



Tabulating and Graphics using SPSS

Dr. Md. Atiqul Islam

M.Sc. (SUST, BD), M.Sc. (UHasselt, BE), PhD (RuG, NL)

Professor

Department of Statistics

Jagannath University, Dhaka-1100

E-mail: atique@stat.jnu.ac.bd

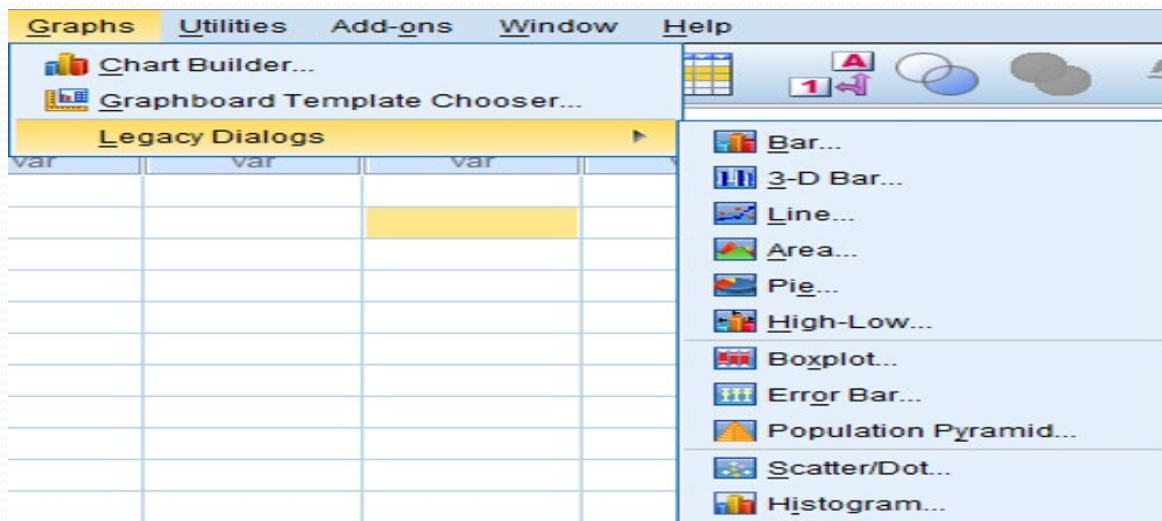
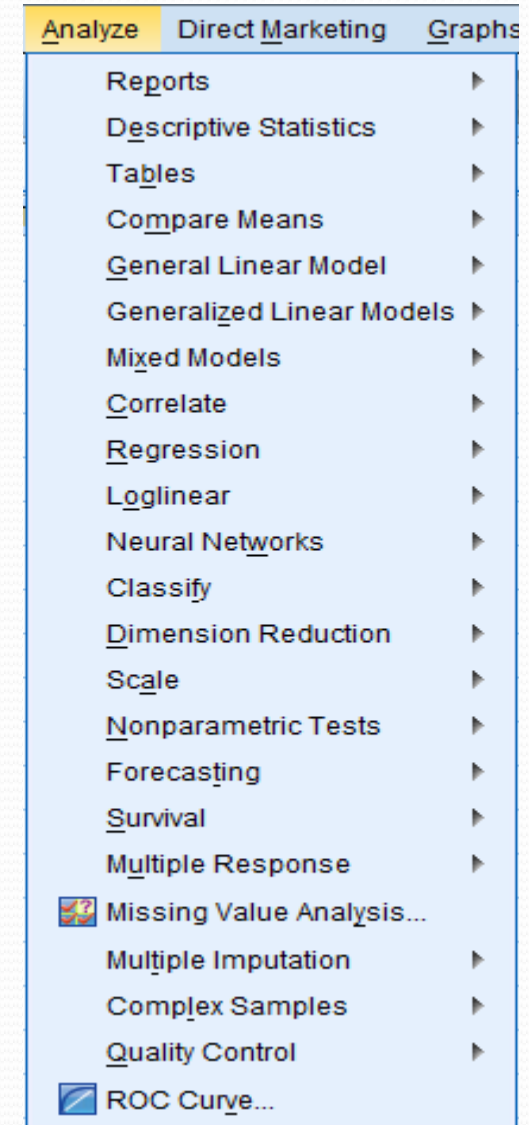
Exploratory Data Analysis



- ▷ Discuss on “Analysis” and “Graph” Menu Bars
- ▷ Data Description
- ▷ Descriptive Statistics for Continuous Data
- ▷ Descriptive Statistics for Categorical Data
- ▷ Graphical Presentation
- ▷ Cross Tabulation

“Analysis” and “Graph” Menu Bar

- ▷ The procedures available in SPSS can be seen in the side figure.
- ▷ Descriptive Statistics using “Descriptive Statistics”
- ▷ Graphical Presentation:
 - ▷ ❶ Bar Diagram ❷ Histogram ❸ Stem & Leaf plot ❹ Box-plot ❺ Pie-chart & Scatter Plot



Data Description and Target



- ▷ The data shows lifespan of two groups of rats
 - ▷ Group 1: A restricted diet (106)
 - ▷ Group 2: Free eating (89)
- ▷ The target of this session
 - ▷ Discuss about the distribution of rats' lifespan
 - ▷ Construct some descriptive tables
 - ▷ Draw some graphs
 - ▷ Compare the distribution of rats' lifespan by diet

Data Description: Sample Data



a) Restricted diet ($n = 105$)

105	193	211	236	302	363	389	390	391	403	530	604	60.5	630	716
718	727	731	749	769	770	789	804	810	811	833	868	871	848	893
897	901	906	907	919	923	931	940	957	958	961	962	974	979	982
1101	1008	1010	1011	1012	1014	1017	1032	1039	1045	1046	1047	1057	1063	1070
1073	1076	1085	1090	1094	1099	1107	1119	1120	1128	1129	1131	1133	1136	1138
1144	1149	1160	1166	1170	1173	1181	1183	1188	1190	1203	1206	1209	1218	1220
1221	1228	1230	1231	1233	1239	1244	1258	1268	1294	1316	1327	1328	1369	1393
1435														

b) *Ad libitum* diet ($n = 89$)

89	104	387	465	479	494	496	514	532	536	545	547	548	582	606
609	619	620	621	630	635	639	648	652	653	654	660	665	667	668
670	675	677	678	678	681	684	688	694	695	697	698	702	704	710
711	712	715	716	717	720	721	730	731	732	733	735	736	738	739
741	743	746	749	751	753	764	765	768	770	773	777	779	780	788
791	794	796	799	801	806	807	815	836	838	850	859	894	963	

Source: Berger, Boss, and Guess, 1988. With permission of the Biometrics Society.

Descriptive Statistics of Rats' Lifespan



▷ Measures of Central Tendency

▷ Mean

▷ Median

▷ Mode

▷ Measures of Location

▷ Quantiles

▷ Measures of Dispersion

▷ Absolute Measures

▷ Range

▷ Standard Deviation

▷ Relative Measures

▷ Coefficient of Variation

▷ Shape Characteristics: Skewness & Kurtosis

Descriptive Statistics of a continuous variable



How can we obtain descriptive statistics of a continuous variable?

