de to behandlingsmetoder kunne udregnes som

$$2 \times (1 - 0.2)^n$$

hvor vi med n betegner det antal personer, der observeres konsekutivt uden komplikationer. Hvis denne P-værdi er mindre end 0.05, vil vi forkaste vores hypotese. Vi skal således løse uligheden

$$2 \times (1 - 0.2)^n < 0.05$$

Dette kan gøres ved "*trial and error*" metoden, eller ved at tage logaritmer på begge sider af ulighedstegnet, hvorved vi får

$$\begin{aligned} \log(2) + n \times \log(0.8) &< \log(0.05) \\ n &> \frac{\log(0.05) - \log(2)}{\log(0.8)} = 16.53 \end{aligned}$$

Vi skal altså op på at observere mindst 17 konsekutive patienter uden at se en eneste komplikation, førend vi med noget, der ligner rimelig

sikkerhed at kunne konkludere, at den nye metode er bedre end den gamle.

Bemærk i øvrigt, at vi i udregningen ovenfor skiftede ulighedstegn fra $<$ til $>$, da vi dividerede med $\log(0.8)$. Det gjorde vi fordi $\log(0.8) = -0.22$, altså negativt.

Ved at ændre 10-tallet i vores data til hhv. 16 og 17, kan vi få SPSS til at bekræfte, at først ved 17 personer, bliver den ensidede P-værdi mindre end 2.5%:

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
antal	Group 1	16	16	1,0	,8	,028
	Total		16	1,0		

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
antal	Group 1	17	17	1,0	,8	,023
	Total		17	1,0		

Opgave 3: Børns sovevaner (DGA s. 274)

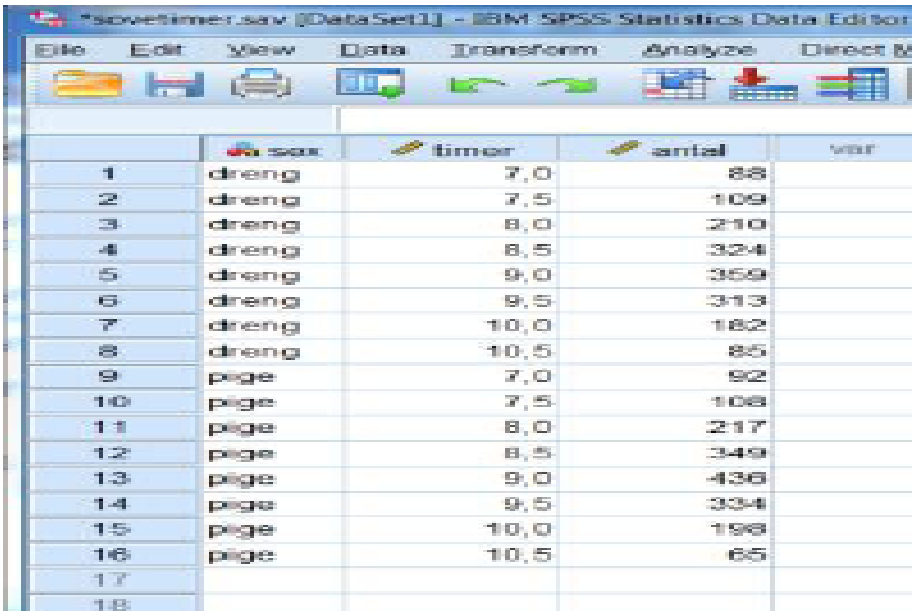
For 3469 børn på 14 år har man registreret antal timer tilbragt i sengen i løbet af et døgn (Macgregor & Balding, Ann. Hum. Biol., 15, 1988).

I nedenstående tabel ses data opdelt efter køn, idet tiderne er afrundet til nærmeste halve time.

Antal timer tilbragt i sengen

	≤ 7.0	7.5	8.0	8.5	9.0	9.5	10.0	> 10.0	Total
Drenge	88	109	210	324	359	313	182	85	1670
Piger	92	108	217	349	436	334	198	65	1799
Total	180	217	427	673	795	647	380	150	3469

