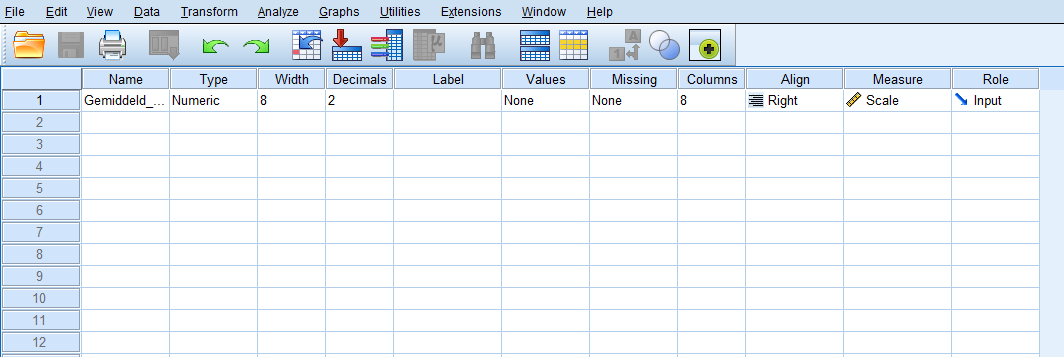
4 Uitvoering in SPSS

Er zijn twee datasets Gemiddeld\_cijfer\_WNS en Gemiddeld\_cijfer\_WNS\_n30 ingeladen met de gemiddelde eindexamencijfers van WNS van eerstejaars Werktuigbouwkunde. Het eerste bestand bevat 124 studenten en het tweede bevat 30 studenten.

4.1 De data bekijken

Klik het tabblad *Data View* aan om de data te bekijken. Klik het tabblad *Variable View* aan om de eigenschappen van alle variabelen te bekijken. Beide ingeladen datasets in dit voorbeeld bevatten slechts 1 variabele: Gemiddeld\_cijfer\_WNS of Gemiddeld\_cijfer\_WNS\_n30. We kunnen onder andere zien dat de variabele in dit geval numeriek is, geen missende waarden bevat en dat deze op schaalniveau is gemeten. Wanneer bepaalde eigenschappen niet blijken te kloppen kunt u deze handmatig aanpassen.



Gebruik head() en tail() om de structuur van de data te bekijken.

*## Eerste 5 observaties*

head(Gemiddeld\_cijfer\_WNS)

*## [1] 7.400000 7.366667 7.366667 7.666667 5.933333 6.500000*

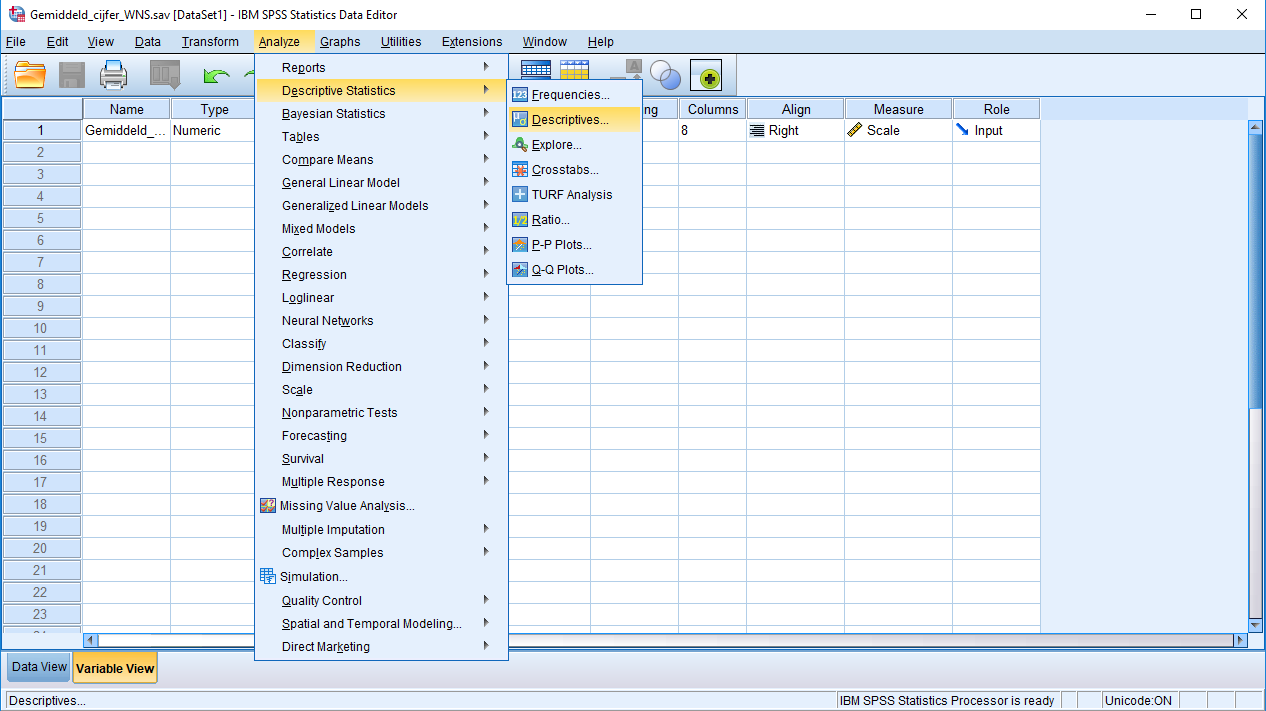
*## Laatste 5 observaties*

tail(Gemiddeld\_cijfer\_WNS)

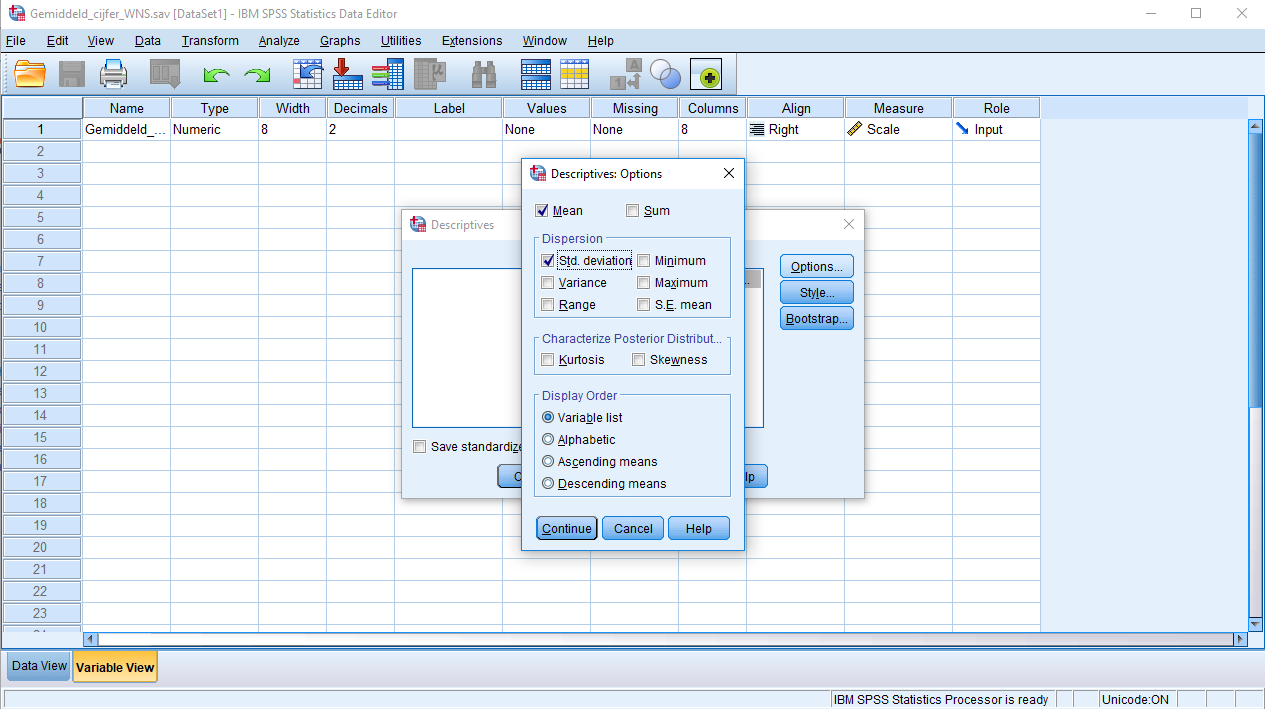
*## [1] 6.333333 6.133333 8.066667 5.166667 7.700000 6.366667*

Krijg meer inzicht in de dataset door het gemiddelde en de spreiding te bekijken. Dit doe je door te klikken op:

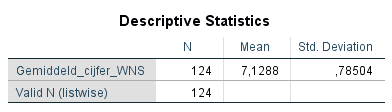
*Analyze* > *Descriptive Statistics* > *Descriptives*



Selecteer de gewenste variabele(n) en plaats deze in het vakje Variable(s), dit doe je middels de  toets. Klik vervolgens op *Options* om de gewenste gegevens te verkrijgen. Vink het gemiddelde (*Mean*) en de standaard deviatie (*Std. deviation*) aan en klik op *Continue*.