Statistics	All	Restricted Diet	Free Diet
Minimum	60.50	60.50	89.00
Maximum	1435.00	1435.00	963.00
Range	1374.00	1374.50	874.00
Mean	836.37	964.30	684.01
SD	274.32	296.27	134.08
Skewness	-0.26	-1.24	-2.01
Kurtosis	0.05	1.20	7.03
IQR	385.00	326.50	121.00
Median	791.00	1042.00	710.00

Descriptive Statistics



Step 1

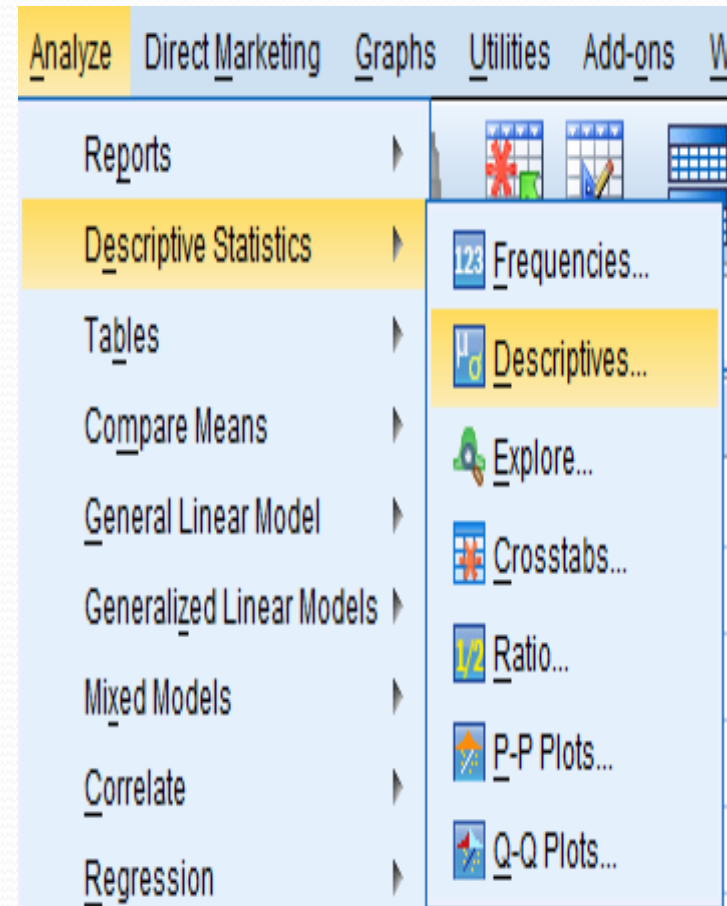
The option “Descriptive Statistics” is used to find the descriptive statistics only for continuous variables.

Manually:

Analysis> Descriptive statistics
> Descriptives > Select Variable> Click
on Statistics Button> Choose the
options> Continue> OK.

Syntax:

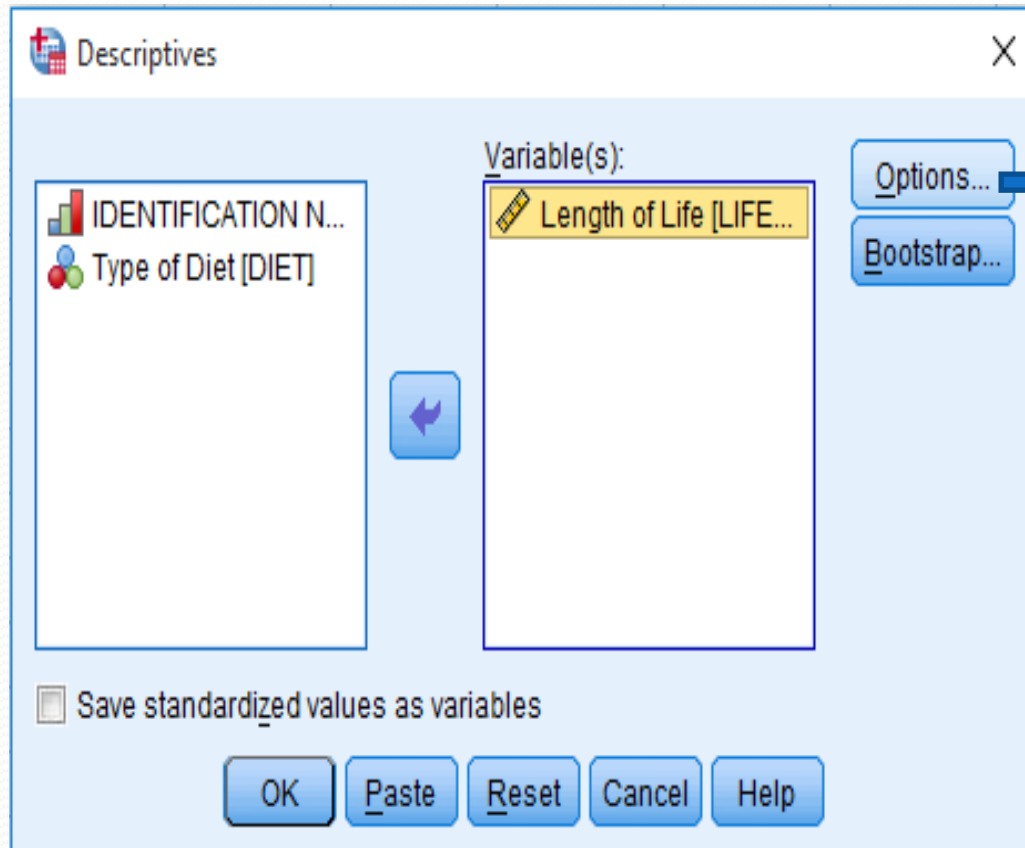
*DESCRIPTIVES VARIABLES=LIFESPAN
/STATISTICS=MEAN STDDEV
VARIANCE RANGE MIN MAX SEMEAN
KURTOSIS SKEWNESS.*



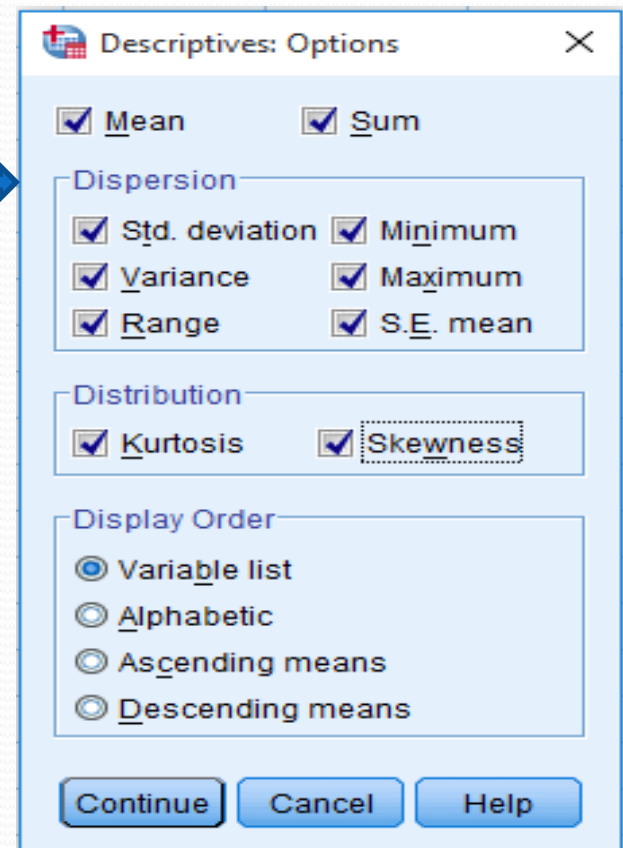
Descriptive Statistics



Step 2



Step 3



Descriptive Statistics



Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Length of Life	195	1374.50	60.50	1435.00	836.3718	19.64465	274.32259	75252.883	-.260	.174	.046	.346
Valid N (listwise)	195											

