```
 GET
  FILE='\\kclad.ds.kcl.ac.
uk\anywhere\UserData\TGStore03\k1759846\My Documents\diabetes - raw data.
sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
```

16. Assess mean BMI. According to the National Health and Nutrition Examination survey, the average BMI in the United States is 26.55 (average of both sexes). Compute a new variable bmi defined as 703×weight/height^2 (a correction factor of 703 is necessary when using pounds and inches), and labelled as 'body mass index'.

```
COMPUTE bmi=703 * weight / (height * height).
EXECUTE.
```

Use a statistical test to assess whether the mean BMI of the subjects in this dataset is significantly different from the general population.

Notes

Output Created		17-NOV-2017 11:34:55
Comments		
Input	Data	\\kclad.ds.kcl.ac.uk\anywhere\UserData\TGStore03\k1759846\My Documents\diabetes - raw data.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	403
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.

Notes

Syntax		T-TEST /TESTVAL=0 /MISSING=ANALYSIS /VARIABLES=bmi /CRITERIA=CI(.95).
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.06

```
T-TEST
 /TESTVAL=26.55
 /MISSING=ANALYSIS
 /VARIABLES=bmi
 /CRITERIA=CI(.95).
```

T-Test

Notes

Output Created		17-NOV-2017 11:36:27
Comments		
Input	Data	\\kclad.ds.kcl.ac.uk\anywhere\UserData\TGStore03\k1759846\My Documents\diabetes - raw data.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	403
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
Syntax		T-TEST /TESTVAL=26.55 /MISSING=ANALYSIS /VARIABLES=bmi /CRITERIA=CI(.95).
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.00

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
bmi	397	28.7846	6.60589	.33154

One-Sample Test

Test Value = 26.55

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
bmi	6.740	396	.000	2.23457	1.5828	2.8864

Explain the output & summarise your conclusions.

The p value here is shown as .000 which means that we should reject the null hypothesis that the sample mean is equal to the hypothesized population mean. There is a significant difference in mean bmi between this sample, and the population. The avergae bmi of this sample is 2.23457 higher than the rest of the population.

Compare BMI between subjects with and without diabetes.

Compare mean BMI between those diagnosed with diabetes and those not diagnosed with diabetes (diabetes is classed as glyhb>7)

Notes

Output Created		17-NOV-2017 11:47:31
Comments		
Input	Data	\\kclad.ds.kcl.ac.uk\anywhere\UserData\TGStore03\k1759846\My Documents\diabetes - raw data.sav
	Active Dataset	DataSet1
	Filter	diabetes = 0 (FILTER)
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	330
Missing Value Handling	Definition of Missing	For each dependent variable in a table, user-defined missing values for the dependent and all grouping variables are treated as missing.
	Cases Used	Cases used for each table have no missing values in any independent variable, and not all dependent variables have missing values.
Syntax		MEANS TABLES=bmi /CELLS=MEAN COUNT STDDEV.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.02

Notes

Output Created		17-NOV-2017 11:51:29
Comments		
Input	Data	\\kclad.ds.kcl.ac.uk\anywhere\UserData\TGStore03\k1759846\My Documents\diabetes - raw data.sav
	Active Dataset	DataSet1
	Filter	diabetes = 1 (FILTER)
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	73
Missing Value Handling	Definition of Missing	For each dependent variable in a table, user-defined missing values for the dependent and all grouping variables are treated as missing.
	Cases Used	Cases used for each table have no missing values in any independent variable, and not all dependent variables have missing values.
Syntax		MEANS TABLES=bmi /CELLS=MEAN COUNT STDDEV.
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.00

Notes

Output Created		17-NOV-2017 11:54:47
Comments		
Input	Data	\\kclad.ds.kcl.ac.uk\anywhere\UserData\TGStore03\k1759846\My Documents\diabetes - raw data.sav
	Active Dataset	DataSet1
	Filter	diabetes = 1 (FILTER)
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	73
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
Syntax		T-TEST GROUPS=diabetes(1.00 0.00) /MISSING=ANALYSIS /VARIABLES=bmi /CRITERIA=CI(.95).
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.01

```
 FILTER OFF.
USE ALL.
EXECUTE.
T-TEST GROUPS=diabetes(1.00 0.00)
  /MISSING=ANALYSIS
  /VARIABLES=bmi
  /CRITERIA=CI(.95).
```