Klik vervolgens op *OK* om meteen naar de output te gaan, klik op *Paste* om de code in je Syntax te zetten (zie onder het kopje Syntax hoe je vervolgens te werk gaat).



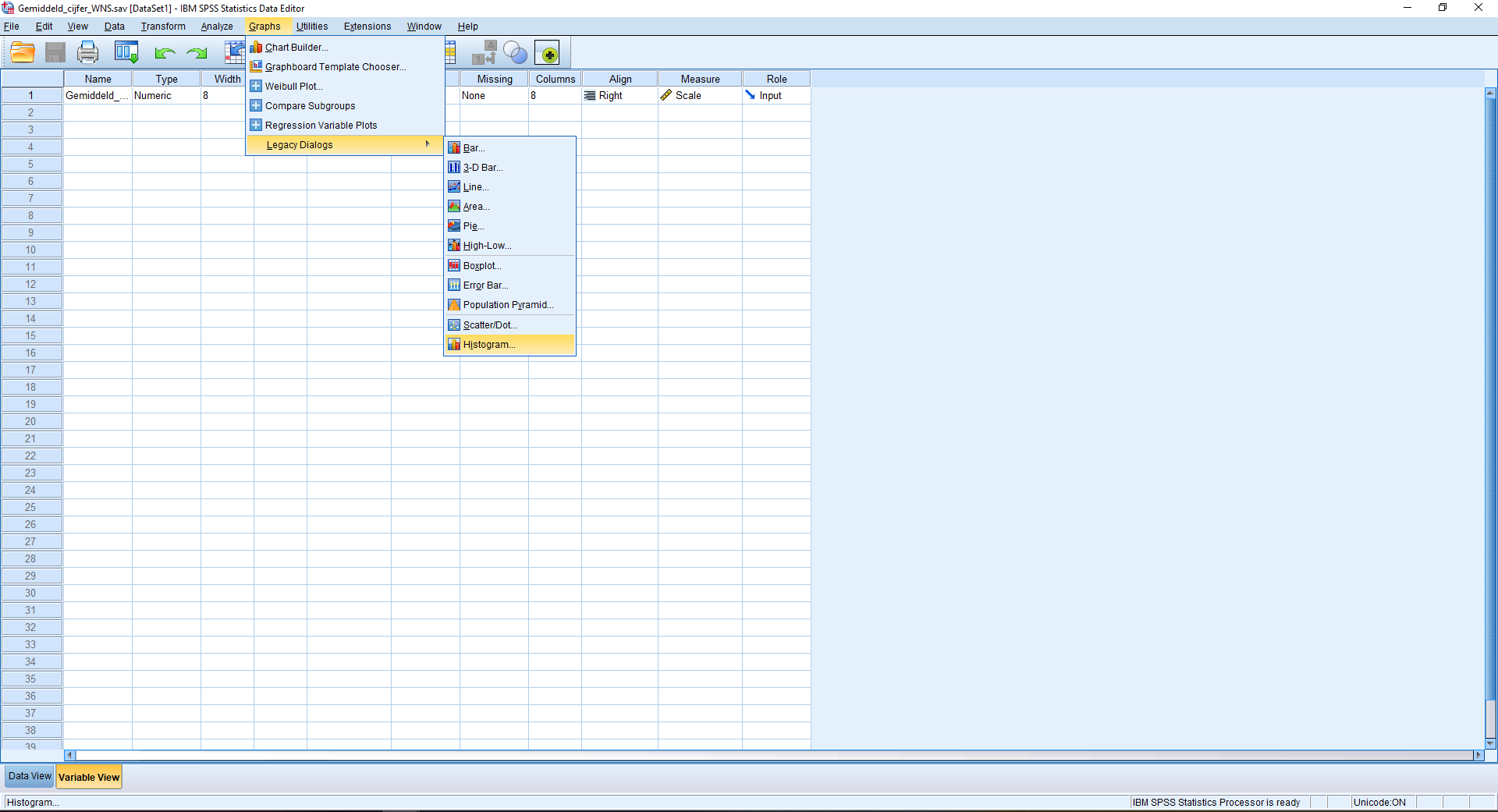
Gemiddeld cijfer WNS (standaardafwijking): 7,13 (0,79). *n* = 124.

4.2 Visuele inspectie van normaliteit

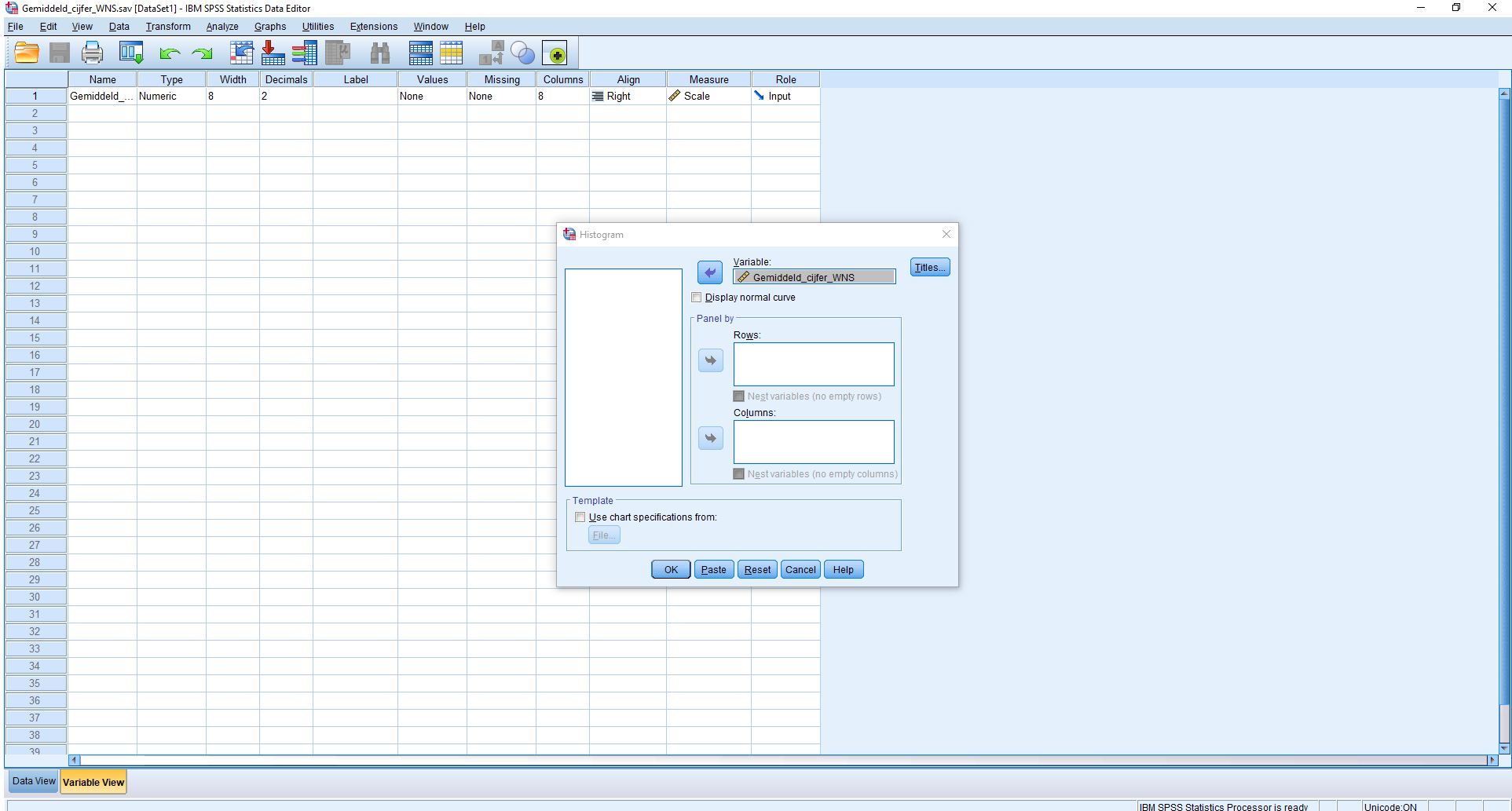
Geef normaliteit visueel weer met een histogram, boxplot of Q-Q plot.

