

SPSS handleiding

Geschreven door:

Saskia le Cessie

Klinische Epidemiologie/Medische Statistiek

LUMC

November 2012

1.	Inleiding.....	3
2.	Invoeren gegevens.....	3
2.1.	Definitie van de variabelen.....	4
2.2.	Coderen van categorische variabelen.....	4
2.3.	Meerdere data windows.....	5
2.4.	Het Output window.....	5
3.	Beschrijven van de gegevens.....	6
3.1.	Aantallen en percentages.....	6
3.2.	Gemiddelden, medianen, standaarddeviaties en andere kengetallen.....	7
4.	Maken van grafieken.....	7
4.1.	Puntenwolk, scatterplot.....	7
4.2.	Meerdere Y-variabelen in een scatterplot.....	8
4.3.	Histogram.....	8
5.	Data bewerkingen.....	9
5.1.	Sorteren.....	9
5.2.	Een deel van de gegevens selecteren.....	9
5.3.	Analyses voor subgroepen apart doen.....	9
6.	Constructie van nieuwe variabelen.....	9
6.1.	Bereken nieuwe variabelen: Compute.....	9
6.2.	Maak nieuwe categorieën: recode.....	10
7.	Numerieke gegevens: twee ongepaarde groepen (vergelijken gemiddelden of medianen).....	10
7.1.	Ongepaarde t-test.....	11
7.2.	Mann-Whitney test (Wilcoxon rank sum test).....	11
8.	Numerieke gegevens: meer dan twee ongepaarde groepen.....	12
8.1.	Variantieanalyse (ANOVA).....	12
8.2.	Kruskal-Wallis test.....	13
9.	Numerieke gegevens: twee gepaarde groepen.....	13
9.1.	Gepaarde t-test.....	13
9.2.	Wilcoxon signed rank test.....	13

10.	Vergelijken twee proporties/percentages, ongepaard	14
10.1.	Chi-kwadraat test	14
10.2.	Relatieve risico's en odds ratio's in een twee bij twee tabel.	14
10.3.	Invoeren bestaande tabel	15
11.	Vergelijken twee proporties/ percentages gepaard, McNemar's test.....	16
12.	Kruistabellen met meer dan twee categorieën	16
13.	Overlevingsdata met gecensureerde gegevens, Kaplan Meier krommes en log rank test	16
14.	Berekenen correlatiecoëfficiënt	16
15.	Overeenstemming	16
15.1.	Overeenstemming tussen twee categorische variabelen: kappa coefficient	16
15.2.	Overeenstemming tussen twee numerieke metingen: Bland Altman plot	17
16.	Lineaire regressie.....	17
17.	Logistische regressie.....	18
18.	Cox regressie	19
19.	Sneller werken met SPSS-Syntax.	20

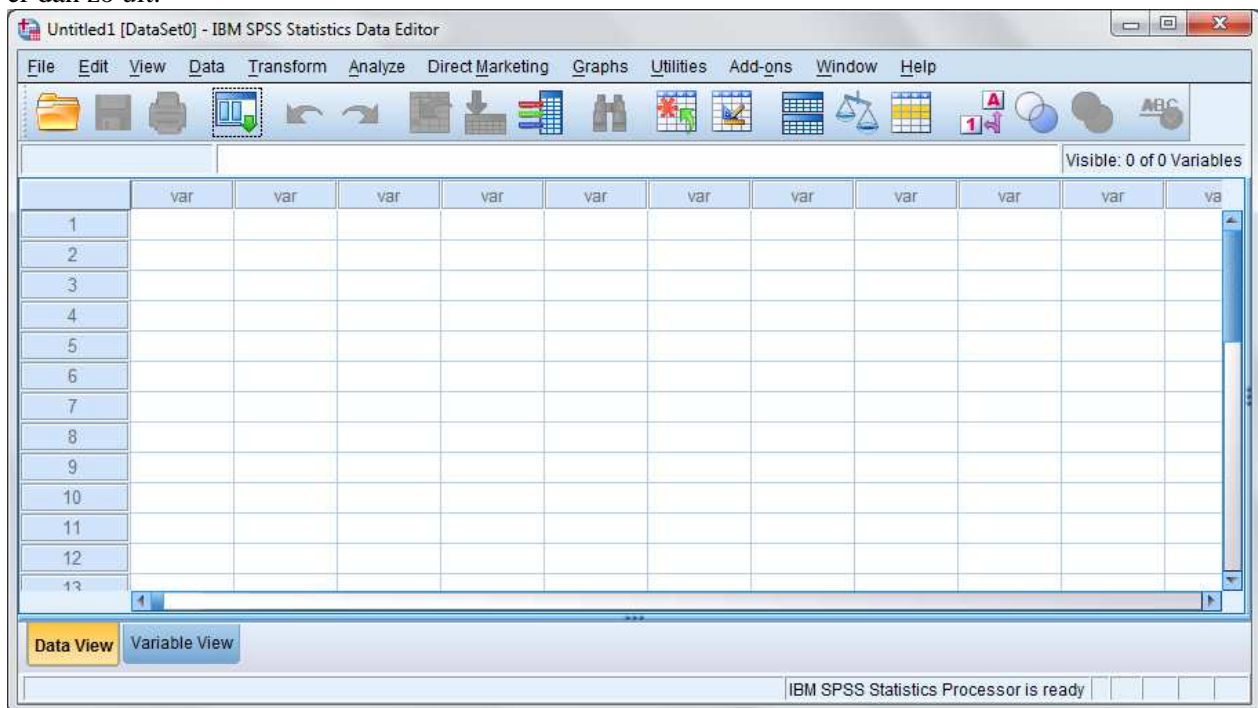
1. Inleiding

Deze handleiding geeft een inleiding in het gebruik van SPSS. De handleiding beperkt zich tot de binnen het medisch onderzoek meest gebruikte statistische methoden zoals beschrijvende statistiek, het maken van grafieken, de statistische methoden voor het vergelijken van twee of meer groepen waarnemingen, het berekenen van correlaties en het uitvoeren van lineaire, logistische en Cox proportional hazard regressie. Voor een uitgebreidere handleiding verwijzen we naar de uitgebreide SPSS help functies en naar de SPSS manuals.

Goede elementaire medisch statistiek boeken zijn: Petrie A en Sabin C. Medical statistics at a glance. Oxford: Blackwell Science, en Swinscow en Campbell Statistics at square one, online beschikbaar via de British Medical Journal

2. Invoeren gegevens

Start SPSS op en geef aan bij het introductiescherm dat u data wilt invoeren (Type in Data). Het scherm ziet er dan zo uit:



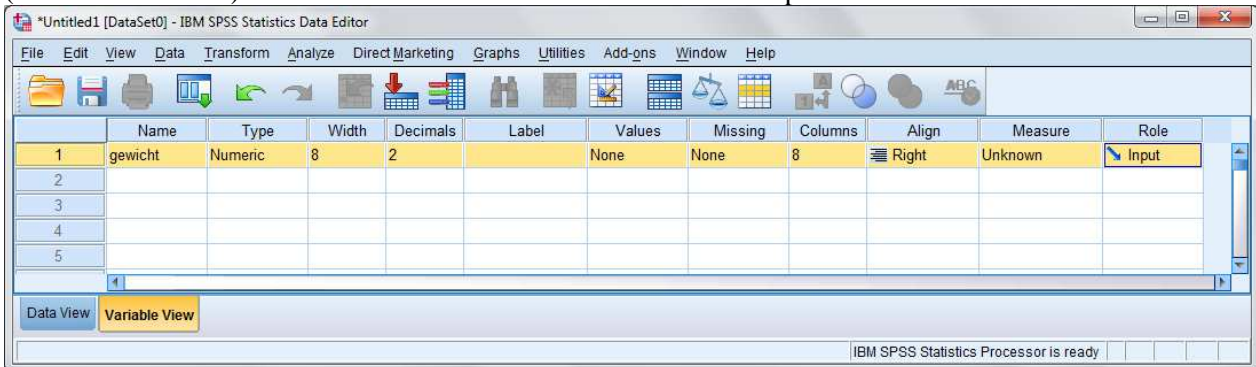
Dit is het data-window. Het invoeren tonen we met behulp van de gegevens van een 6-tal zwangerschappen.