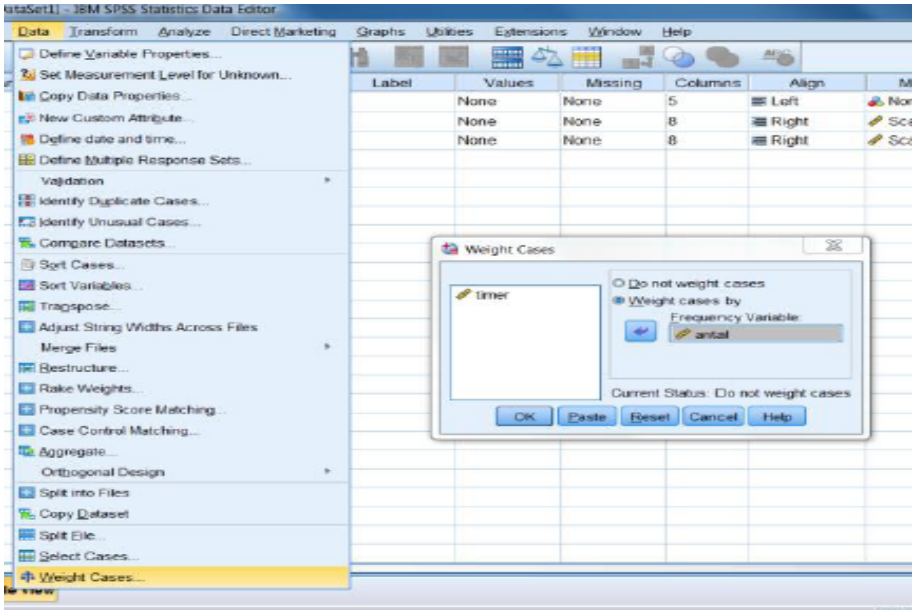
Data er indlæst i SPSS, så de ser således ud:



1. *Hvilke metoder kunne anvendes til at sammenligne fordelingen for piger og drenge?*

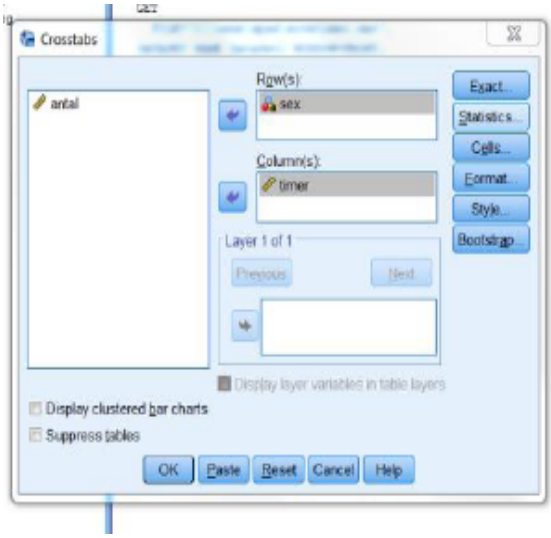
De to grupper (drenge hhv. piger) skal sammenlignes m.h.t. fordelingen af antal timer i sengen pr. døgn. Den naturlige sammenligning vil basere sig på den kumulative fordeling for hver af de to grupper, men for at lave et egentligt test, kunne man vælge at se på et sædvanligt χ^2-test, der dog ikke er særligt stærkt, dels på grund af de mange kategorier, og dels fordi det helt ignorerer ordningen mellem disse kategorier.

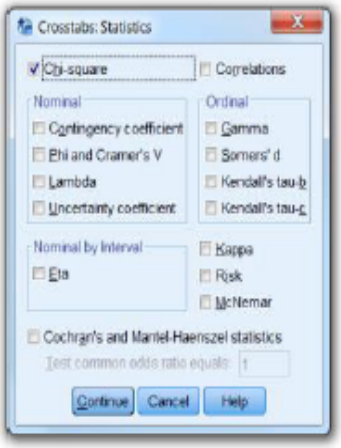
Når vi skal lave tabelanalyse i SPSS, skal vi vægte med `antal`, ligesom vi gjorde det i den første opgave. Vi går altså ind i `Data/Weight Cases`, afkrydser `Weight cases by` og sætter `antal` over i `Frequency Variable`:



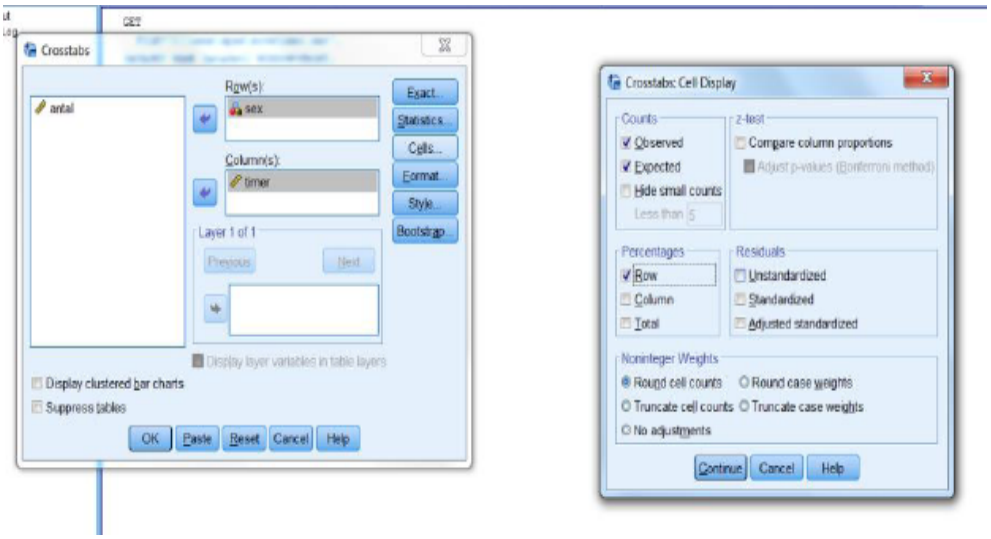
Herefter kan vi udføre χ^2-testet for uafhængighed (dvs. et test for om fordelingen i sovetime-kategorier er ens for drenge og piger) ved at gå ind i `Analyze/Descriptive Statistics/Crosstabs/`, sætte `sex` over i `Row(s)` og `timer` i `Column(s)`

Vi går derefter ind i `Statistics` og afkrydser `Chi-square`:





Herefter går vi ind i `Cells` for at afkrydse, at vi, foruden de observerede antal (`Observed`), gerne vil se de forventede antal (`Expected`), samt under `Percentages`: `Row`, for at få rækkeprocenter, dvs. fordelingen af sove-kategorier for hvert køn for sig (overvej, hvorfor søjleprocenter ikke ville være ligeså informative).



Vi finder så outputtet

sex * timer Crosstabulation

			timer				
			7,0	7,5	8,0	8,5	9,0
sex	dreng	Count	88	109	210	324	359
		Expected Count	86,7	104,5	205,6	324,0	382,7
		% within sex	5,3%	6,5%	12,6%	19,4%	21,5%
	pige	Count	92	108	217	349	436
		Expected Count	93,3	112,5	221,4	349,0	412,3
		% within sex	5,1%	6,0%	12,1%	19,4%	24,2%
Total		Count	180	217	427	673	795
		Expected Count	180,0	217,0	427,0	673,0	795,0
		% within sex	5,2%	6,3%	12,3%	19,4%	22,9%

sex * timer Crosstabulation

			timer			
			9,5	10,0	10,5	Total
sex	dreng	Count	313	182	85	1670
		Expected Count	311,5	182,9	72,2	1670,0
		% within sex	18,7%	10,9%	5,1%	100,0%
	pige	Count	334	198	65	1799
		Expected Count	335,5	197,1	77,8	1799,0
		% within sex	18,6%	11,0%	3,6%	100,0%
Total		Count	647	380	150	3469
		Expected Count	647,0	380,0	150,0	3469,0
		% within sex	18,7%	11,0%	4,3%	100,0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	7,831[a]	7	,348
Likelihood Ratio	7,839	7	,347
N of Valid Cases	3469		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 72,21.

Hvis man sammenligner procenterne i tabellen, ses ikke de store forskelle, og samme konklusion giver χ^2-testet (P=0.348), ret overbevisende i betragtning af de meget store tal, det drejer sig om.

2. *Er der nogen forskel på piger og drenge i denne henseende?*

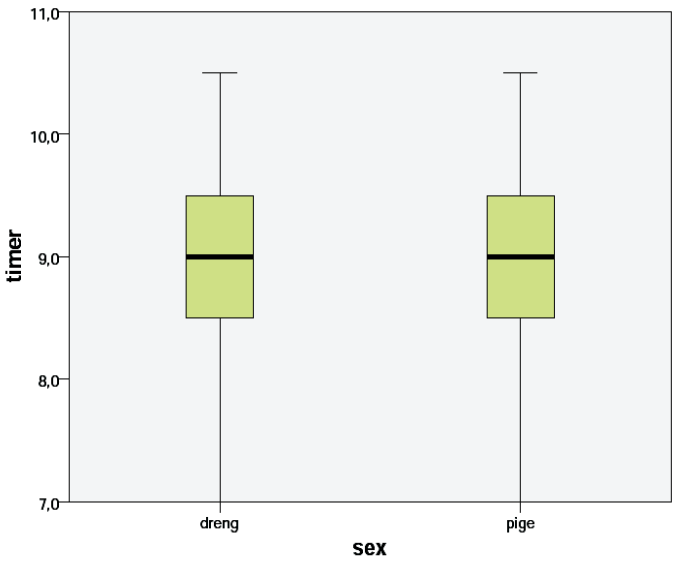
Det ser altså ikke sådan ud, men nu er χ^2-testet jo ikke særligt stærkt, når det drejer sig om så store tabeller.