4.2.1 Histogram

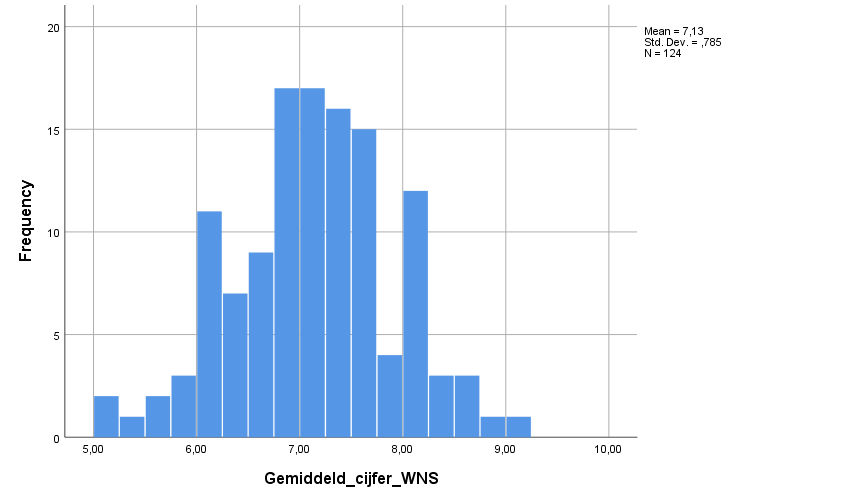
Om een figuur te maken in SPSS klik je op *Graphs* > *Legacy Dialogs* en vervolgens op het gewenste figuur, in dit geval dus *Histogram..*.



Selecteer de gewenste variabele en plaats deze in het vakje *Variable*, dit doe je middels de  toets.



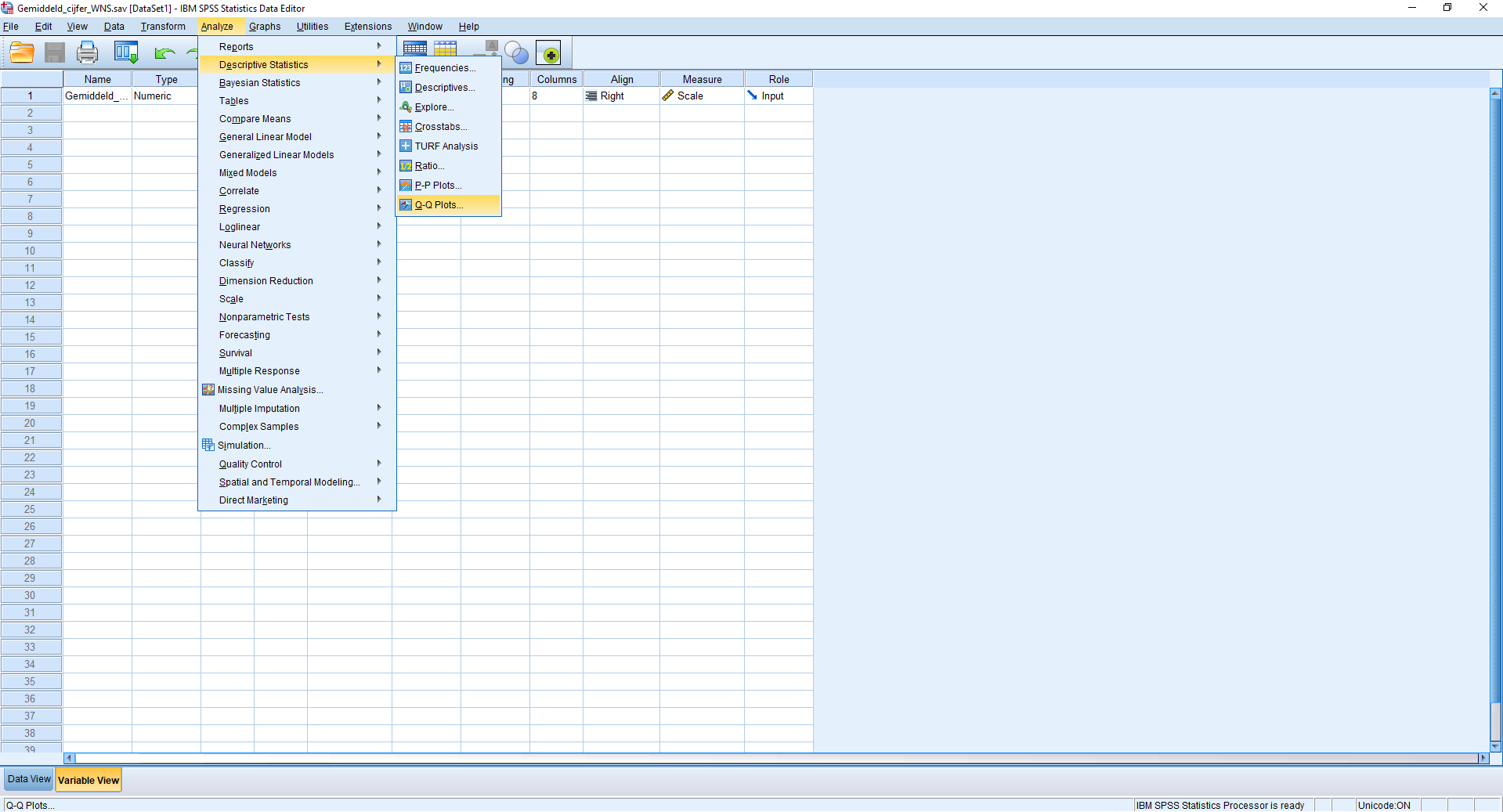
Klik vervolgens op *OK* om meteen naar de output te gaan, klik op *Paste* om de code in je Syntax te zetten.



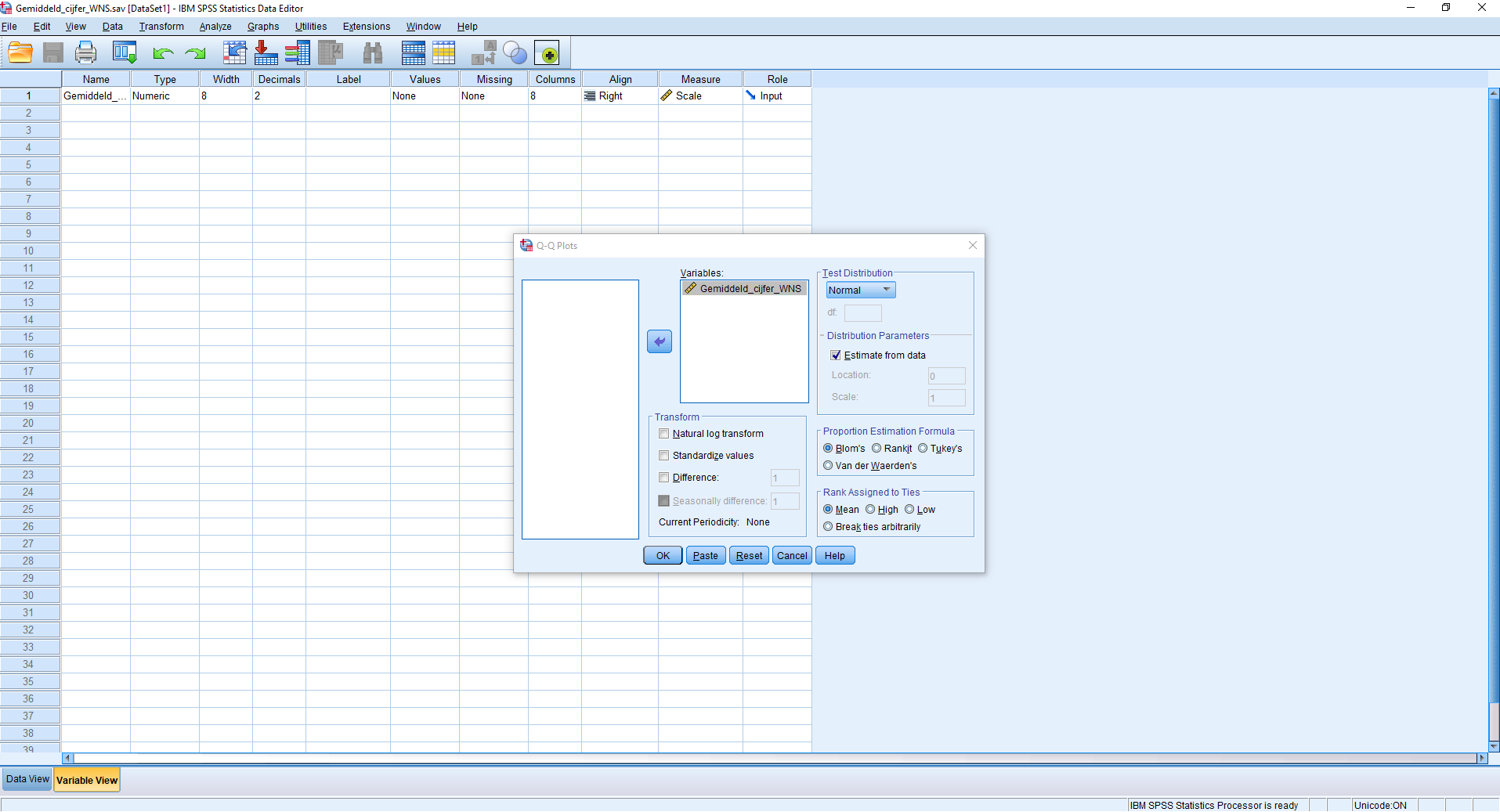
Het histogram laat een Bell Curve zien vergelijkbaar met een normale verdeling: veel waardes liggen rondom het gemiddelde en weinig bij de staarten van de verdeling. Daarnaast ziet de verdeling er in grote mate symmetrisch uit.