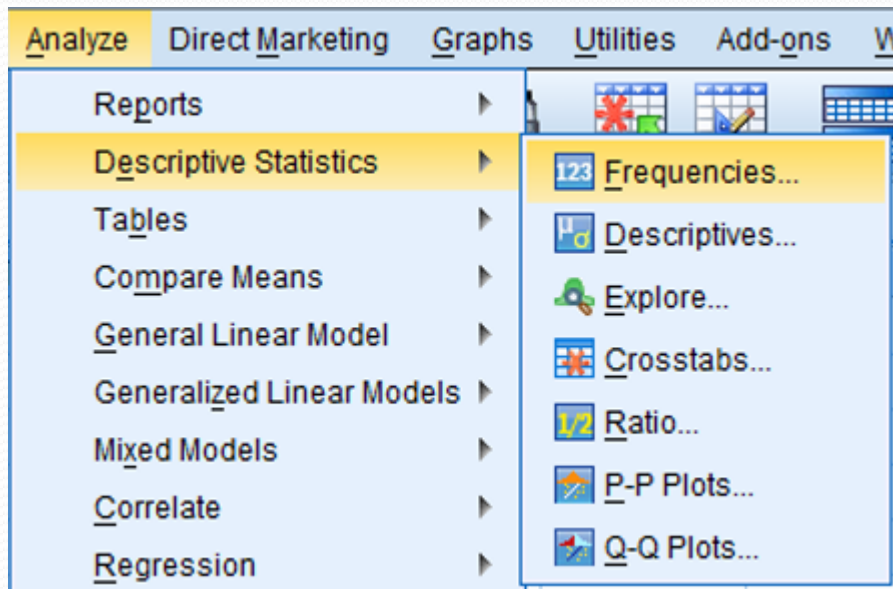
Sampling distribution of mean. Its standard deviation is standard error

► How can we calculate Median, Mode, and IQR? What about quantiles such as quartiles and percentiles?

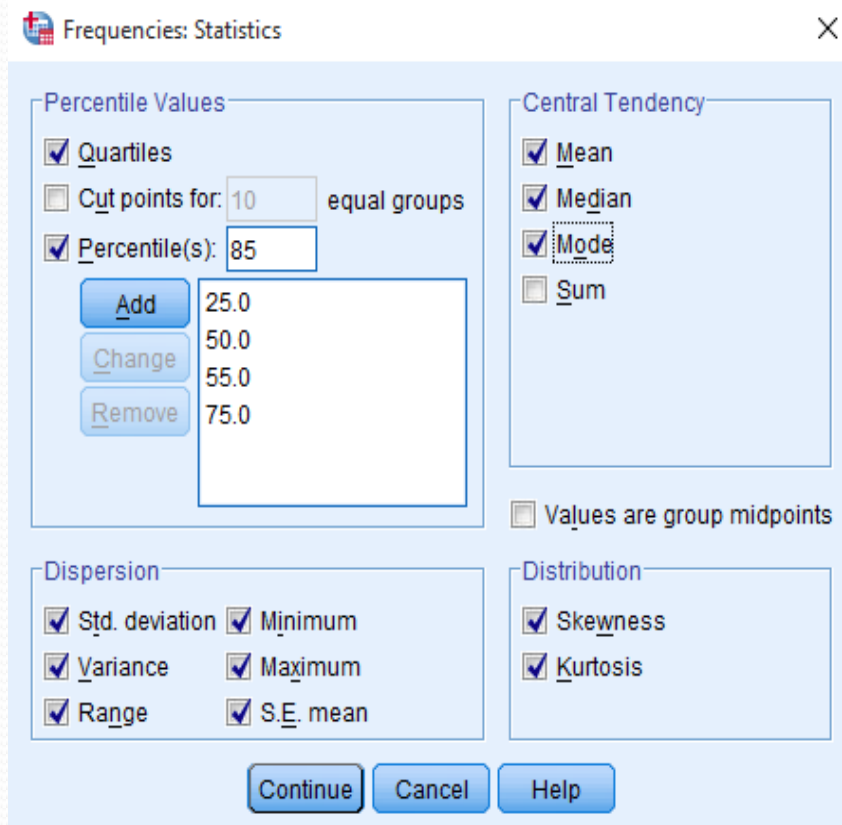
Descriptive Statistics: Alternative Way



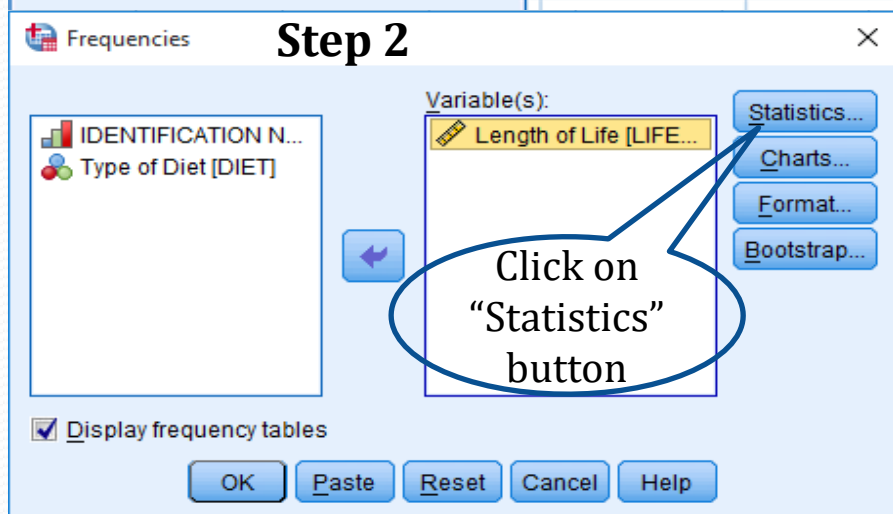
Step 1



Step 3



Step 2



Descriptive Statistics of a continuous variable



Statistics		
Length of Life		
N	Valid	195
	Missing	0
Mean		836.3718
Std. Error of Mean		19.64465
Median		791.0000
Mode		630.00 ^a
Std. Deviation		274.32259
Variance		75252.883
Skewness		-.260
Std. Error of Skewness		.174
Kurtosis		.046
Std. Error of Kurtosis		.346
Range		1374.50
Minimum		60.50
Maximum		1435.00
Percentiles	25	678.0000
	50	791.0000
	55	814.2000
	75	1063.0000
	89	1195.7200

Acceptable

Acceptable

Mean lifetime of the data is 836.37 days with standard deviation 274.32 days

Median and mode of lifespan are 791 and 630 days

678 days is the 25th percentile indicating 25% observations are lying under 678 days.

a. Multiple modes exist. The smallest value is shown

Descriptive Statistics of a continuous variable



Values should be less than ± 1.0 to be considered normal.