Gewicht	Leeftijd	Geslacht
3036	28	meisje
3005	31	meisje
3152	32	meisje
3073	20	jongen
2882	30	meisje
2943	30	jongen

Elke regel bevat de gegevens van een bevalling: het geboortegewicht in grammen, de leeftijd van de moeder tijdens de bevalling in jaren en het geslacht van het kind.

2.1. Definitie van de variabelen

Klik voor het invoeren van de bovenstaande gegevens linksonder aan het scherm op Variable View. U krijgt dan een scherm met een aantal kolommen en rijen. Typ in de eerste rij onder het kopje 'name' de naam (maximaal 8 letters) van de eerste variabele: GEWICHT en druk op enter.



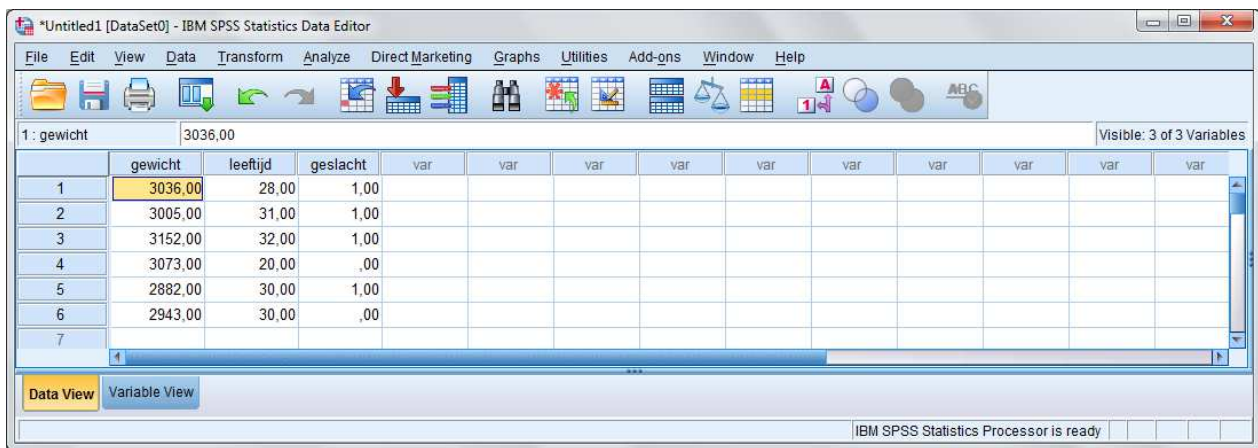
U ziet nu dat onder de andere kolommen ook informatie komt. Hier kunt u meer informatie over de variabele aangeven:

- Name: naam van de variabele.
- Type: het soort variabele (bv numeriek (getal), date (datum), of string (tekst)).
- Width: Totaal aantal cijfers of letters dat in de computeruitvoer getoond wordt
- Decimals: aantal cijfers achter de komma dat in het scherm getoond wordt.
- Label: een uitgebreidere beschrijving van de variabele. U kunt hier bijvoorbeeld "geboortegewicht van kind" invullen.
- Values: Bij categorische variabelen kunt u hier aangeven welke numerieke waarde er bij welke categorie komt (bv man=0, vrouw=1). Zie verderop voor meer informatie.
- Missing: Hierbij kunt u aangeven welke waarden missende gegevens representeren (U kunt bijvoorbeeld missende waarden met -9 aangeven)
- Measure: hier kunt u aangeven of de variabele numeriek (scale), ordinaal, of categorisch (nominal) is.
- Role: dit hoeft verder niet ingevuld te worden

2.2. Coderen van categorische variabelen

De variabele GESLACHT is een categorische variabele, met de categorieën 'jongen' en 'meisje'. Het werkt vaak makkelijker als dit met getallen gecodeerd wordt, dus bijvoorbeeld 1 voor meisje en 0 voor jongen. Het is dan wel handig om ervoor te zorgen dat u weet welk getal bij welke categorie hoort. Dit doet u door '**value labels**' te definiëren. Ga hiervoor met de muis naar het vakje onder het kopje Values en klik op het kleine vierkantje met de 3 puntjes dat rechts verschijnt. U krijgt nu het Value Labels venster. Vul bij 'value' '1' in en bij 'value label' 'meisje'. Klik 'ADD' aan. Vul verder in: 'value' '0', 'value label' 'jongen' en klik ADD, en vervolgens 'OK'.

Om de gegevens in te voeren moet u terug naar het Data View window (door links onderin op Data View te klikken). In dit window zijn de kolommen variabelen en de rijen individuen. Ga met de muis naar links boven (eerste rij, eerste kolom) en voer bovenstaande gegevens in. Gebruik hierbij voor de variabele GESLACHT cijfers en geen letters. Het scherm moet er als u alle data hebt ingevoerd als volgt uitzien:



U kunt uw data opslaan met de commando's File; Save en weer terughalen met de commando's File; Open. Zorg ervoor dat u als u data invoert regelmatig uw file savet.

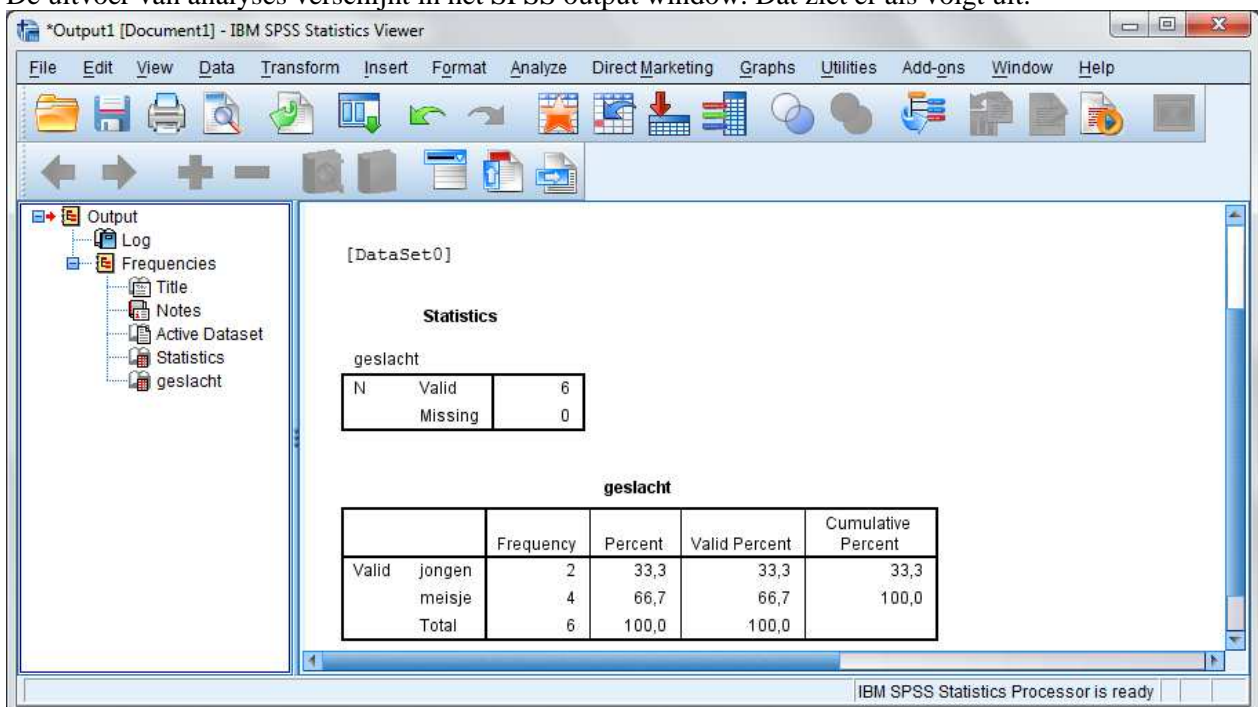
Wanneer er gegevens missen kunt u het hokje leeg laten. Een alternatief is om een missende waarde met een getal aan te geven (bv -9). In het Variable-view window moet u dan onder de kolom missing aangeven welke waarden missende gegevens representeren.