Man kunne få den tanke at anvende et såkaldt **trend test**, da en af siderne i tabellen er en ordinal inddeling (tid i sengen). Det ville imidlertid svare til at vende årsagsammenhængen og undersøge om sandsynligheden for at være en dreng afhang af hvor lang tid man tilbragte i sengen.

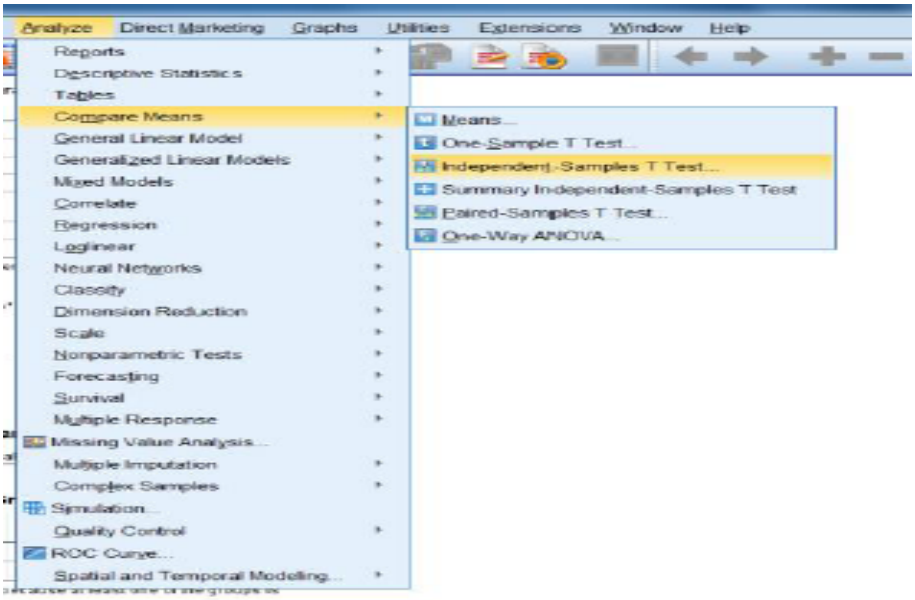
Sammenlign f.eks. til en situation, hvor et trend test *ville* være relevant, nemlig hvis inddelingerne var skostørrelse vs. risikoen for få kejsersnit. Humlen er: Hvad er outcome, og hvad er forklarende variabel?

En bedre ide ville være at lave et t-test, hvor man så må lade som om alle personer i samme celle i tabellen have samme værdi for antal timer i sengen. Det svarer blot til at at have timer som responsvariabel i en regressionsanalyse hvor man anvender sex som forklarende variabel. Nu opfatter vi altså timerne som en kvantitativ information, selv om den er afrundet (og derfor ikke giver en flot fordeling.

Når vi fastholder vægtningen med `antal`, kan vi nemt lave f.eks. et Boxplot til sammenligning af drenge og piger:



og for at lave det tilhørende T-test, går vi ind i
`Analyze/Compare Means/Independent Samples T Test`:



hvorved vi finder resultatet

Group Statistics

	sex	N	Mean	Std. Deviation	Std. Error Mean
timer	dreng	1670	8,853	,8838	,0216
	pige	1799	8,847	,8506	,0201

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
timer	Equal variances assumed	3,892	,049	,199	3467
	Equal variances not assumed			,199	3423,587

Independent Samples Test

		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence ... Lower
timer	Equal variances assumed	,842	,0059	,0295	-,0519
	Equal variances not assumed	,843	,0059	,0295	-,0520

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the ...	
		Upper	
timer	Equal variances assumed	,0636	
	Equal variances not assumed	,0637	

Forskellen mellem det antal timer drenge og piger tilbringer i sengen er gennemsnitligt 0.0059 timer eller ca. 21 sekunder. Et konfidensinterval for dette middeltal er (-0.0520, 0.0637) timer, dvs. (-3.1, 3.8) minutter. Af testet ses at der ikke er nogen signifikant forskel på drenge og piger hvad angår den gennemsnitlige tid de tilbringer i sengen. Af figuren så vi desuden, at der heller ikke visuelt er nogen forskel i fordelingen af den tid de to populationer tilbringer i sengen.