Statistics

Length of Life		
N	Valid	195
	Missing	0
Mean		836.3718
Std. Error of Mean		19.64465
Median		791.0000
Mode		630.00 ^a
Std. Deviation		274.32259
Variance		75252.883
Skewness		-.260
Std. Error of Skewness		.174
Kurtosis		.046
Std. Error of Kurtosis		.346
Range		1374.50
Minimum		60.50
Maximum		1435.00
Percentiles	25	678.0000
	50	791.0000
	55	814.2000
	75	1063.0000
	89	1195.7200

Values should be less than ± 1.0 to be considered normal.

a. Multiple modes exist. The smallest value is shown

Descriptive Statistics of LIFESPAN by Group



Use “**Explore**” command **under** “Descriptive Statistics”

- ▷ Need one or more continuous variable
- ▷ Need one categorical variable (Factor) to insert in Factor List

Manually:

Analysis> Descriptive statistics >Explore> Select Dependent Variable> Select Categorical Variable in Factor List>Click on Statistics & Plot Buttons> Choose the options> Continue>Ok

Syntax:

EXAMINE VARIABLES=LIFESPAN BY DIET

/PLOT BOXPLOT STEMLEAF HISTOGRAM NPLOT

/COMPARE GROUP

/PERCENTILES(5,10,25,50,75,90,95) HAVERAGE

/STATISTICS DESCRIPTIVES EXTREME

/CINTERVAL 95

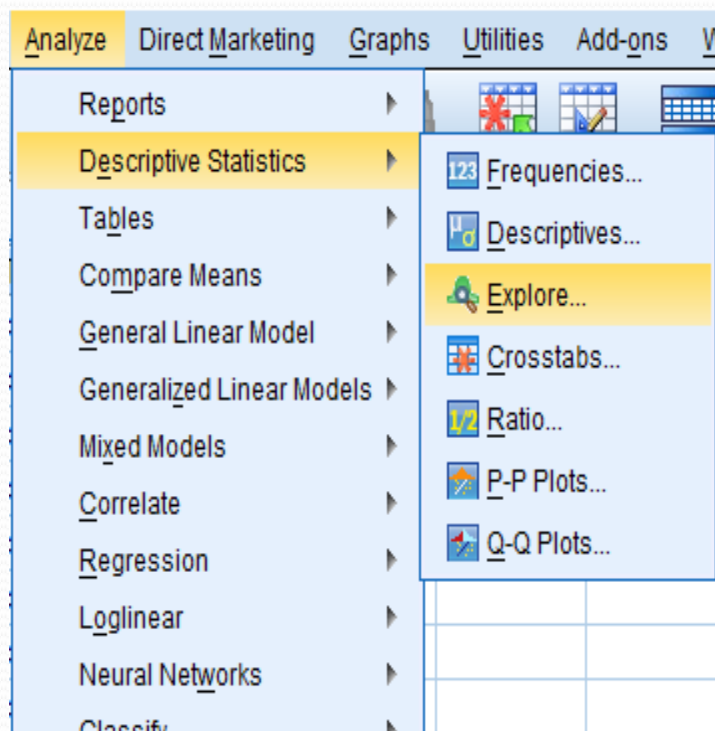
/MISSING LISTWISE

/NOTOTAL.