2.3. Meerdere data windows.

In SPSS kunnen meerdere windows met data tegelijk geopend worden. Deze worden aangegeven met DataSet0, DataSet1, DataSet2, enz. Met een groen kruisje wordt aangegeven welk van de dataset "active" is. SPSS commando's en berekeningen worden met de gegevens uit de actieve dataset uitgevoerd. Beginnende SPSS gebruikers wordt aangeraden om maar één data window te gebruiken.

2.4. Het Output window

De uitvoer van analyses verschijnt in het SPSS output window. Dat ziet er als volgt uit:



Dit window is in twee stukken verdeeld. Rechts vinden we de volledige uitvoer. Je kunt daar in bladeren m.b.v. de balk en de pijltjes uiterst rechts, of met de toetsen Page Up en Page Down. Dubbelklikken in het rechtergedeelte zorgt ervoor dat de uitvoer 'ge-edit' kan worden; er kunnen veranderingen in worden aangebracht.

Het linker gedeelte is een overzicht van de uitvoer. We zien daar dat de Frequency uitvoer uit vijf stukken bestaat: een Title, Notes, Active Dataset; Statistics en GESLACHT. Snel door de uitvoer bladeren kan dus ook door in het linker gedeelte op het juiste stuk te klikken. Een gedeelte selecteren en daarna op de Delete toets drukken zorgt ervoor dat het betreffende stuk permanent verdwijnt. Stukken kunnen ook tijdelijk uit de uitvoer weggelaten worden door op het minteken voor de uitvoer te klikken.

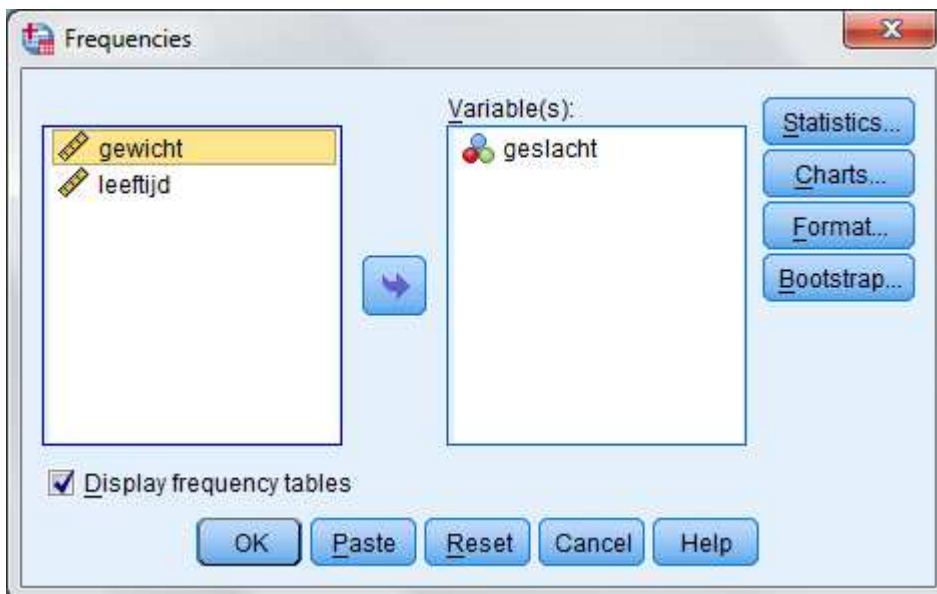
Een gedeelte van uw uitvoer kopiëren naar een ander pakket, zoals MS-Word gaat als volgt. Selecteer met de muis de uitvoer die gekopieerd moet worden. Kies dan in SPSS de commando's Edit; Copy (of gebruik de toetsencombinatie CTRL C). Ga dan naar het andere programma en kies daar de commando's Edit; Paste (of de toetsencombinatie CTRL V).

3. Beschrijven van de gegevens

3.1. Aantallen en percentages

Gebruik hiervoor de commando's Analyze; Descriptive Statistics; Frequencies

Verplaats de variabelen waar u een frequentieverdeling van wilt hebben naar variable(s) Om bijvoorbeeld een frequentieverdeling van geslacht te maken moet het scherm er als volgt uitzien.



Klik nu op het hokje OK. U ziet dan per categorie het aantal personen en het percentage van het totaal. Hieronder zie je de frequentieverdeling van de leeftijd.

Aantal observaties per categorie		% van totaal	% van totaal zonder de missende waarden	Som van de percentages van huidige en voorafgaande categorieën	
	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid	20,00	1	16,7	16,7	16,7
	28,00	1	16,7	16,7	33,3
	30,00	2	33,3	33,3	66,7
	31,00	1	16,7	16,7	83,3
	32,00	1	16,7	16,7	100,0
Total		6	100,0	100,0	

3.2. Gemiddelden, medianen, standaarddeviaties en andere kengetallen

Het gemiddelde, mediaan, standaarddeviatie, minimum, maximum en andere kengetallen kunnen op verschillende manieren berekend worden:

- Met de commando's Analyze, Descriptive Statistics, Frequencies, Kies de optie Statistics en geef aan welke kengetallen u berekend wilt hebben.
- Met de commando's Analyze, Descriptive Statistics, Descriptives. Bij het subcommando Options kunt u aanklikken wat u berekend wilt hebben.
- Met de commando's Analyze, Descriptive Statistics; Explore. Verplaats de variabele waarvan u de kengetallen wilt berekenen naar het hokje Dependent list. Wanneer u gemiddelden voor subgroepen wilt berekenen kunt u de variabele die de subgroepen aangeeft verplaatsen naar het hokje Factor List.

De commando's Frequencies en Explore kunnen ook percentielen berekenen. Het commando Explore berekent ook een 95% betrouwbaarheidsinterval voor het gemiddelde.

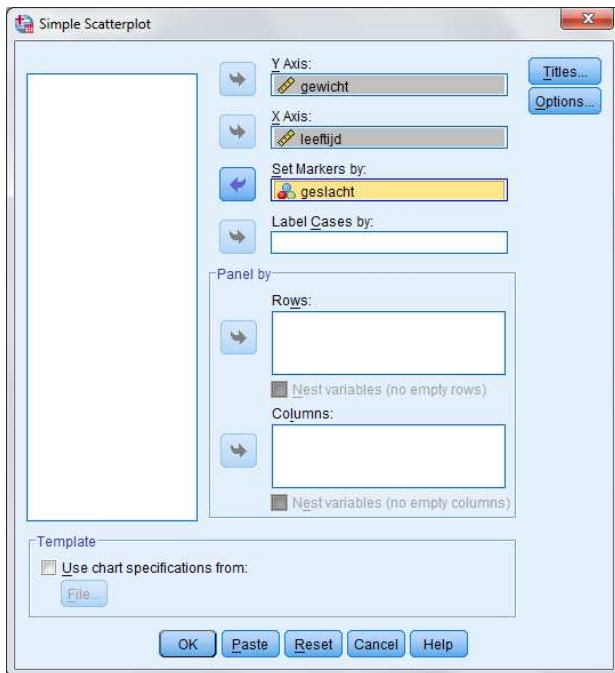
4. Maken van grafieken

In SPSS 20 kunnen grafieken op twee verschillende manieren gemaakt worden. Met de opties Graphs; Chart Builder kunnen interactief grafieken gemaakt worden. Maar vaak is het veel makkelijker om met de opties Graphs; Legacy Dialogs; te werken en dan het gewenste type grafiek te kiezen.