4.2.2 Q-Q plot

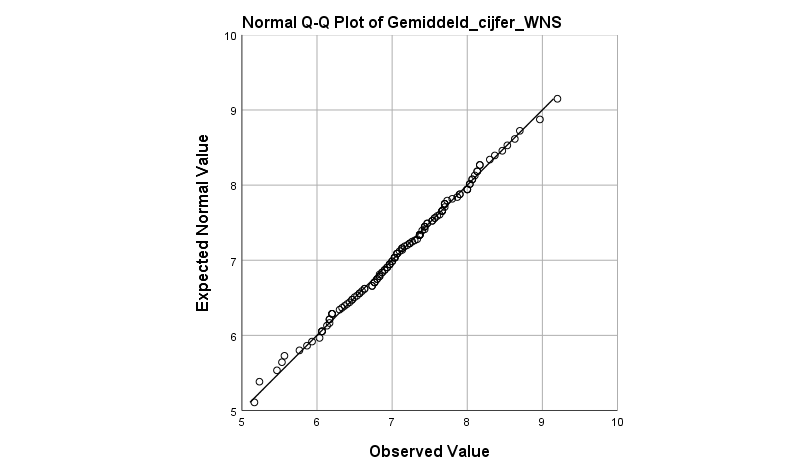
De Q-Q plot is niet te vinden onder *Graphs*. Klik voor de Q-Q plot op *Analyze* > *Desctiptive Statistics* > *Q-Q Plots…*



Selecteer de gewenste variabele(n) en plaats deze in het vakje Variables, dit doe je middels de  toets.



Klik vervolgens op *OK* om meteen naar de output te gaan, klik op *Paste* om de code in je Syntax te zetten. Als over het algemeen de meeste datapunten op de lijn liggen, kan aangenomen worden dat de data normaal verdeeld zijn.

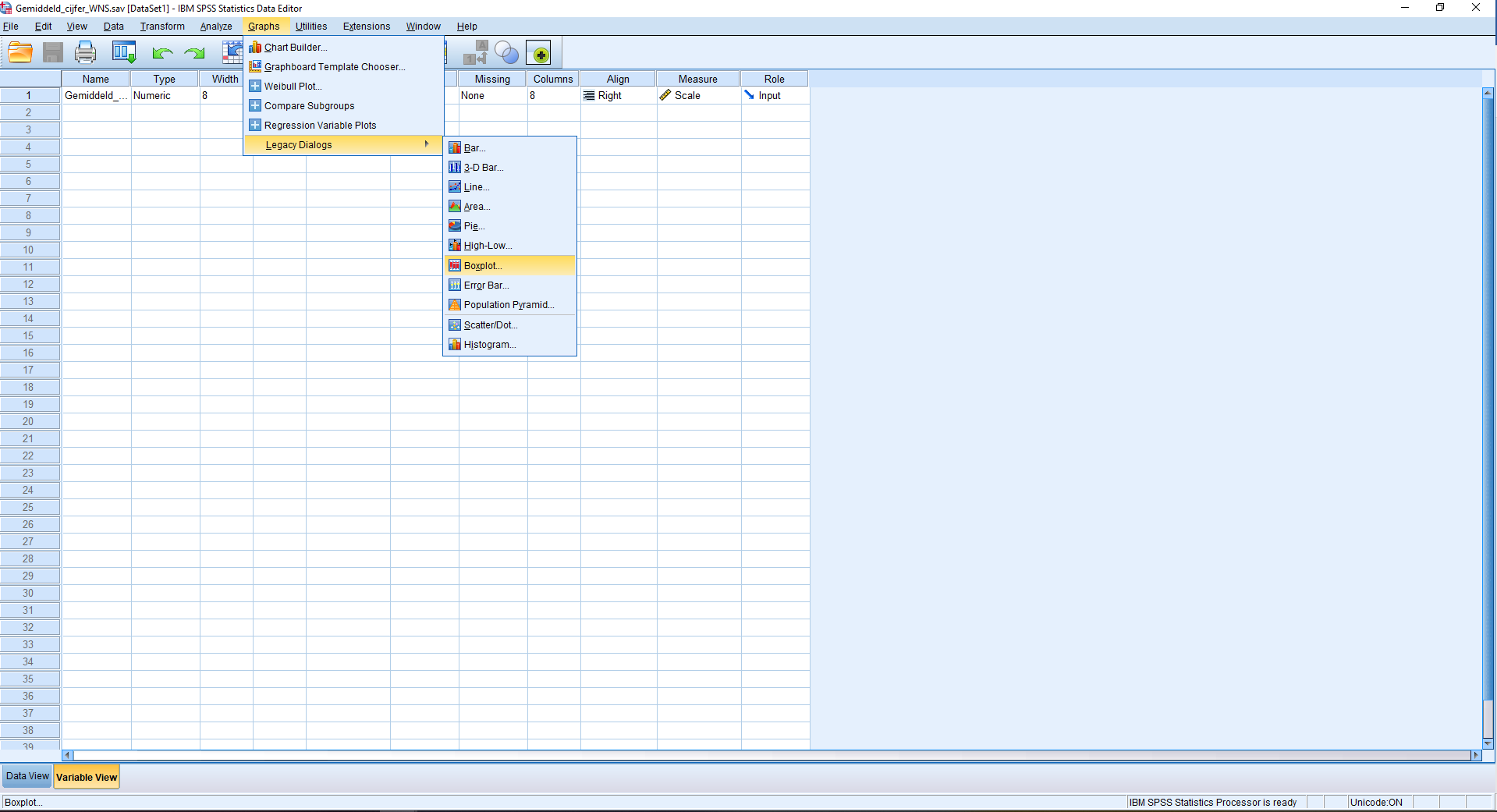


In deze casus liggen de meeste punten op de lijn. Bij de uiteinden liggen de punten dichtbij de lijn.

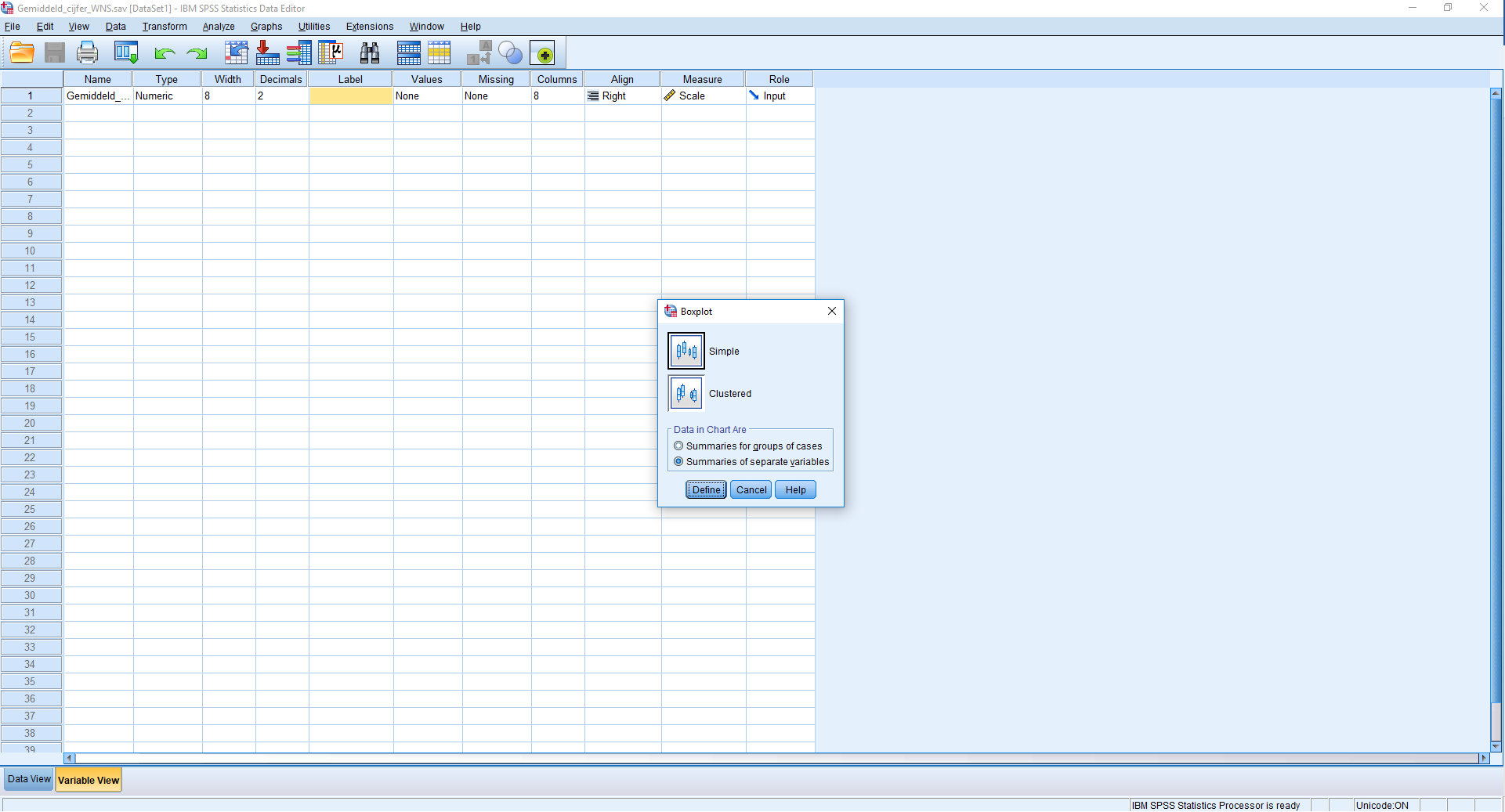
4.2.3 Boxplot

De box geeft de middelste 50% van de tentamencijfers weer. De zwarte lijn binnen de box is de mediaan. In de staarten zitten de eerste 25% en de laatste 25%. Cirkels visualiseren uitbijters.[8](file:///C:\Users\jdk840\AppData\Local\Microsoft\Windows\INetCache\Content.Outlook\OEX47LOI\01-One-sample-t-toets.html#fn8)