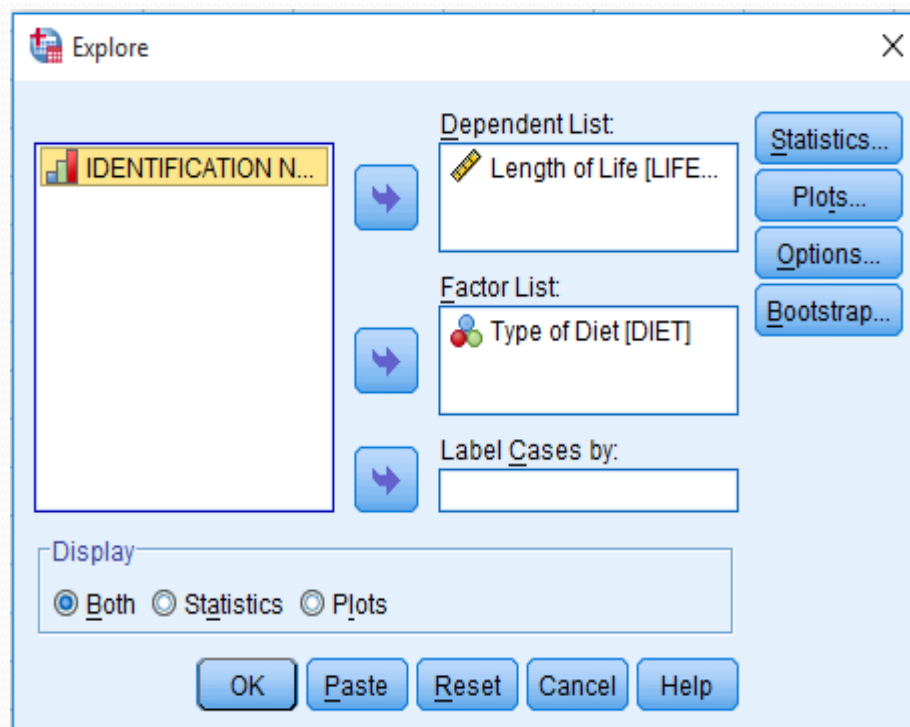
Descriptive Statistics by Group



Step 1



Step 2

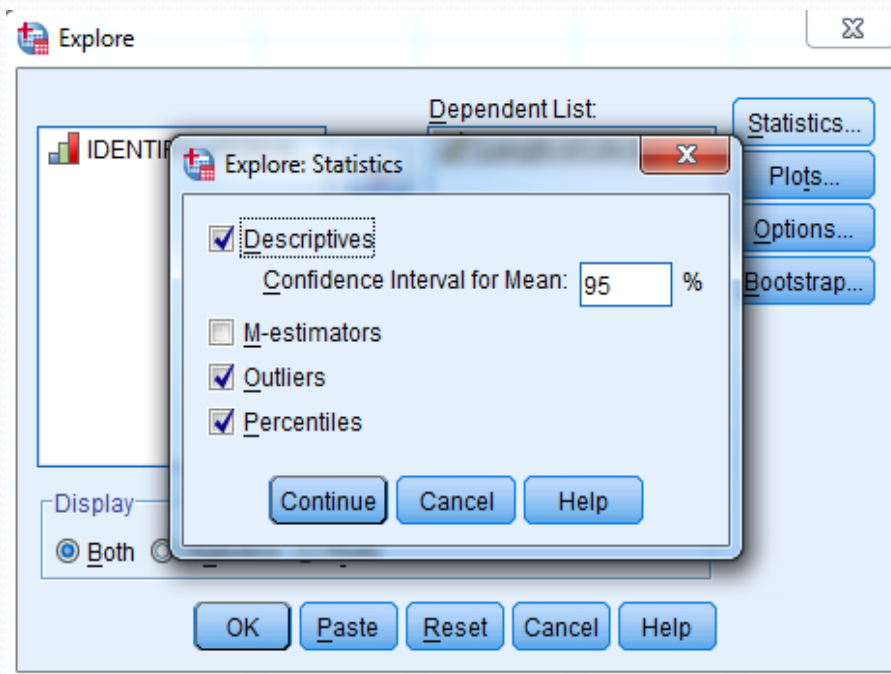


Descriptive Statistics by Group

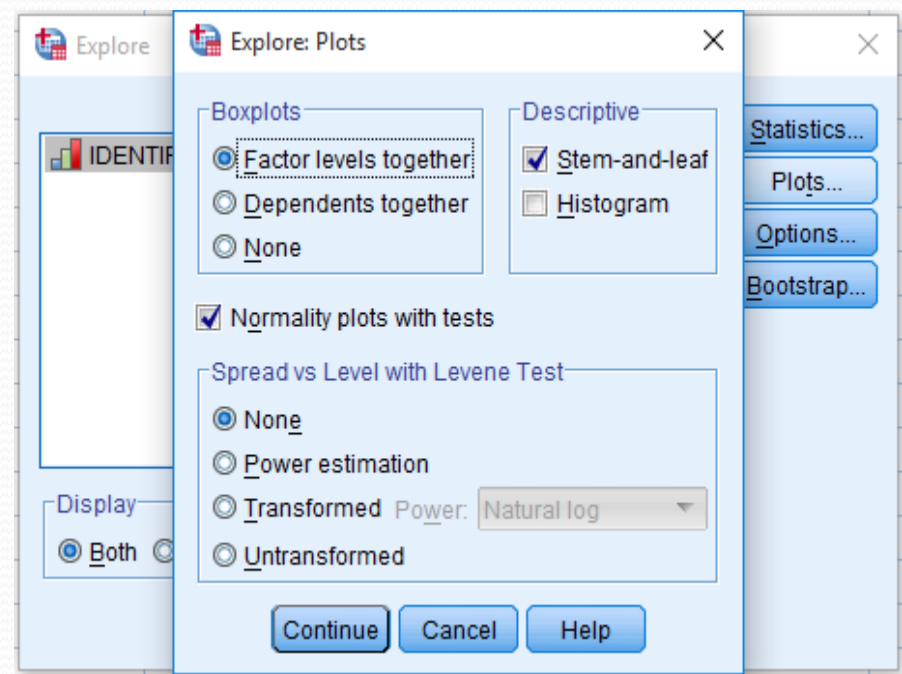


Click on Statistics (for step 3) and Plots (for step 4)

Step 3



Step 4



Case Processing Summary

		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Length of Life	Restricted Diet	106	100.0%	0	.0%	106	100.0%
	Free Diet	89	100.0%	0	.0%	89	100.0%

Percentiles

			Percentiles						
			5	10	25	50	75	90	95
Weighted Average (Definition 1)	Length of Life	Restricted Diet	259.1000	399.4000	844.2500	1042.0000	1170.7500	1240.5000	1323.1500
		Free Diet	472.0000	532.0000	643.5000	710.0000	764.5000	806.0000	844.0000
Tukey's Hinges	Length of Life	Restricted Diet			848.0000	1042.0000	1170.0000		
		Free Diet			648.0000	710.0000	764.0000		

Output



Descriptives

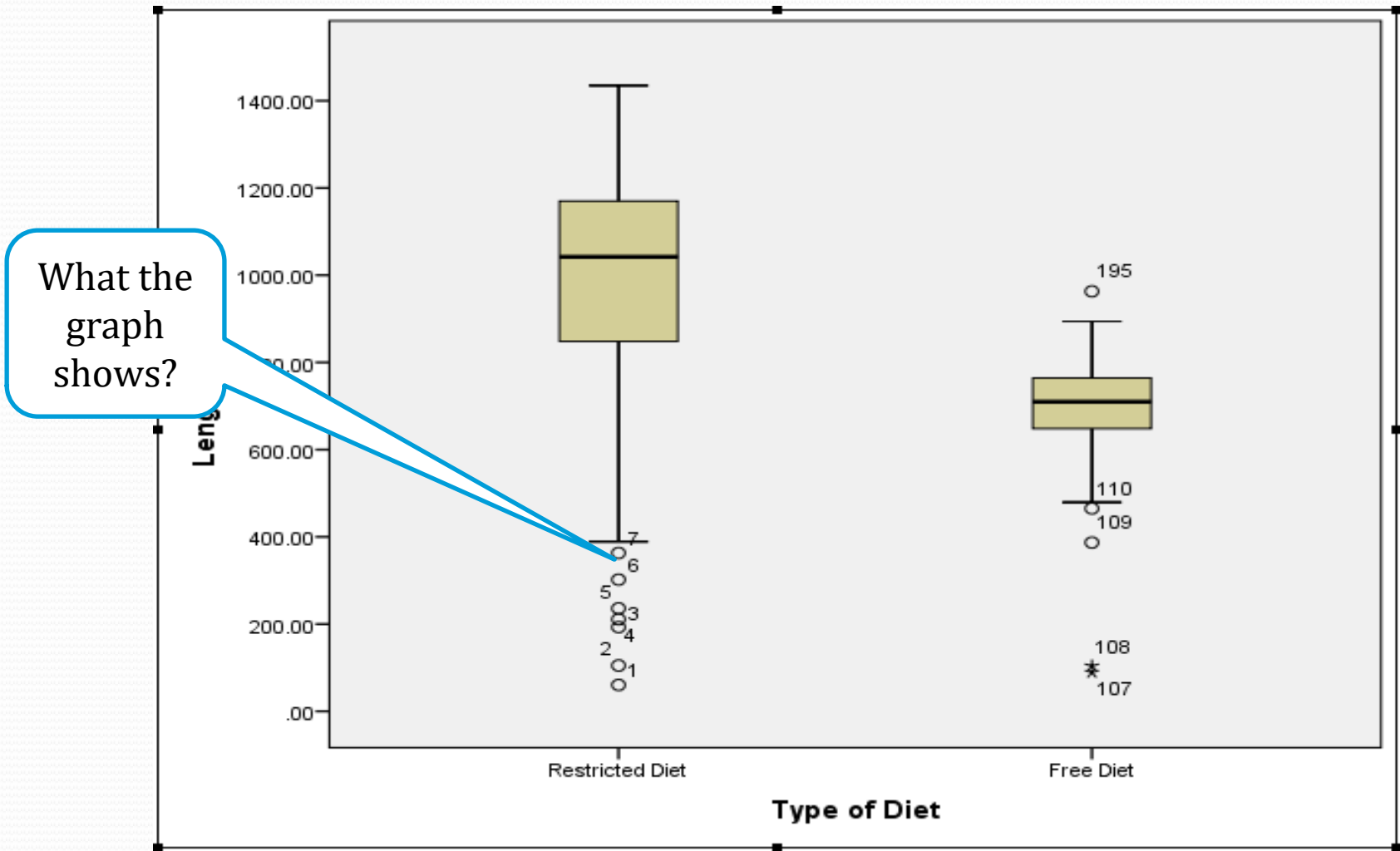
Type of Diet				Statistic	Std. Error
Length of Life	Restricted Diet	Mean		964.2972	28.77593
		95% Confidence Interval for Mean	Lower Bound	907.2398	
			Upper Bound	1021.3545	
		5% Trimmed Mean		986.0860	
		Median		1042.0000	
		Variance		87773.732	
		Std. Deviation		296.26632	
		Minimum		60.50	
		Maximum		1435.00	
		Range		1374.50	
		Interquartile Range		326.50	
		Skewness		-1.244	.235
		Kurtosis		1.199	.465
	Free Diet	Mean		684.0112	14.21290
		95% Confidence Interval for Mean	Lower Bound	655.7661	
			Upper Bound	712.2564	
		5% Trimmed Mean		695.0481	
		Median		710.0000	
		Variance		17978.579	
		Std. Deviation		134.08423	
		Minimum		89.00	
		Maximum		963.00	
		Range		874.00	
		Interquartile Range		121.00	
		Skewness		-2.010	.255
		Kurtosis		7.027	.506

5% trimmed mean indicates the mean of the observations by excluding lower and upper 5% of the observations

If *skewness* > +1, the distribution is right skewed.
If *skewness* < -1, the distribution is left skewed.