4.1. Puntenwolk, scatterplot




Kies de commando's Graphs; Legacy Dialogs; Scatter/Dots; Kies de optie Simple scatter en klik op Define. Verplaats de variabele die u op de x-as wil hebben naar het hokje X-axis en de variabele voor de Y-as naar het hokje Y-axis. Klik dan op OK.

Om verschillende subgroepen met aparte symbolen aan te geven moet u na de commando's Graphs; Legacy Dialogs; Scatter; Define, de variabele die de groepsindeling representeert naar het hokje Set Markers By verplaatsen.



Het plaatje kunt u editen, door twee keer op het plaatje te klikken. U komt dan in de SPSS-chart editor terecht. Met de knoppen X en Y boven het plaatje, of door twee keer op de x-as of de y-as te klikken de assen aanpassen en een goede schaalverdeling kiezen.

Verder opties zijn:

- Punten met een rechte lijn verbinden. Dit doet u door de knop 'Add interpolation line'  boven het plaatje gebruiken of met de opties Elements; Interpolation line.
- Regressielijn aanpassen door de punten
Gebruik de knop 'Add fit line at total'  om een regressielijn voor alle punten aan te passen, of de knop 'Add fit line at Subgroups'  wanneer aparte lijnen voor subgroepen wilt.

4.2. Meerdere Y-variabelen in een scatterplot

Soms zijn er meerdere Y-variabelen die samen in een plaatje getekend moeten worden. De gegevens zien er dan bijvoorbeeld als volgt uit:

X	Y ₁	Y ₂
1	1.9	1.96
2	3.2	2.46
3	4.1	3.07
4	4.2	3.80
5	4.7	4.64
...

Kies de commando's **Graph; Scatter;** ; Legacy Dialogs; **Overlay; Define**. Verhuis het eerste X-Y paar naar Y-X, en herhaal dat voor het tweede X-Y paar. Zorg ervoor dat de X, Y volgorde goed is ((X en Y kunt u omdraaien met swap-pair). Klik daarna op ok.

4.3. Histogram.

Een histogram kan op de volgende twee manieren gemaakt worden.

1. Met de commando's *Analyze, Descriptive Statistics, Frequencies, Chart*. Kies dan de optie *Histogram*. Eventueel kunt u een normale verdelingskromme in het histogram tekenen.
2. Met de commando's *Graphs; Legacy Dialogs; Histogram*.

5. Data bewerkingen

5.1. Sorteren.

Met de commando's *Data; Sort cases* kunt u variabelen sorteren. Verplaats de variabele(n) waarop u wilt sorteren naar *Sort by*.

5.2. Een deel van de gegevens selecteren

In het data window kan een deel van de gegevens geselecteerd worden met de commando's *Data; Select cases*. Kies vervolgens de optie *If condition is satisfied* en klik op de knop *If*. Typ in het scherm dat nu verschijnt welke gegevens u wilt behouden. Een paar voorbeelden:

leeftijd = 30	selecteer alleen de personen van 30 jaar
leeftijd <15	selecteer 0-14.999 jaar
leeftijd <= 15	selecteer 0-15 jaar
leeftijd >=30 and leeftijd < 40	selecteer 30-39,999 jaar

Klik daarna op *Continue*.

U kunt nog kiezen of u de niet geselecteerde gegevens tijdelijk wilt verwijderen (*filtered*) of permanent (*deleted*). Pas op: als u *deleted* kiest en daarna de dataset *save* zijn de observaties ook echt verdwenen.

5.3. Analyses voor subgroepen apart doen

Om analyses apart per subgroepen te doen (bv apart voor mannen en vrouwen een plaatje maken van gewicht tegen lengte) kiest men in het data-window de opties *Data; Split file*. Kies nu de optie *Compare groups* en verplaats de variabele die de groepsindeling weergeeft naar het hokje *Groups Based on*.

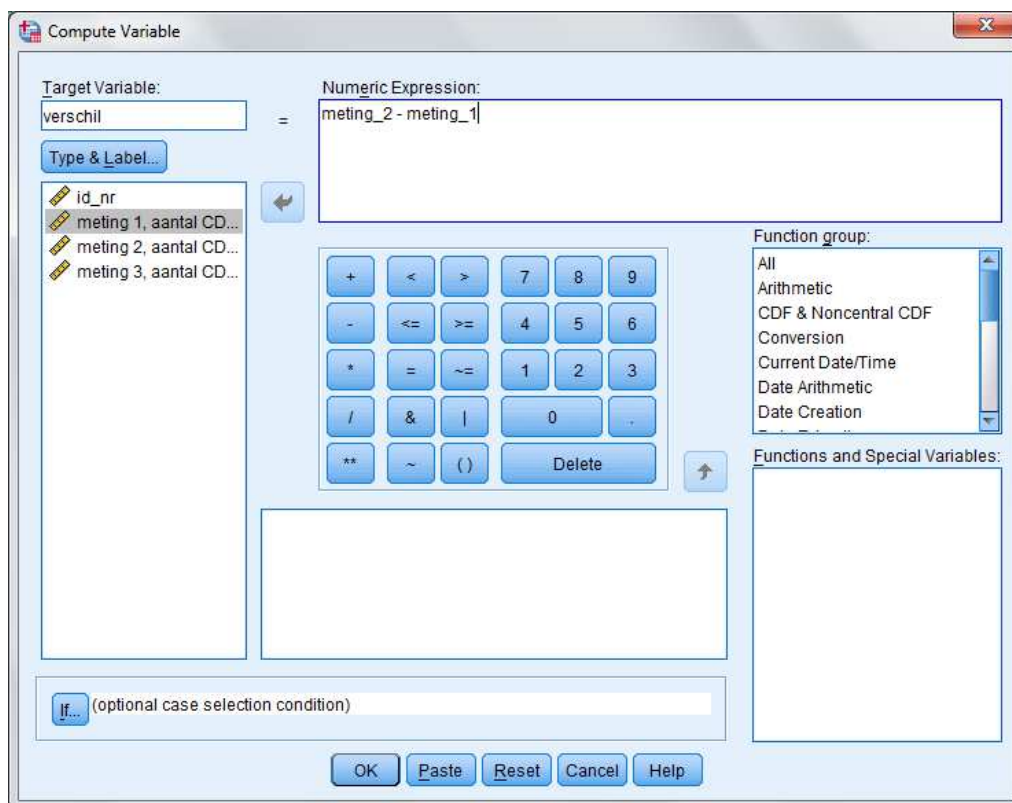
! Vergeet niet om *Split File* weer uit te zetten als u analyses weer op het hele bestand wilt doen. !

6. Constructie van nieuwe variabelen

6.1. Bereken nieuwe variabelen: Compute

Met de commando's *Transform; Compute* kunt u nieuwe variabele berekenen. In het hokje *Target Variable* typt u de naam van de nieuwe variabele, in het hokje *Numeric Expression* de berekening.

Stel bijvoorbeeld dat u twee variabelen *meting_1* en *meting_2* hebt en u wilt het verschil uitrekenen in een nieuwe variabele die u de naam 'verschil' wilt geven. Typ dan 'verschil' in het hokje *Target Variable* en *meting_1-meting_2* in het hokje *Numeric Expression*.