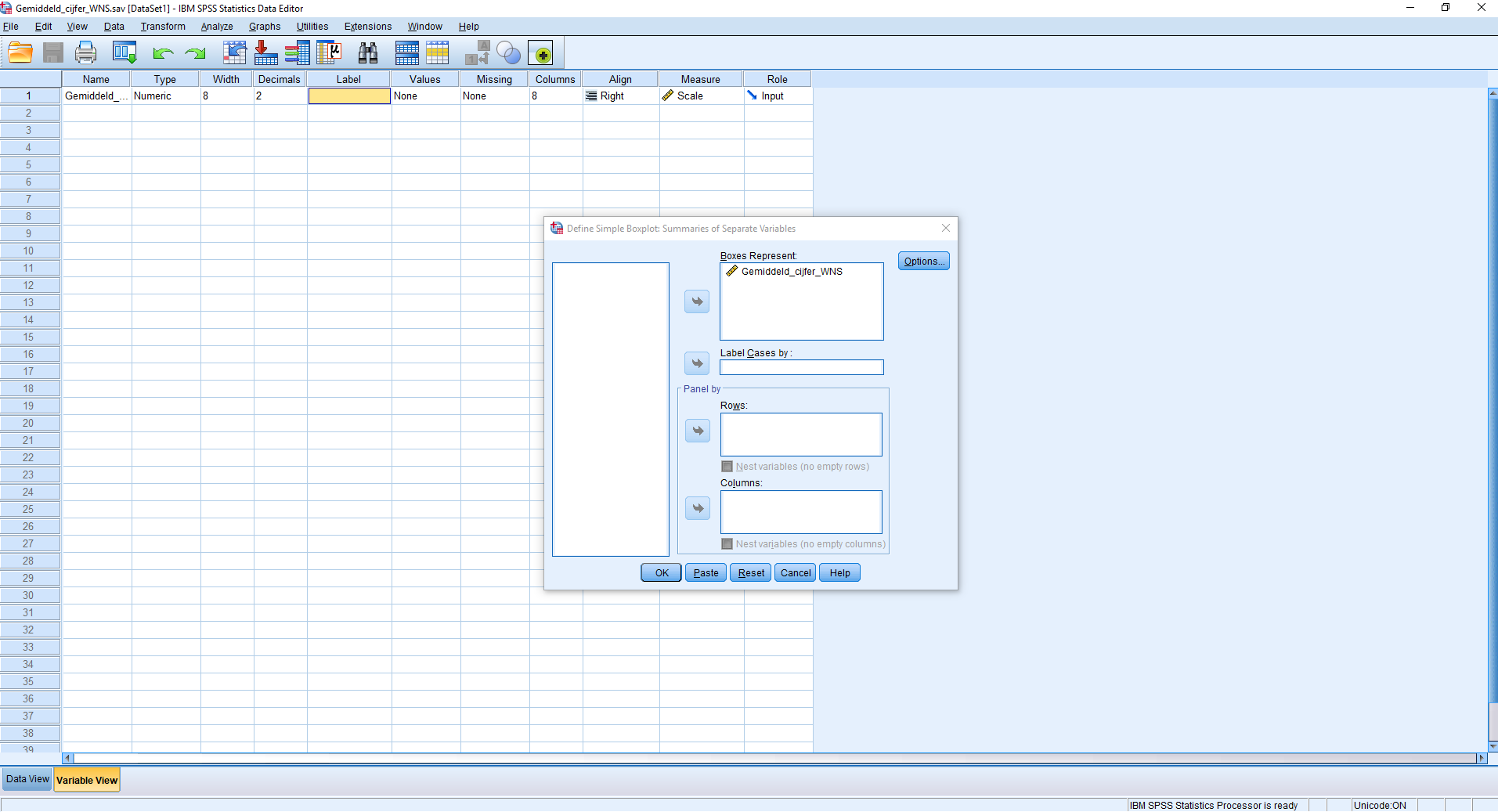
Klik op *Graphs* > *Legacy Dialogs* > *Boxplot…*



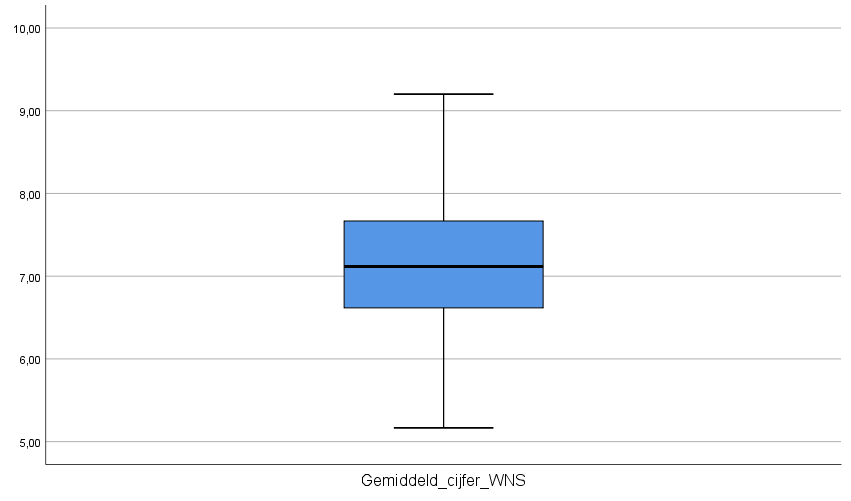
Kies vervolgens voor een simpele variant en aangezien we in dit voorbeeld naar 1 variabele kijken en niet naar groepen, kies je voor *Summaries of separate variables*. Klik dan op *Define*.



Selecteer de gewenste variabele(n) en plaats deze in het vakje Boxes Represent, dit doe je middels de  toets.



Klik vervolgens op *OK* om meteen naar de output te gaan, klik op *Paste* om de code in je Syntax te zetten.



De boxplot geeft de spreiding van het gemiddelde eindexamencijfer voor de exacte vakken weer van de studenten Werktuigbouwkunde. De box en staarten zien er symmetrisch uit, wat een indicatie is van een normale verdeling.