If *kurtosis* > +1, the distribution is leptokurtic.
If *kurtosis* < -1, the distribution is Platykurtic.

Box-plot



Stem and Leaf Plot



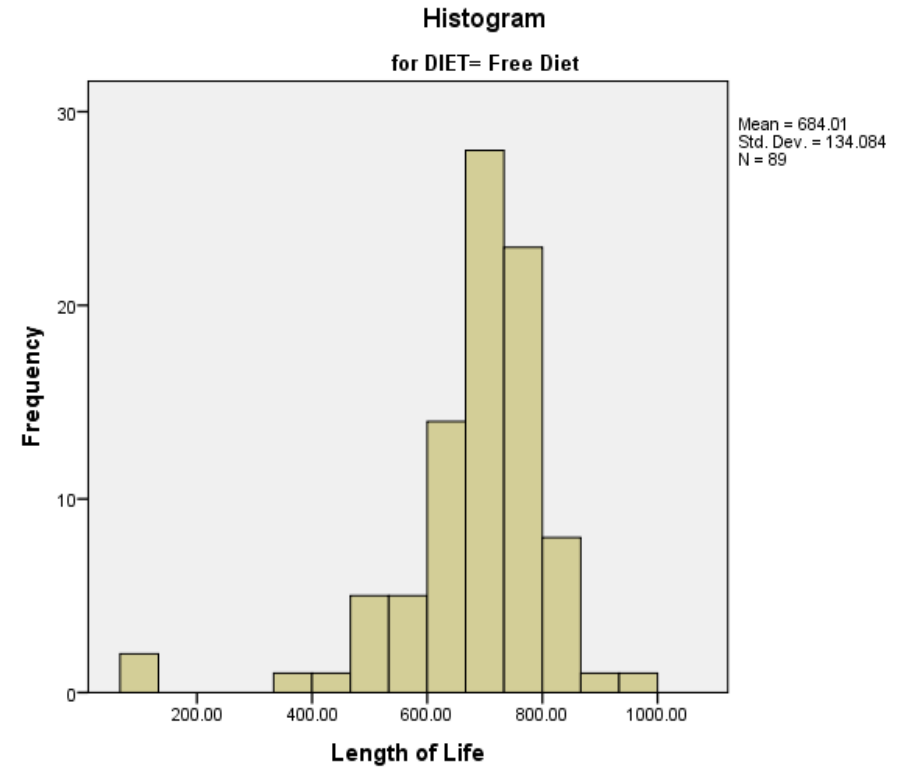
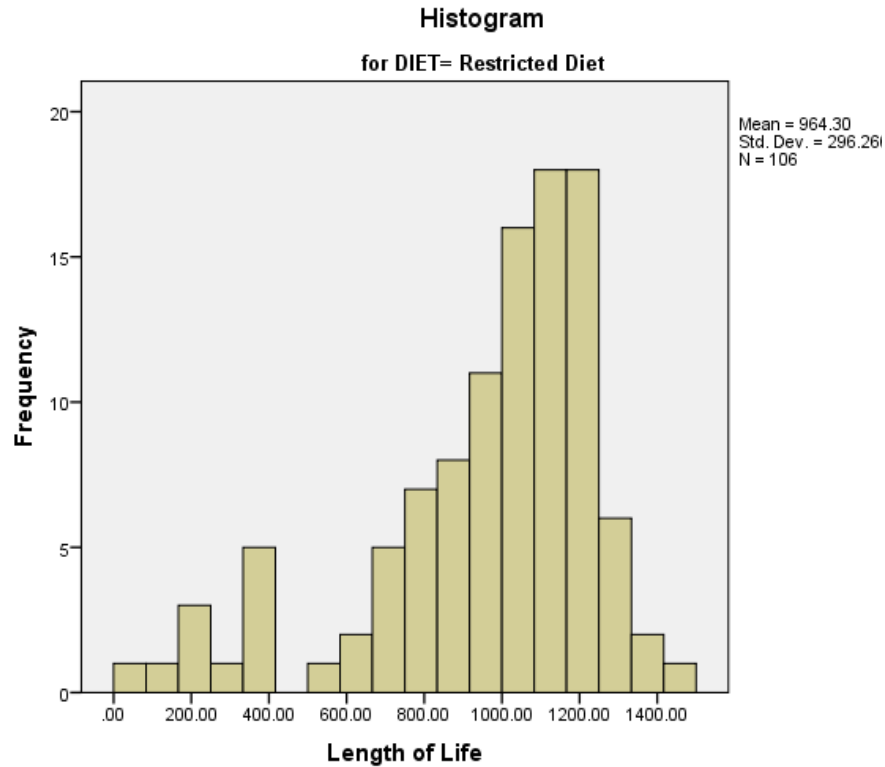
Length of Life Stem-and-Leaf Plot for
DIET= Restricted Diet

Frequency	Stem &	Leaf
7.00	Extremes	(=<363)
3.00	3 .	899
1.00	4 .	0
1.00	5 .	3
2.00	6 .	03
8.00	7 .	11234678
9.00	8 .	011346799
14.00	9 .	00012345566778
20.00	10 .	011111133444567778999
20.00	11 .	00122233334466778889
15.00	12 .	0001222333334569
5.00	13 .	12269
1.00	14 .	3
Stem width: 100.00		
Each leaf: 1 case(s)		

Length of Life Stem-and-Leaf Plot for
DIET= Free Diet

Frequency	Stem &	Leaf
4.00	Extremes	(=<465)
3.00	4 .	799
6.00	5 .	133444
1.00	5 .	8
9.00	6 .	001223334
19.00	6 .	5556666777778889999
22.00	7 .	0011111122333333334444
15.00	7 .	556667777889999
6.00	8 .	000133
3.00	8 .	559
1.00	Extremes	(>=963)
Stem width: 100.00		
Each leaf: 1 case(s)		

Histogram



Descriptive Statistics: Categorical Variable



Class Interval	Frequency	Cumulative Freq.	Percentage	Relative Percentage
<500				
500-1000				
1000-1500				

Conversion of LIFESPAN into Three (03) Categories

SYNTAX

RECODE LIFESPAN (0 thru 500=1) (501 thru 1000=2) (1001 thru 1500=3) INTO category.

VARIABLE LABELS category "Categorical Lifespan".

VALUE LABELS category 1 "<=500" 2 "501-1000" 3 "1001-1500".