Ook kunt u functies van een variabele uitrekenen. Zo kunt u de wortel van een variabele x uitrekenen in een nieuwe variabele x wortel door de naam van de nieuwe variabele x wortel in het hokje Target Variable te typen en \sqrt{x} in het hokje Numeric Expression. Een lijst met de verschillende functies die SPSS kan uitrekenen staat in het Compute Variable-scherm onder Function group.

6.2. Maak nieuwe categorieën: recode

Met de commando's Transform; Recode into different variables, kunt u continue variabelen opdelen in categorieën (bv geboortegewicht in < 2500 , $2500-4000$ en > 4000 gram), en categorieën van categorische variabelen samenvoegen. Verplaats de variabele die u wilt transformeren naar het hokje Numeric Variable -> Output Variable, typ de naam van de nieuwe variabele in het hokje Output Variable en klik op change, en klik vervolgens op Old and New Variables. Geef nu een specifieke waarde van de oude variabele die gehercodeerd moet worden, of een range van waarden en geeft bij New Value de nieuwe waarde aan. Klik vervolgens op Add en herhaal dit tot alle waarden (inclusief missende) gehercodeerd zijn. Klik dan op Continue.

7. Numerieke gegevens: twee ongepaarde groepen (vergelijken gemiddelden of medianen)

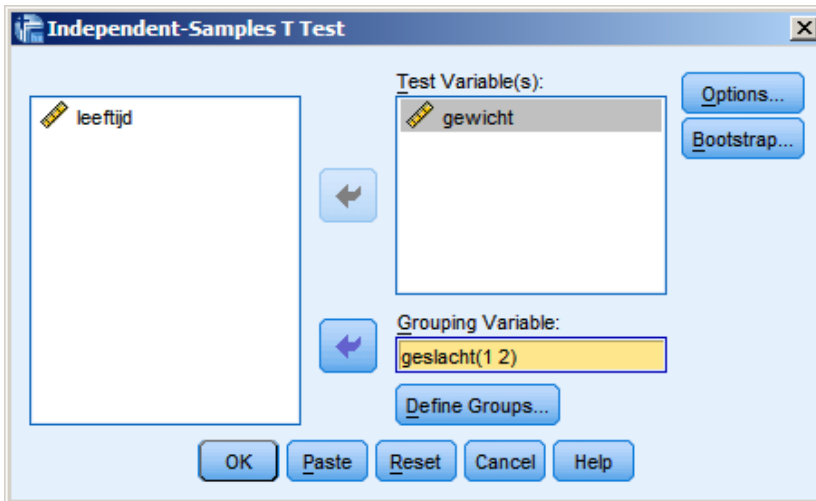
Allereerst moet u besluiten of u een twee steekproeven t-toets uitvoert of een niet parametrische toets (Mann Whitney). Een t-toets gebruikt u als

- Groepen groot zijn (vuistregel: meer dan 25 observaties per groep)
- De standaarddeviatie van de variabele die tussen de groepen vergeleken wordt in beide groepen ongeveer even groot is (vuistregel: standaarddeviatie in de ene groep is niet groter dan 4 keer die in de andere) .

Bij kleine aantallen moet de verdeling van de variabele redelijk normaal verdeeld zijn, met in beide groepen gelijke standaarddeviatie. Dat wil zeggen dat de mediaan ongeveer gelijk is aan het gemiddelde en er geen extreme uitschieters zijn.

7.1. Ongepaarde t-test

Kies de commando's Analyze; Compare Means; Independent Samples t-test. Verplaats de variabele die u tussen de groepen wilt vergelijken naar het hokje Test Variable(s) en de variabele die de groepen aangeeft naar het hokje Grouping Variable. Voor de variabele die de groepen aangeeft, moet u met Define Groups, aangeven met welke numerieke waarden de twee groepen aangeduid worden (u vult bijvoorbeeld 1 en 2 in als u mannen met vrouwen vergelijkt en de waarde 1 een man aangeeft en 2 een vrouw). Het scherm ziet er dan zo uit:



De uitvoer van een ongepaarde t-test ziet er zo uit:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
UITKOMST	Equal variances assumed	.597	.450	-1.782	18	.092	-1.5041	.8442	-3.2777	.2696
	Equal variances not assumed			-1.782	17.389	.092	-1.5041	.8442	-3.2822	.2740

↑

test voor gelijke sd's in de groepen

↑

p-waarde t-test

↑

95% betrouwbaarheidsinterval voor verschil gemiddelden

7.2. Mann-Whitney test (Wilcoxon rank sum test)

In SPSS 20 is een nieuwe routine voor niet-parametrische testen gekomen die “automatisch” de “correcte” test kiest. Het is af te raden om SPSS te laten bepalen welke statistische test er uitgevoerd gaat worden, dat kun je veel beter zelf bepalen. Vandaar dat we deze nieuwe routines niet gebruiken.

Om de Mann Whitney test uit te voeren kies je de commando's Analyze; Nonparametric Tests; LEGACY DIALOGS; 2 Independent Samples. Verplaats de variabele die u tussen de groepen wilt vergelijken naar het hokje Test Variable List en de variabele die de groepen aangeeft naar het hokje Grouping Variable. Met

Define Groups moet u aangeven met welke numerieke waarden de twee groepen aangeduid worden (dus bijvoorbeeld een 1 voor de mannen en een 2 voor de vrouwen). De uitvoer ziet er als volgt uit:


Ranks

	GROEP	N	Mean Rank	Sum of Ranks
UITKOMST	1.00	10	8.40	84.00
	2.00	10	12.60	126.00
	Total	20		

Test Statistics^b

	UITKOMST
Mann-Whitney U	29.000
Wilcoxon W	84.000
Z	-1.587
Asymp. Sig. (2-tailed)	.112
Exact Sig. [2*(1-tailed Sig.)]	.123 ^a

p-waarde Mann-Whitney test



a. Not corrected for ties.

b. Grouping Variable: GROEP

8. Numerieke gegevens: meer dan twee ongepaarde groepen

Ook hier is er weer de keuze tussen parametrische (variantie-analyse) of niet parametrische (Kruskal-Wallis) methoden. Bij de variantie-analyse wordt verondersteld dat de uitkomstvariabele bij benadering normaal verdeeld is en dat de spreiding in de verschillende groepen gelijk is.

8.1. Variantieanalyse (ANOVA)

Variantie-analyse kan in SPSS op verschillende manieren uitgevoerd worden


- Met de commando's Analyze, Compare Means; Means. Verplaats de variabele waar u de gemiddelden wilt vergelijken naar het hokje Dependent List en de variabele die de groepsindeling aangeeft naar Independent List. Klik op Options en vink vervolgens de optie Anova table and eta aan.
- Met de commando's Analyze, Compare Means; One-way ANOVA. Verplaats de variabele waar u de gemiddelden wilt vergelijken naar het hokje Dependent List en de variabele met de groepsindeling naar Factor.

De uitvoer ziet er als volgt uit:

ANOVA

UITKOMST	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	128.211	2	64.105	16.385	.000
Within Groups	105.635	27	3.912		
Total	233.846	29			

p-waarde



Om de gemiddelden paarsgewijs te vergelijken kunt u na de commando's Analyze, Compare Means; One-way ANOVA, met het subcommando Post Hoc, correcties voor herhaald toetsen (multiple testing) uitvoeren. Klik de correctie aan die u wilt uitvoeren.

8.2. Kruskal-Wallis test

Kies de commando's Analyze; Non parametric tests; LEGACY DIALOGS; K-independent samples. Verplaats de variabele waar u de gemiddelden wilt vergelijken naar het hokje Test Variable List en de groepsindeling naar Grouping Variable. Bij Define Range moet de hoogste en laagste waarde die de groepsvariabele aanneemt ingevuld worden.