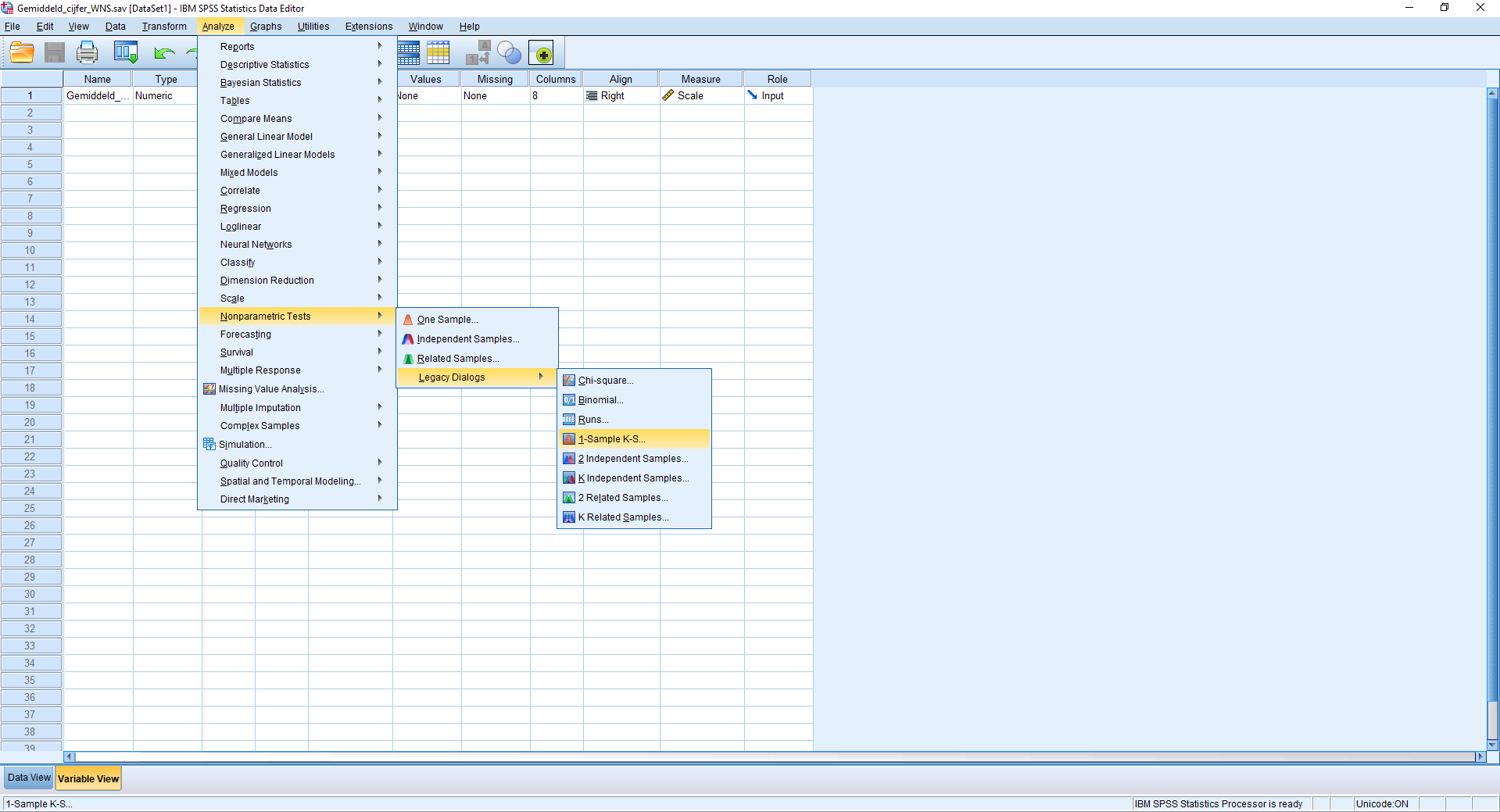
4.3 Toetsen van normaliteit

Om te controleren of de data normaal verdeeld zijn, kan de normaliteit getoetst worden. Hierbij een toelichting bij twee veelgebruikte toetsen: de *Kolmogorov-Smirnov test* en de *Shapiro-Wilk test*.

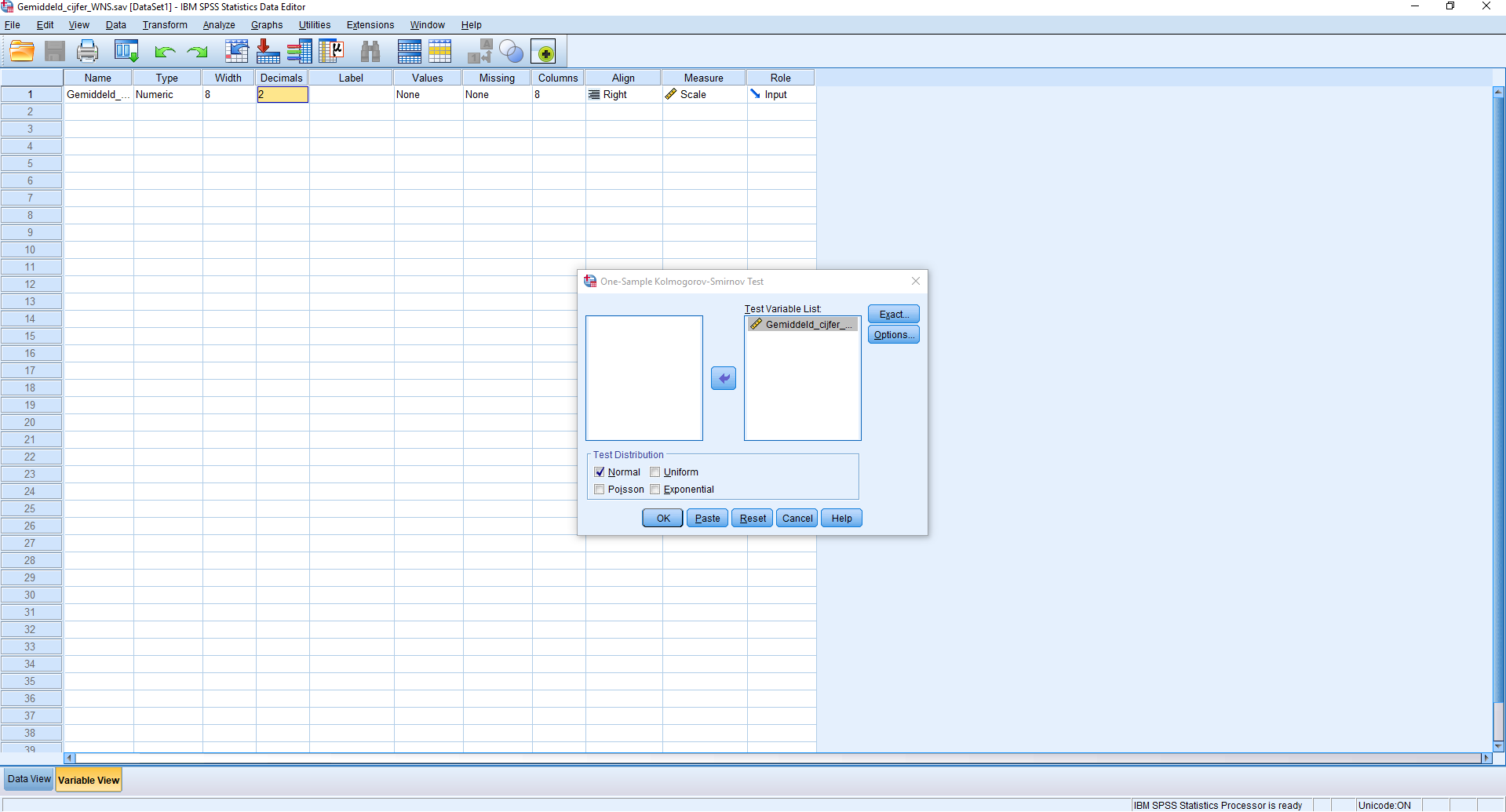
4.3.1 Kolmogorov-Smirnov

De *Kolmogorov-Smirnov test* toetst het verschil in vorm tussen twee verdelingen. Standaard toetst deze test het verschil tussen een normale verdeling en de verdeling van de steekproef. De Lilliefors correctie wordt gebruikt als het gemiddelde niet 0 is en de standaardafwijking niet 1 is. Als de p-waarde < 0,05 is de verdeling van de data statistisch significant verschillend van de normale verdeling.

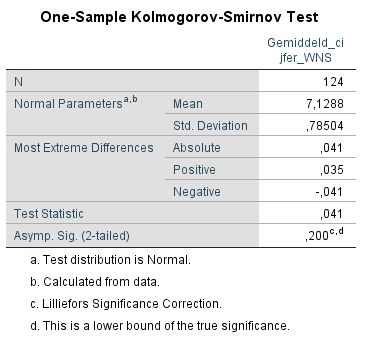
Klik op *Analyze* > *Nonparametric Tests* > *Legacy Dialogs* > *1-Sample K-S…*



Selecteer de gewenste variabele(n) en plaats deze in het vakje Test Variable List, dit doe je middels de  toets. Vink bij Test Distribution het vakje voor Normal aan, omdat we willen testen of de steekproef normaal verdeeld is.



Klik vervolgens op *OK* om meteen naar de output te gaan, klik op *Paste* om de code in je Syntax te zetten.