EXECUTE.

Try Manually if interested?

Descriptive Statistics: Categorical Variable



Manually

Analysis> Descriptive statistics >Frequency> Select Variable> Ok

Syntax

FREQUENCIES VARIABLES=category.

CATEGORICAL

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	<500	18	9.2	9.2	9.2
	501-1000	116	59.5	59.5	68.7
	1001-1500	61	31.3	31.3	100.0
Total		195	100.0	100.0	

Let's Play with Graphs

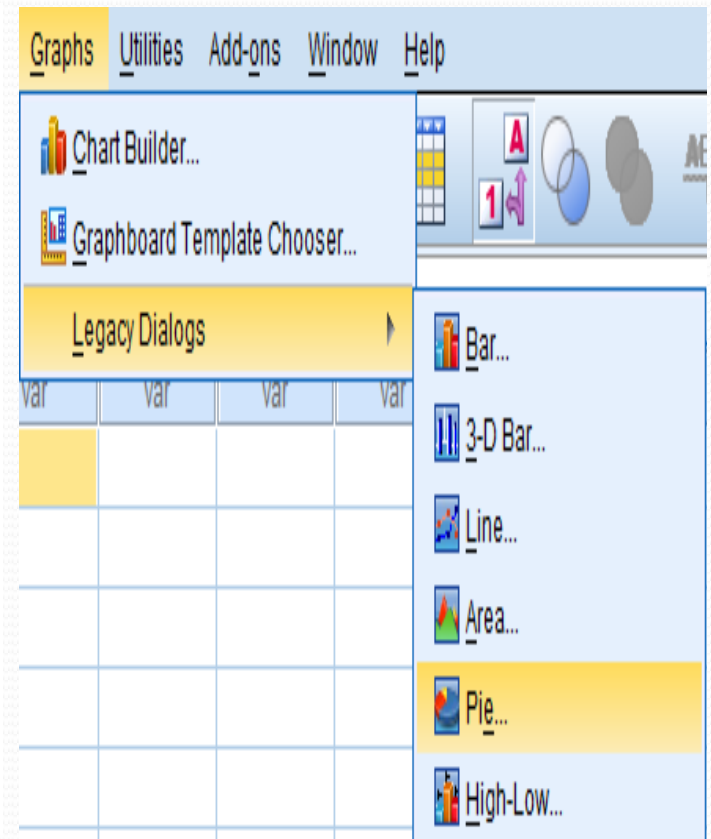
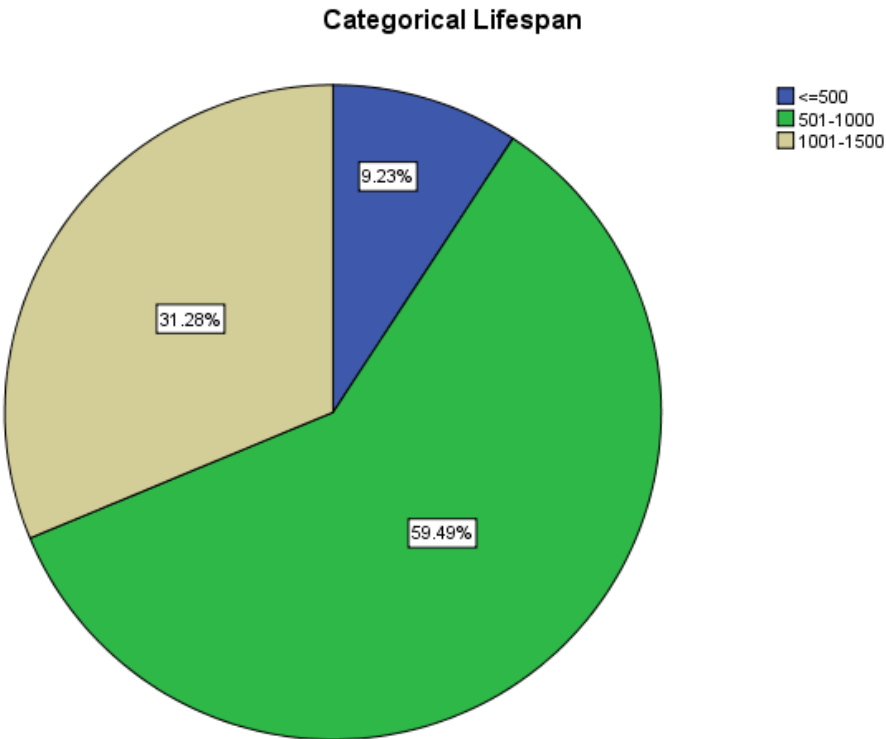


- ▷ How do we show the distribution of rats by lifespan category?
 - ▷ Bar Diagram
 - ▷ Pie Chart
- ▷ How do we show the distribution of lifespan? How do you compare the distributions of lifespan by diet?
 - ▷ Histogram
 - ▷ Boxplot
- ▷ How do we plot a scatterplot of lifespan by diet?
 - ▷ Scatter plot

PIE DIAGRAM



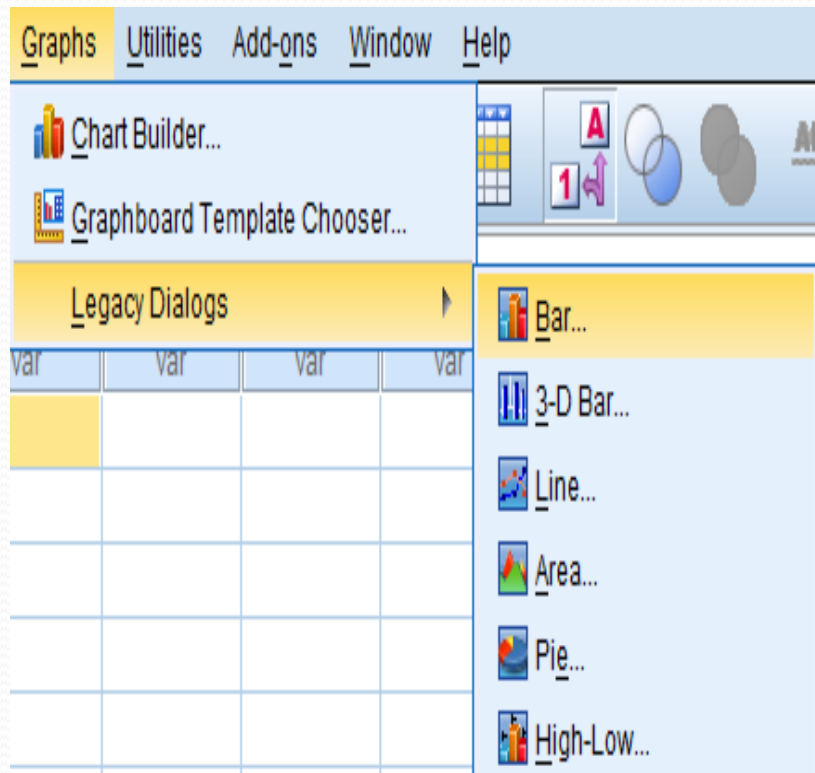
Manually : Graphs -> Legacy dialogs -> Pie