9. Numerieke gegevens: twee gepaarde groepen

Ook hier moet u besluiten of u een gepaarde t-toets uitvoert of een niet parametrische toets (Wilcoxon signed rank toets). Bereken daartoe per paar het verschil en kijk naar de verdeling van de verschillen. Een gepaarde t-toets gebruikt u als

- Aantal paren groot is (vuistregel: meer dan 25)
- Bij kleine aantallen moet de verdeling van het verschil redelijk normaal verdeeld zijn, zonder extreme uitschieters.

9.1. Gepaarde t-test

We gaan er hier van uit dat de gepaarde waarnemingen naast elkaar staan. Kies de commando's Analyze; Compare Means; Paired Samples t-test en verplaats de twee gepaarde variabelen naar het hokje Paired Variables. De uitvoer ziet er als volgt uit:

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	VOOR - NA	-.5956	1.0570	.2364	-1.0903	-.1009	-2.520	19	.021

p-waarde

↓

95 % b. i. voor verschil
in gemiddelden

↑

9.2. Wilcoxon signed rank test

We gaan er weer vanuit dat de gepaarde waarnemingen naast elkaar staan. Kies de commando's: Analyze; Non parametric tests; LEGACY DIALOGS; 2 Related Samples. Verplaats de twee gepaarde variabelen naar het hokje Test pairs. In de uitvoer is "asyp. sig (2-tailed)" de p-waarde die bij de Wilcoxon signed rank test hoort.

10. Vergelijken twee proporties/percentages, ongepaard

Een kruistabel maakt u met de commando's Analyze; Descriptive Statistics; Crosstabs. Verplaats de variabelen voor de rijen en de kolommen naar de desbetreffende hokjes. Met het subcommando Cells kunt u rij en/of kolompercentages berekenen.

10.1. Chi-kwadraat test

Met de commando's Analyze; Descriptive Statistics; Crosstabs en daarna de optie Statistics kunt u verschillende statistische bewerkingen uitvoeren. Klik het hokje chi-square aan om de chi-kwadraat toets en de Fisher exact test te berekenen. De uitvoer voor de chi-kwadraattoets ziet er als volgt uit:

p-waardes

↓

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.049 ^b	1	.825		
Continuity Correction ^a	.000	1	1.000		
Likelihood Ratio	.048	1	.826		
Fisher's Exact Test				1.000	.568
Linear-by-Linear Association	.047	1	.828		
N of Valid Cases	38				

a. Computed only for a 2x2 table

b. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 2.74.

Enige uitleg:

Pearson Chi-Square:	de gewone chi-kwadraat test
Continuity Correction	De chi-kwadraat test met continuïteitscorrectie.
Likelihood Ratio:	Deze test is voor grote aantallen gelijk aan de gewone chi-kwadraat test
Fisher's Exact Test	Fisher exact test
Linear-by-Linear Association	Toetst of er een lineair verband is tussen de rijvariabele en de kolomvariabele. Kan nuttig zijn als de rij of kolomvariabele meer dan twee categorieën heeft

10.2. Relatieve risico's en odds ratio's in een twee bij twee tabel.

Met de commando's Analyze; Descriptive Statistics; Crosstabs, de optie Statistics en daarna het subcommando risk kunt u oddsratio's en relatieve risico's berekenen met 95% betrouwbaarheidsintervallen. Voor het relatieve risico worden de rij-percentages op elkaar gedeeld. Zorg er daarom voor bij het berekenen van een relatief risico dat de blootstelling als rijvariabele, en de uitkomst als kolomvariabele. Een simpel voorbeeld:

geslacht * ziek Crosstabulation

Count

		ziek		Total
		,00	1,00	
geslacht	.00 jongen	1	1	2
	1.00 meisje	3	1	4
Total		4	2	6

In deze tabel zien we snel dat het risico op ziekte voor meisjes 0,25 is en voor jongens 0,50. Jongens hebben dus een twee keer zo groot risico op ziekte. De uitvoer die SPSS geeft is:

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for geslacht (.00 jongen / 1.00 meisje)	,333	,009	11,939
For cohort ziek = ,00	,667	,149	2,979
For cohort ziek = 1,00	2,000	,224	17,894
N of Valid Cases	6		

Bij “For cohort ziek = 1,00” staat het relatief risico op ziek=1 voor de bovenste rij (geslacht = 0, jongen) versus de onderste rij (meisje). Bij “For cohort ziek = 0,00” staat het relatief risico op ziek=0 voor de bovenste rij (jongen) versus de onderste rij (meisje). Voor jongens is de kans om niet ziek te zijn 0,50, voor meisjes 0,75. Dat geeft het relatieve risico op gezond zijn van 0,667. Dit kan soms verwarrend zijn dus controleer deze relatieve risico’s altijd met de hand.

10.3. Invoeren bestaande tabel

Hieronder een voorbeeld van het invoeren van een bestaande kruistabel. In een onderzoek naar hersenvliesontsteking bij kinderen heeft men gegevens over etiologie en leeftijd in onderstaande tabel samengevat.

Etiologie	leeftijd in jaren		totaal
	1	2	
meningococcus	96	57	153
streptococcus	7	13	20
palumococcus	15	9	24
Total	118	79	197

Invoeren in SPSS kan op twee manieren.

De eerste manier is om voor elk kind afzonderlijk de gegevens in te voeren (197 !). Veel sneller is het om evenveel regels in het data window te gebruiken als er rijen en kolommen zijn (hier $3 \times 2 = 6$). Per regel geeft men dan de code voor rij en kolom aan en hoe vaak deze combinatie van rij en kolom is voorgekomen. De invoer van de data wordt dan:

etiologie	leeftijd	aantal
1	1	96

1	2	57
2	1	7
2	2	13
3	1	15
3	2	9

Voer deze data in in het data-window. Kies na het invoeren op de commandobalk Data en vervolgens Weight Cases. Klik nu op het rondje voor Weight Cases by, en verplaats de variabele aantal naar het Frequency Variable hokje. Klik dan op OK.

Nu weet SPSS dat er 96 kinderen waren van 1 jaar met een meningococcus infectie (etiologie=1), etc.

11. Vergelijken twee proporties/ percentages gepaard, McNemar's test

Hierbij moeten de gepaarde waarnemingen naast elkaar staan. Voer McNemar's toets uit door op de commandobalk aan te klikken: Analyze; Nonparametric Tests; LEGACY DIALOGS; 2-related samples. Verplaats de twee gepaarde variabelen naar het Test pair(s) list hokje. Vink vervolgens het hokje voor McNemar aan en het hokje voor Wilcoxon af.

12. Kruistabellen met meer dan twee categorieën

Een kruistabel maken en een chi-kwadraat test uitvoeren gaat op dezelfde manier als in paragraaf 10 beschreven staat voor een 2 bij 2 tabel.

13. Overlevingsdata met gecensureerde gegevens, Kaplan Meier krommes en log rank test

Kaplan-Meier krommes maakt u met de commando's Analyze; Survival; Kaplan-Meier. Verplaats de variabele die de overlevingsduur aangeeft naar het hokje Time en de censurerings variabele (de variabele die aangeeft of een individu overleden is of gecensureerd) naar het hokje Status. Met Define Event moet u aangeven welke waarde correspondeert met een echt event. Door bij Options de optie Plot; Survival, zorgt u er voor dat SPSS ook een plaatje van de geschatte overlevingskromme maakt.

Om de Kaplan-Meier krommes van meerdere groepen in een figuur te tekenen, moet u de variabele die de groepsindeling aangeeft verplaatsen naar het hokje Factor. Een log rank toets voert u uit met de optie Compare factor. Vink vervolgens log rank test aan.

14. Berekenen correlatiecoëfficiënt

Kies Analyze; Correlate; Bivariate; om correlaties uit te rekenen. Verplaats de variabelen die u wilt correleren naar Variables, en klik aan welk type correlatie u wilt berekenen (de Pearson correlatiecoëfficiënt als de variabelen redelijk normaal verdeeld zijn, of de Spearman correlatiecoëfficiënt bij niet normaal verdeelde variabelen).