T-Test

Notes

Output Created		17-NOV-2017 11:55:31
Comments		
Input	Data	\\kclad.ds.kcl.ac.uk\anywhere\UserData\TGStore03\k1759846\My Documents\diabetes - raw data.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	403
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
Syntax		T-TEST GROUPS=diabetes(1.00 0.00) /MISSING=ANALYSIS /VARIABLES=bmi /CRITERIA=CI(.95).
Resources	Processor Time	00:00:00.00
	Elapsed Time	00:00:00.00

Group Statistics

	diagnosis of diabetes	N	Mean	Std. Deviation	Std. Error Mean
bmi	yes	71	30.6165	6.31956	.74999
	no	326	28.3856	6.60881	.36603

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
bmi	Equal variances assumed	.313	.576	2.597	395
	Equal variances not assumed			2.673	106.022

Independent Samples Test

		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence ... Lower
bmi	Equal variances assumed	.010	2.23096	.85893	.54231
	Equal variances not assumed	.009	2.23096	.83455	.57640

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the ...	
		Upper	
bmi	Equal variances assumed	3.91962	
	Equal variances not assumed	3.88553	

From this analysis we can see that the p values of the test are given as 0.10 and 0.009. As these values are less than 0.05, we can reject the null hypothesis, which was that "the mean values of BMI are not significantly different for subjects with the diagnosis of diabetes and for subjects without a diagnosis of diabetes". Essentially saying that the BMIs of the two groups ARE significantly different.

18. Investigate relationship between diabetes and hypertension. Hypertension is defined as systolic blood

pressure > 140. Compute a new variable hypertension as first systolic blood pressure (bp.1s) > 140, and assign suitable value labels.

```
 IF  (bp.1s > 140) hypertension=1.00.
EXECUTE.
IF  (bp.1s  <=  140) hypertension=0.00.
EXECUTE.
```

Perform a suitable test to assess the relationship between diabetes and hypertension.

```
CROSSTABS
  /TABLES=diabetes BY hypertension
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI CORR
  /CELLS=COUNT ROW COLUMN TOTAL
  /COUNT ROUND CELL.
```

Crosstabs

Notes

Output Created		17-NOV-2017 12:25:38
Comments		
Input	Data	\\kclad.ds.kcl.ac.uk\anywhere\UserData\TGStore03\k1759846\My Documents\diabetes - raw data.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	403
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.
Syntax		CROSSTABS /TABLES=diabetes BY hypertension /FORMAT=AVALUE TABLES /STATISTICS=CHISQ PHI CORR /CELLS=COUNT ROW COLUMN TOTAL /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.03
	Dimensions Requested	2
	Cells Available	524245

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
diagnosis of diabetes * diagnosis of hypertension	398	98.8%	5	1.2%	403	100.0%

diagnosis of diabetes * diagnosis of hypertension Crosstabulation

			diagnosis of hypertension		
			No	Yes	Total
diagnosis of diabetes	no	Count	232	93	325
		% within diagnosis of diabetes	71.4%	28.6%	100.0%
		% within diagnosis of hypertension	86.9%	71.0%	81.7%
		% of Total	58.3%	23.4%	81.7%
	yes	Count	35	38	73
		% within diagnosis of diabetes	47.9%	52.1%	100.0%
		% within diagnosis of hypertension	13.1%	29.0%	18.3%
		% of Total	8.8%	9.5%	18.3%
Total		Count	267	131	398
		% within diagnosis of diabetes	67.1%	32.9%	100.0%
		% within diagnosis of hypertension	100.0%	100.0%	100.0%
		% of Total	67.1%	32.9%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	14.832[a]	1	.000		
Continuity Correction[b]	13.789	1	.000		
Likelihood Ratio	14.110	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	14.795	1	.000		
N of Valid Cases	398				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 24.03.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]
Nominal by Nominal	Phi	.193		
	Cramer's V	.193		
Interval by Interval	Pearson's R	.193	.053	3.915
Ordinal by Ordinal	Spearman Correlation	.193	.053	3.915
N of Valid Cases		398		

Symmetric Measures

		Approximate Significance
Nominal by Nominal	Phi	.000
	Cramer's V	.000
Interval by Interval	Pearson's R	.000[c]
Ordinal by Ordinal	Spearman Correlation	.000[c]
N of Valid Cases		

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Looking at the Chi squared test results, we can see that the asymptotic significance of the Pearson chi squared test is .000, which tells us that there is a statistically significant association between people in this sample having diabetes, and having hypertension. In this sample, having diabetes is associated with having hypertension.

What percentage of subject with diabetes also have hypertension?

From the cross-table above, we can see that 52.1% of subjects with diabetes also have hypertension.