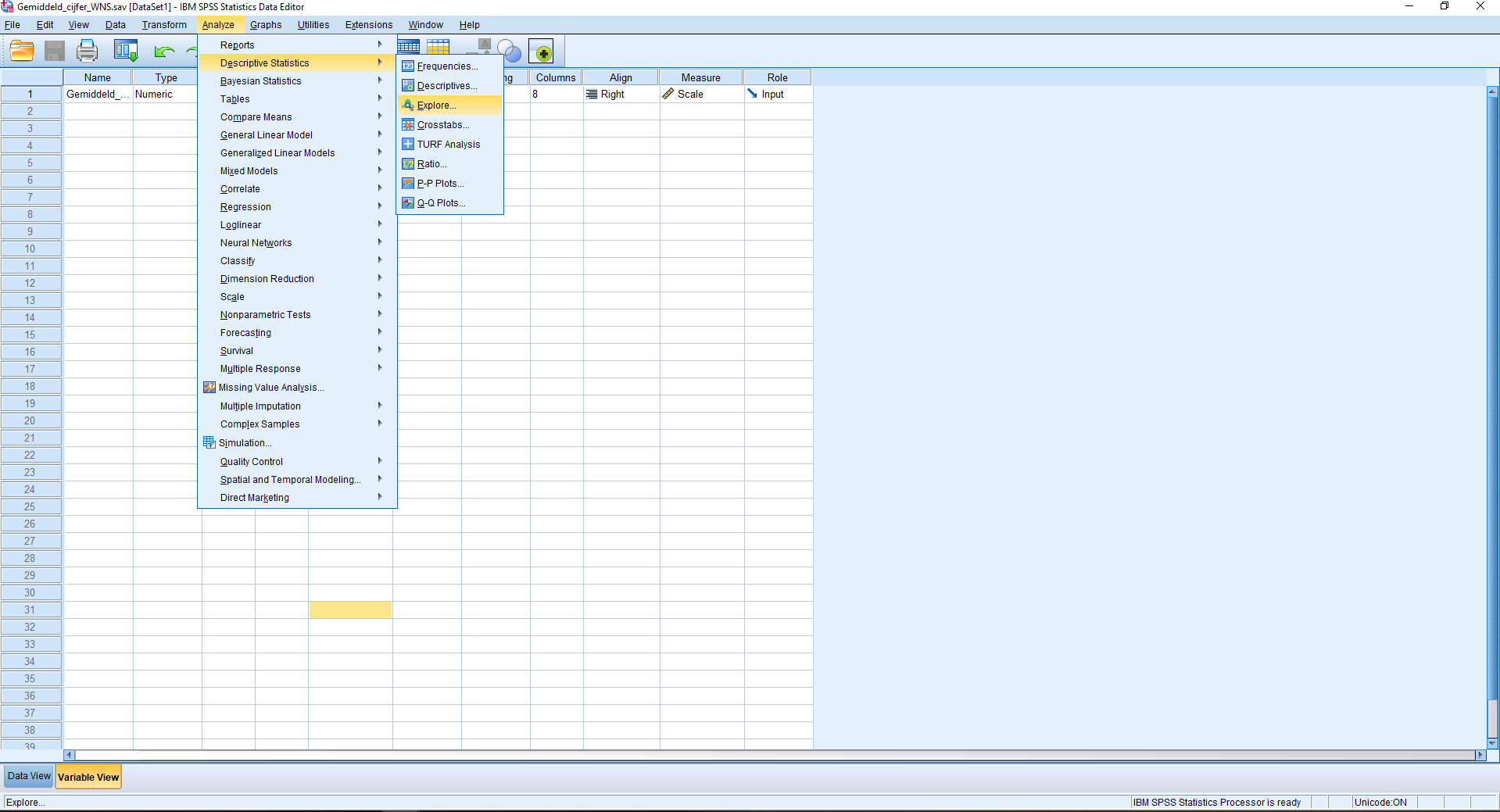
De p-waarde is *p* > 0,05, dus er is geen statistisch significant verschil gevonden tussen de verdeling van de steekproef en de normale verdeling. De *one sample-t-toets* kan uitgevoerd worden.

4.3.2 Shapiro-Wilk Test

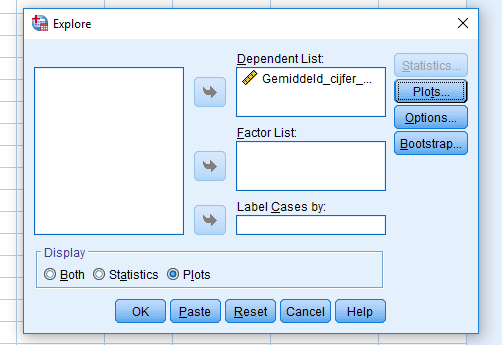
De Shapiro-Wilk test is een soortgelijke test als de Kolmogorov-Smirnov test en wordt vooral gebruikt bij kleine steekproeven (n < 50). Als de p-waarde < 0,05 is de verdeling van de data significant verschillend van de normale verdeling.

De dataset met Gemiddeld\_cijfer\_WNS\_n30 is ingeladen met de gemiddelde eindexamencijfers van WNS van eerstejaars Werktuigbouwkunde, *n* = 30.

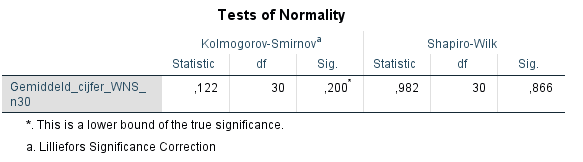
Klik op *Analyze* > *Descriptive Statistics* > *Explore…*



Selecteer de gewenste variabele(n) en plaats deze in het vakje Test Variable List, dit doe je middels de  toets. Vink bij Test Distribution het vakje voor Plots aan.



Klik vervolgens op *OK* om meteen naar de output te gaan, klik op *Paste* om de code in je Syntax te zetten.



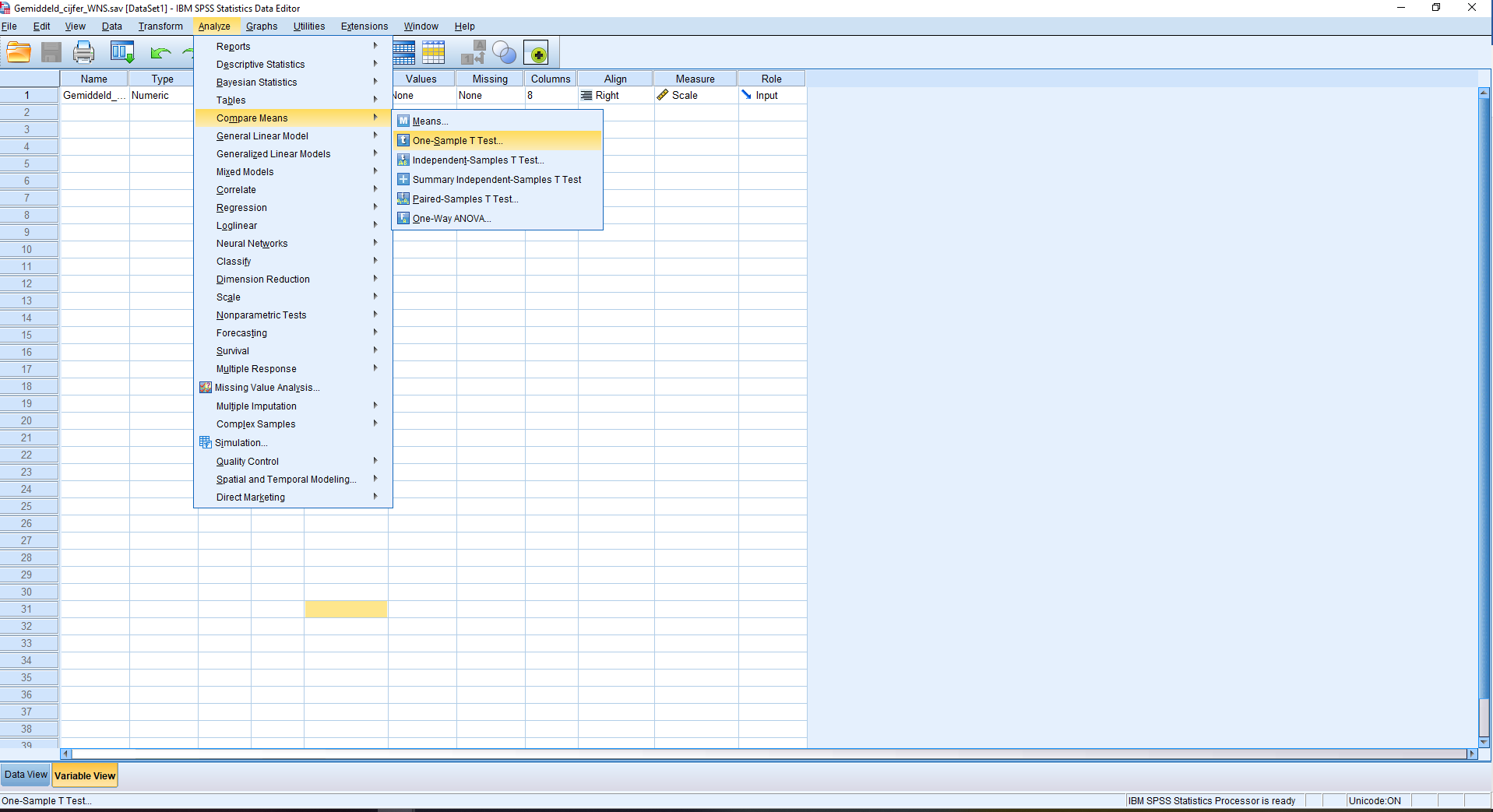
De p-waarde is *p* > 0,05, dus er is geen statistisch significant verschil gevonden tussen de verdeling van de steekproef en de normale verdeling. De *one sample-t-toets* kan uitgevoerd worden.