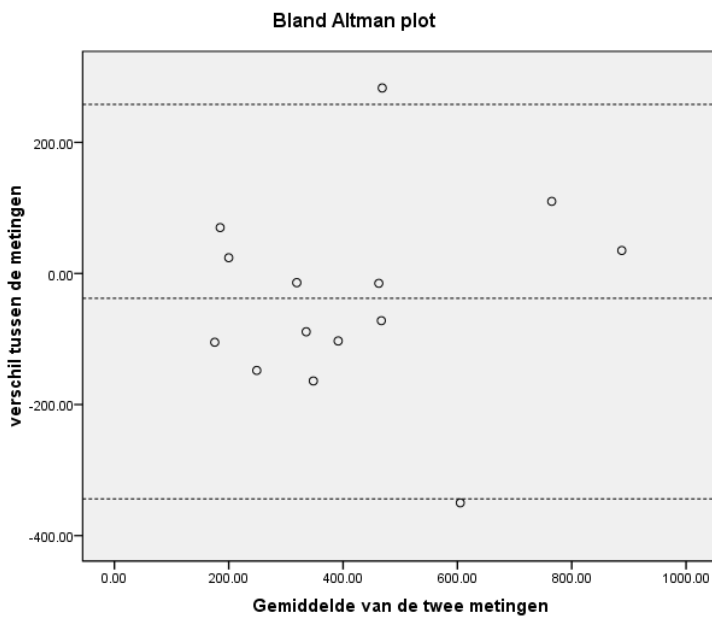
15. Overeenstemming

15.1. Overeenstemming tussen twee categorische variabelen: kappa coefficient

De kappacoëfficiënt kan berekend worden door een kruistabel te maken (Analyze; Descriptive Statistics; Crosstabs), de optie Statistics te kiezen en daar Kappa aan te vinken.

15.2. Overeenstemming tussen twee numerieke metingen: Bland Altman plot

Overeenstemming tussen twee numerieke variabelen kan in een Bland Altman plot weergegeven worden. Er is hiervoor geen optie in SPSS. Om een Bland-Altman plot te maken moeten twee nieuwe variabelen berekend worden, een variabele voor het gemiddelde van de twee metingen en een variabele voor het verschil van de twee numerieke variabelen. Dat kunt u doen zoals beschreven in paragraaf 6.1. Daarna kan het verschil tegen het gemiddelde uitgezet worden in een scatterplot (zie voor het maken van een scatterplot paragraaf 4.1). De grenzen van overeenstemming (limits of agreement) zijn gedefinieerd als $d - 2SD$ en $d + 2SD$ met d het gemiddelde verschil en SD de standaarddeviatie (SD) van de verschillen. Deze grenzen kunnen ook in de scatterplot gezet worden, door het plaatje te editten (dubbelklikken op plaatje) en vervolgens met Options; Y Axis reference line, twee horizontale lijnen te tekenen op de plaats van de onder- en boven limit of agreement.



16. Lineaire regressie

Kies de commando's **Analyze**; **Regression**: **Linear**. Voer vervolgens de uitkomstvariabele (ook wel Y of afhankelijke variabele genoemd) naar het hokje Dependent en de onafhankelijke variabele(n) (ook wel X, exposure, of risicovariabele) naar het hokje Independent.

Klik op OK. Onder het kopje coëfficiënt vindt u de schatting van richtingscoëfficiënt en intercept van de regressielijn:

Model		Coefficients ^a		Toets of coëfficiënten significant van 0	
		Unstandardized Coefficients	Standardized Coefficients	t	Sig.
1	(Constant)	1.098		9.345	.000
	X	2.196E-02	.417	2.101	.048

a. Dependent Variable: Y

schattingen intercept en richtingscoef.
geschatte regressielijn is
 $y = 1.098 + .02196 X$

Niet belangrijk

In de uitvoer onder het kopje Model summary vindt u de verklaarde variatie (R^2) en de schatting van de residuele spreiding rond de regressielijn (σ_{res}).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.417 ^a	.174	.134	.21670

a. Predictors: (Constant), X

R^2

σ_{res}

Om 95% betrouwbaarheidsintervallen rond de regressiecoëfficiënten laten berekenen, kiest u in het linear regression schermje de optie Statistics en vinkt u vervolgens Confidence Intervals aan.

17. Logistische regressie

Om een logistische regressie uit te voeren kiest u de opties: Analyze; Regression; Binary Logistic. Verplaats vervolgens de binaire uitkomst variabele naar het hokje Dependent en de X-variabele(n) naar het hokje independent. Wanneer er X-variabelen categorisch zijn (bv haarkleur) kunt u dat aangeven door op categorical te klikken en vervolgens de categorische variabelen naar Categorical Covariates te verplaatsen.

Standaard worden oddsratio's t.o.v. de laatste categorie berekend. Door bij Change Contrasts, het bolletje voor First aan te klikken kunt u dit veranderen in de eerste categorie.

Wanneer u de logistische regressie uitvoert verschijnt een heleboel uitvoer. Vooral van belang zijn de schattingen van de regressiecoëfficiënten en de oddsratio's:

Variables in the Equation							Odds ratio voor leeftijd
		B	S.E.	Wald	df	Sig.	Exp(B)
Step	LEEFTIJD	-.376	.115	10.750	1	.001	.687
1	Constant	4.428	1.049	17.805	1	.000	83.755

a. Variable(s) entered on step 1: LEEFTIJD.

Regressiecoëfficiënten

Met de optie Options kunt u 95% betrouwbaarheidsintervallen rond de oddsratio's laten berekenen.

Wanneer u categorische X-variabelen in het model heeft is het belangrijk om te controleren welke codering van de categorieën gebruikt is. De codering vindt u onder Categorical Variable Coding:

Categorical Variables Codings

	Frequency	Parameter coding			
		(1)	(2)	(3)	(4)
HAARKLEU .00 rood	51	1.000	.000	.000	.000
1.00 groen	42	.000	1.000	.000	.000
2.00 blond	89	.000	.000	1.000	.000
3.00 bruin	127	.000	.000	.000	1.000
4.00 zwart	49	.000	.000	.000	.000

Hier is zwart de referentiecategorie want daar zijn alle parameter codings 0. Dat betekent dat alle oddsratio's ten opzichte van de haarkleur zwart uitgerekend. Haarkleur(1) geeft de odds ratio van rood versus zwart weer, haarkleur(2) van groen versus zwart, enz.

In dit voorbeeld gaf dit:

p-waarde toets of de variabele haarkleur het model significant verbetert

oddsratio rood t.o.v. zwart

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step	HAARKLEU			10.135	4	.038	
1	HAARKLEU(1)	.495	.474	1.090	1	.296	1.640
	HAARKLEU(2)	-.725	.443	2.682	1	.101	.484
	HAARKLEU(3)	.259	.403	.412	1	.521	1.295
	HAARKLEU(4)	.347	.382	.824	1	.364	1.414
	Constant	.916	.316	8.396	1	.004	2.500

a. Variable(s) entered on step 1: HAARKLEU.

18. Cox regressie

Een Cox proportional hazard model voert u uit met de commando's Analyze; Survival; Cox regression.

Verplaats de variabele die de overlevingsduur aangeeft naar het hokje Time en de censureringsvariabele naar het hokje Status. Met Define Event moet u aangeven welke waarde correspondeert met een echt event. De X-variabelen in het model verplaatst u naar het hokje Covariates. Wanneer er X-variabelen categorisch zijn (bv haarkleur) kunt u dat aangeven door op categorical te klikken en vervolgens de categorische variabelen naar Categorical Covariates te verplaatsen.

Standaard worden hazardratios t.o.v. de laatste categorie berekend. Door bij Change Contrasts, het bolletje voor First aan te klikken kunt u dit veranderen in de eerste categorie.

Wanneer u de Cox regressie uitvoert verschijnt veel uitvoer. Vooral van belang zijn de schattingen van de regressiecoëfficiënten en de hazardratio's:

	B	SE	Wald	df	Sig.	Exp(B)
LEEFTIJD	-.001	.109	.000	1	.993	.999

Met Options kunt u 95% betrouwbaarheidsintervallen rond de hazardratio's laten berekenen.

Wanneer u categorische X-variabelen in het model heeft is het belangrijk om te controleren welke codering van de categorieën gebruikt is. De codering vindt u onder Categorical Variable Coding:

Categorical Variable Codings^{a,b}

	Frequency	(1)	(2)	(3)	(4)
HAARKLEU .00=rood	51	1.000	.000	.000	.000
1.00=bruin	42	.000	1.000	.000	.000
2.00=zwart	89	.000	.000	1.000	.000
3.00=grijs	127	.000	.000	.000	1.000
4.00=blond	49	.000	.000	.000	.000

a. Indicator Parameter Coding

b. Category variable: HAARKLEU

Hier is blond de referentiecategorie want daar zijn alle parameter codings 0. Dat betekent dat alle hazardratio's ten opzichte van de haarkleur blond uitgerekend. Haarkleur(1) geeft de hazardratio van rood versus blond weer, haarkleur(2) van bruin versus blond, enz.

In dit voorbeeld gaf dit:

**p-waarde toets of de variabele
haarkleur het model significant
verbetert**

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
HAARKLEU			9.901	4	.042	
HAARKLEU(1)	.076	.379	.041	1	.840	1.075
HAARKLEU(2)	-.860	.329	6.811	1	.009	.423
HAARKLEU(3)	-.520	.317	2.685	1	.101	.594
HAARKLEU(4)	-.196	.291	.453	1	.501	.822

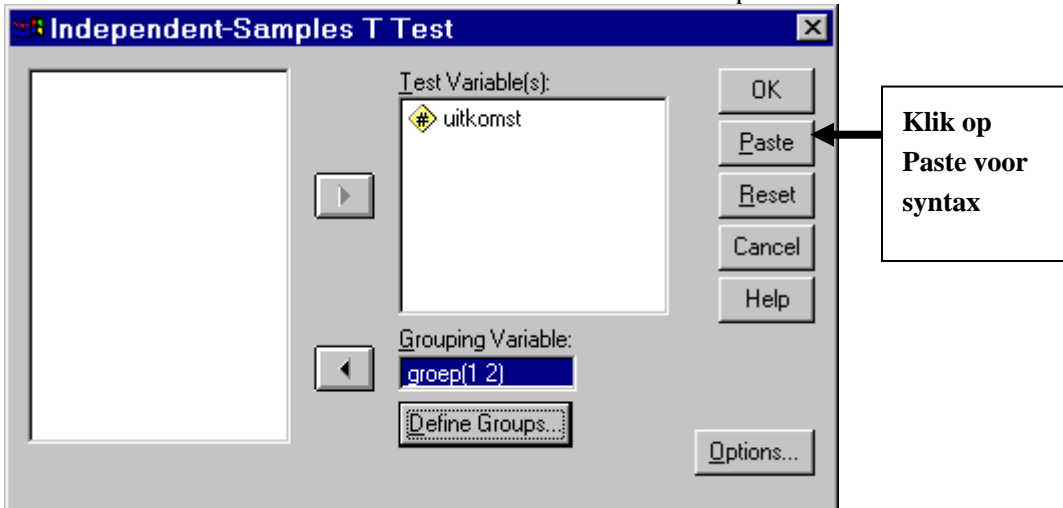
**hazardratio rood
t.o.v. blond**

19. Sneller werken met SPSS-Syntax.

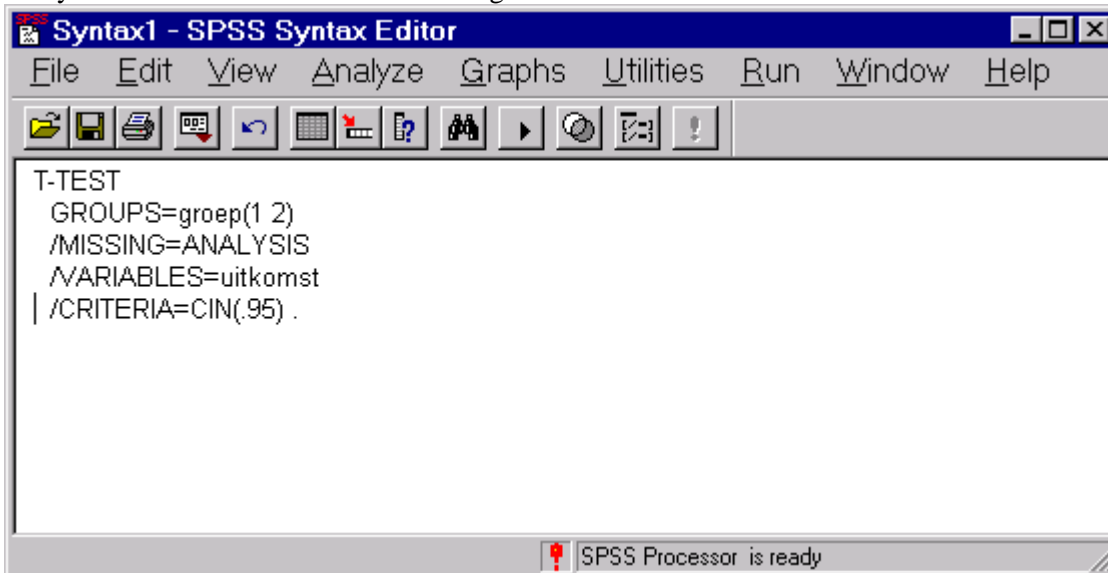
De menu structuur van SPSS is gebruikersvriendelijk maar wanneer er vaker dezelfde analyses uitgevoerd moeten worden wordt het geklik en geselecteer van alle subcommando's erg omslachtig. Achter het menu-gestuurde SPSS zitten SPSS commando's, de zogenaamde syntax. Standaard ziet u deze syntax ook in de uitvoer staan.

Om te documenteren wat u gedaan heeft en om er voor te zorgen dat u analyses eenvoudig kunt herhalen, is het erg handig om de syntax van commando's te bewaren. Dit kunt u doen door na het invullen van een analysescherm niet op OK te klikken maar op Paste. In plaats van de opdracht uit te voeren verschijnt nu de opdracht in het syntax window.

Hieronder ziet u als voorbeeld het uitvoeren van de twee steekproeven t-toets:



De syntax voor deze t-toets ziet er als volgt uit:



De syntax kunt u markeren en met de > knop uitvoeren.

De syntax kunt u opslaan en bij een nieuwe sessie van SPSS weer binnenhalen en uitvoeren. De syntax kan ook ge-edit worden. Houd er wel rekening mee dat SPSS commando's altijd eindigen met een punt. U kunt uw syntax ook van commentaar voorzien door een regel met een * te beginnen. Hieronder een voorbeeld van een SPSS-syntax programma.

* inlezen van de data.

```
get file= 'c:\files\voorbeeld.sav'.
```

*frequentieverdeling van geslacht.

```
FREQUENCIES
VARIABLES=geslacht
/ORDER= ANALYSIS .
```

* kruistabel geslacht versus groepsindeling, met chi-kwadraattest.

```
CROSSTABS
```

```
/TABLES=geslacht BY groep  
/FORMAT= AVALUE TABLES  
/STATISTIC=CHISQ  
/CELLS= COUNT ROW .
```