4.4 One sample t-toets

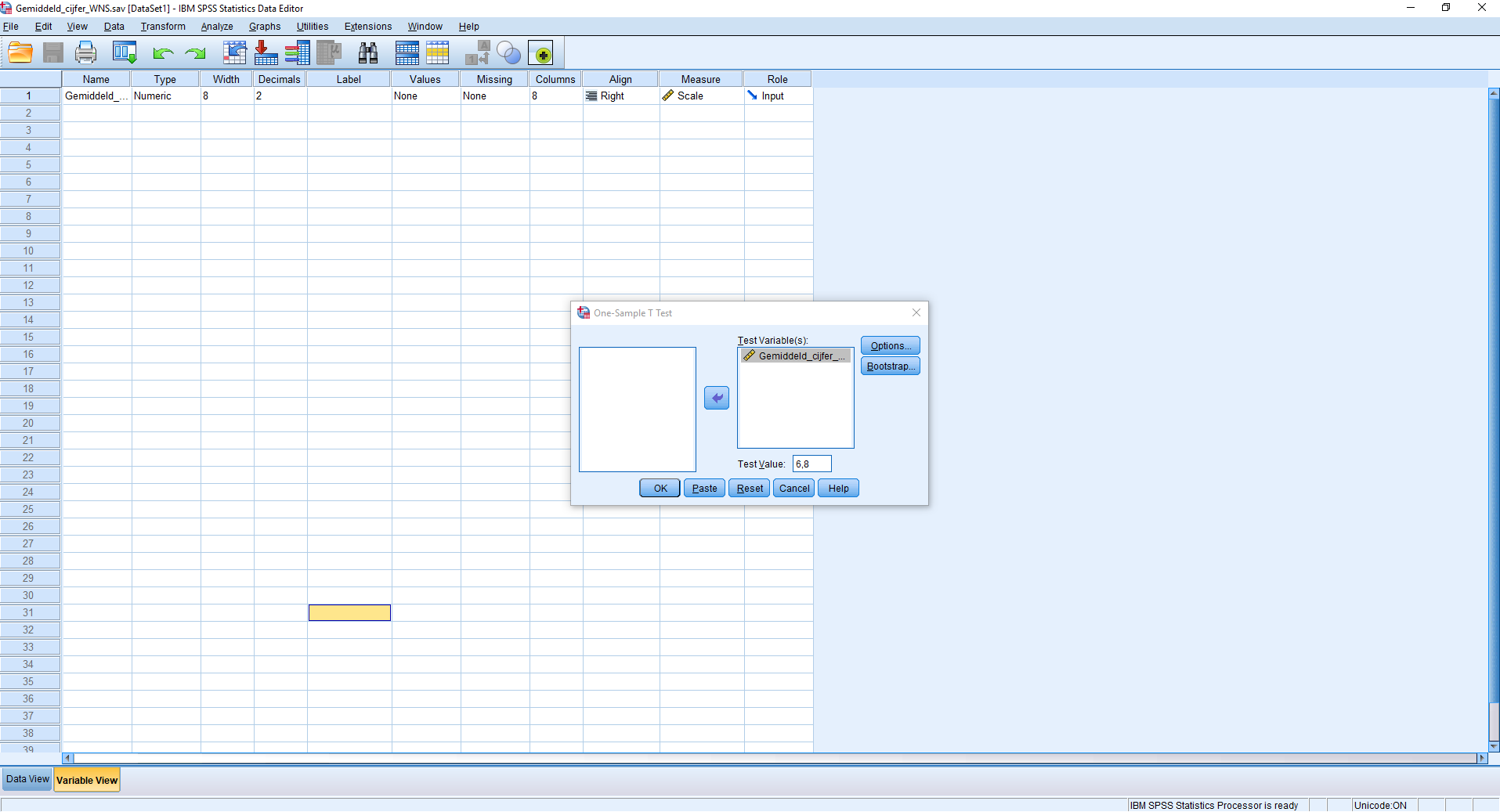
Gebruik t.test()om een t-toets uit te voeren. Geef het gemiddelde waarmee wordt vergeleken en specifieer welke alternatieve hypothese er getoetst wordt.

De dataset met Gemiddeld\_cijfer\_WNS is ingeladen met de gemiddelde eindexamencijfers van WNS van eerstejaars Werktuigbouwkunde, *n* = 124. De *one sample t-toets* is uitgevoerd om de vraag te beantwoorden of het gemiddelde eindexamencijfer voor de exacte vakken van vwo studenten die beginnen aan de Bachelor Werktuigbouwkunde hoger is dan het landelijk gemiddelde (µ = 6,8). De verwachting is dat de studenten hoger scoren, maar omdat het relevant om te weten of de studenten ook lager scoren dan het landelijk gemiddelde is er voor gekozen om tweezijdig te toetsen.

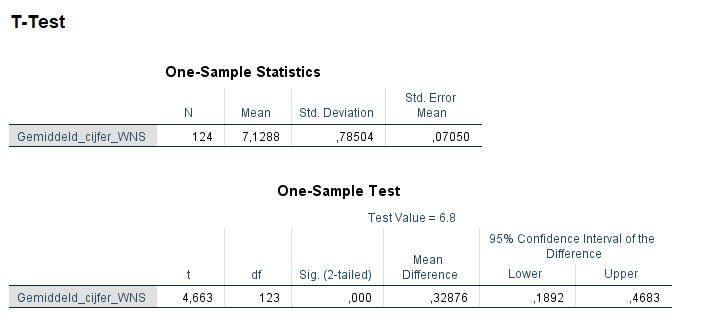
Klik op *Analyze* > *Compare Means* > *One-Sample T Test…*



Selecteer de gewenste variabele(n) en plaats deze in het vakje Test Variable(s), dit doe je middels de  toets. Vul bij Test Value het gemiddelde in waarmee je de data wil vergelijken, in dit geval 6,8. (Let op: gebruik een komma en geen punt).



Klik vervolgens op *OK* om meteen naar de output te gaan, klik op *Paste* om de code in je Syntax te zetten.



De output bestaat uit twee tabellen waarvan de bovenste tabel de statistieken bevat en de onderste tabel de One-Sample Test.

* Vrijheidsgraden, *df* = *n* -1 = 124-1 = 123
* *t* (123) = 4,66, *p* < 0,01
* p-waarde < 0,05, dus de H0 wordt verworpen en de HA wordt aangenomen [9](file:///C:\Users\jdk840\AppData\Local\Microsoft\Windows\INetCache\Content.Outlook\OEX47LOI\01-One-sample-t-toets.html#fn9)
* 95%-betrouwbaarheidsinterval: bij het herhalen van het onderzoek zal in 95% van de gevallen de µ in het interval vallen. In deze casus is het interval tussen 6,99 en 7,27.
* Het gemiddelde van de steekproef is 7,13

Syntax

Het uitvoeren van opdrachten in SPSS kan ook via het SPSS Syntaxbestand. Het gebruik van Syntax kan voordelig zijn als je analyses meermaals wil uitvoeren. Het syntaxbestand kun je namelijk opslaan en zo vaak runnen als je wil zonder de stappen opnieuw uit te hoeven voeren. Je opent een nieuw Syntax bestand door te klikken op *File* > *Open* > *Syntax…*

Een ingeladen dataset moet je vervolgens activeren in de Syntax middels DATA ACTIVATE gevolgd door de naam van de dataset

DATASET ACTIVATE DataSet1.

Wanneer je een geschreven commando wil uitvoeren klik je op de groene ‘play’ knop: .

Als de dataset is geactiveerd kun je opdrachten zelf schrijven of plakken in de Syntax. Wanneer je een opdracht wil plakken voer je dezelfde stappen uit als hierboven beschreven. In plaats van te klikken op OK klik je op Paste. De opdracht wordt overgenomen in je Syntax, hier kun je deze vervolgens runnen door op de  knop te drukken.

Hoe je zelf de bovenstaande opdrachten kunt schrijven in Syntax is hieronder beschreven. Een asterix (\*) geeft commentaar aan.

\*Activeer de dataset

DATASET ACTIVATE DataSet1.

\*Bekijk het gemiddelde en de spreiding

DESCRIPTIVES VARIABLES=Gemiddeld\_cijfer\_WNS

/STATISTICS=MEAN STDDEV.

\*Maak een histogram

GRAPH

/HISTOGRAM=Gemiddeld\_cijfer\_WNS.

\*Maak een Q-Q Plot

PPLOT

/VARIABLES=Gemiddeld\_cijfer\_WNS

/NOLOG

/NOSTANDARDIZE

/TYPE=Q-Q

/FRACTION=BLOM

/TIES=MEAN

/DIST=NORMAL.

\*Maak een Boxplot

EXAMINE VARIABLES=Gemiddeld\_cijfer\_WNS

/COMPARE VARIABLE

/PLOT=BOXPLOT

/STATISTICS=NONE

/NOTOTAL

/MISSING=LISTWISE.

\*Doe de Kolmogorov-Smirnov test

NPAR TESTS

/K-S(NORMAL)=Gemiddeld\_cijfer\_WNS

/MISSING ANALYSIS.

\*Doe de Shapiro-Wilkinson test

…

\*Doe de One-Sample t-test

T-TEST

/TESTVAL=6.8

/MISSING=ANALYSIS

/VARIABLES=Gemiddeld\_cijfer\_WNS

/CRITERIA=CI(.95).

Notes

1. Valid N (listwise) duidt het aantal gevallen aan die *geen* missende waarden bevatten.