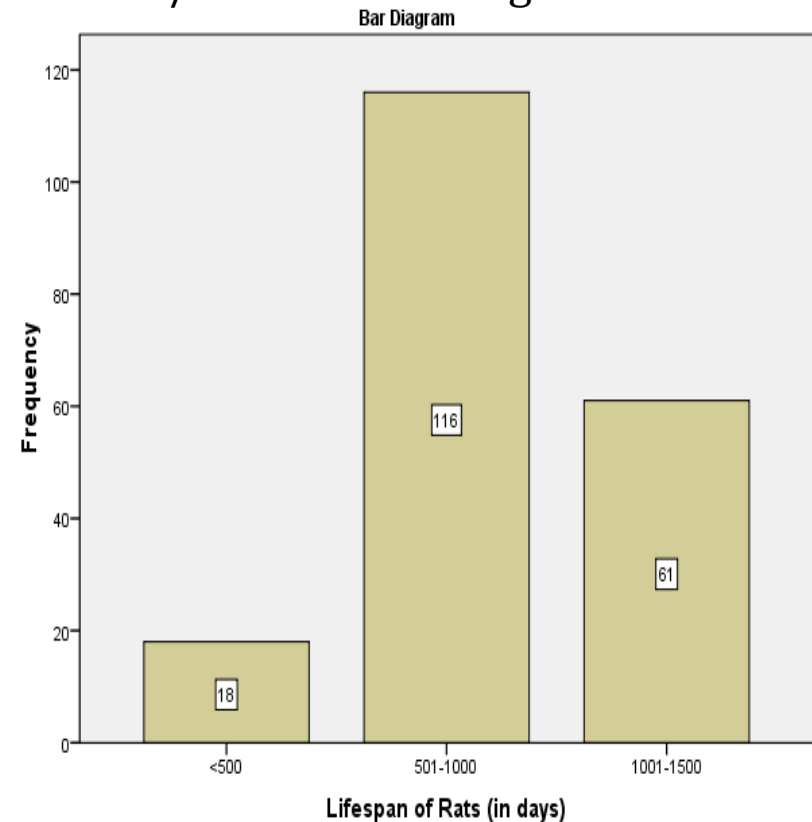


Bar DIAGRAM

Manually: Graphs -> Legacy
dialogs -> bar



Syntax:
GRAPH
/BAR(SIMPLE)=CUFREQ BY
category
/TITLE='Bar Diagram'.

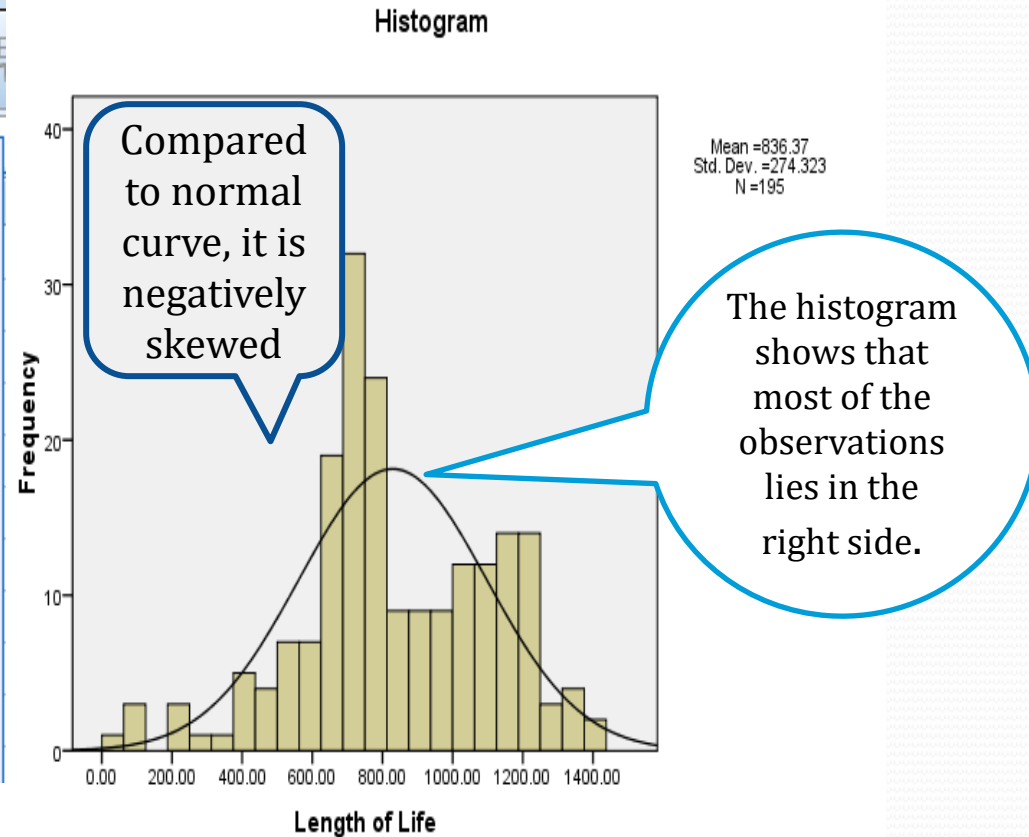
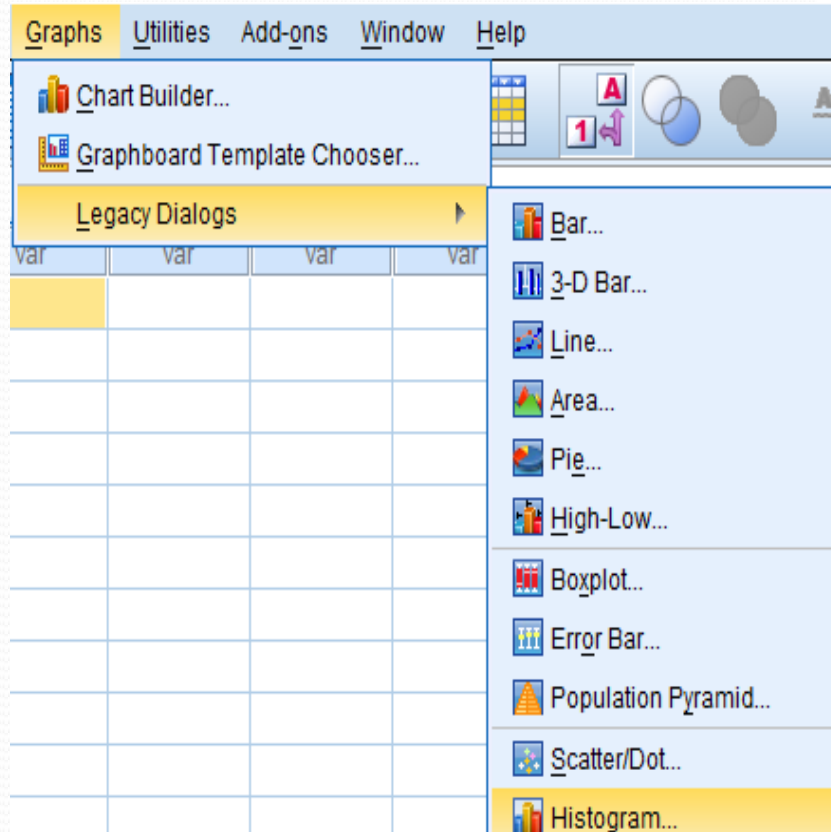


HISTOGRAM

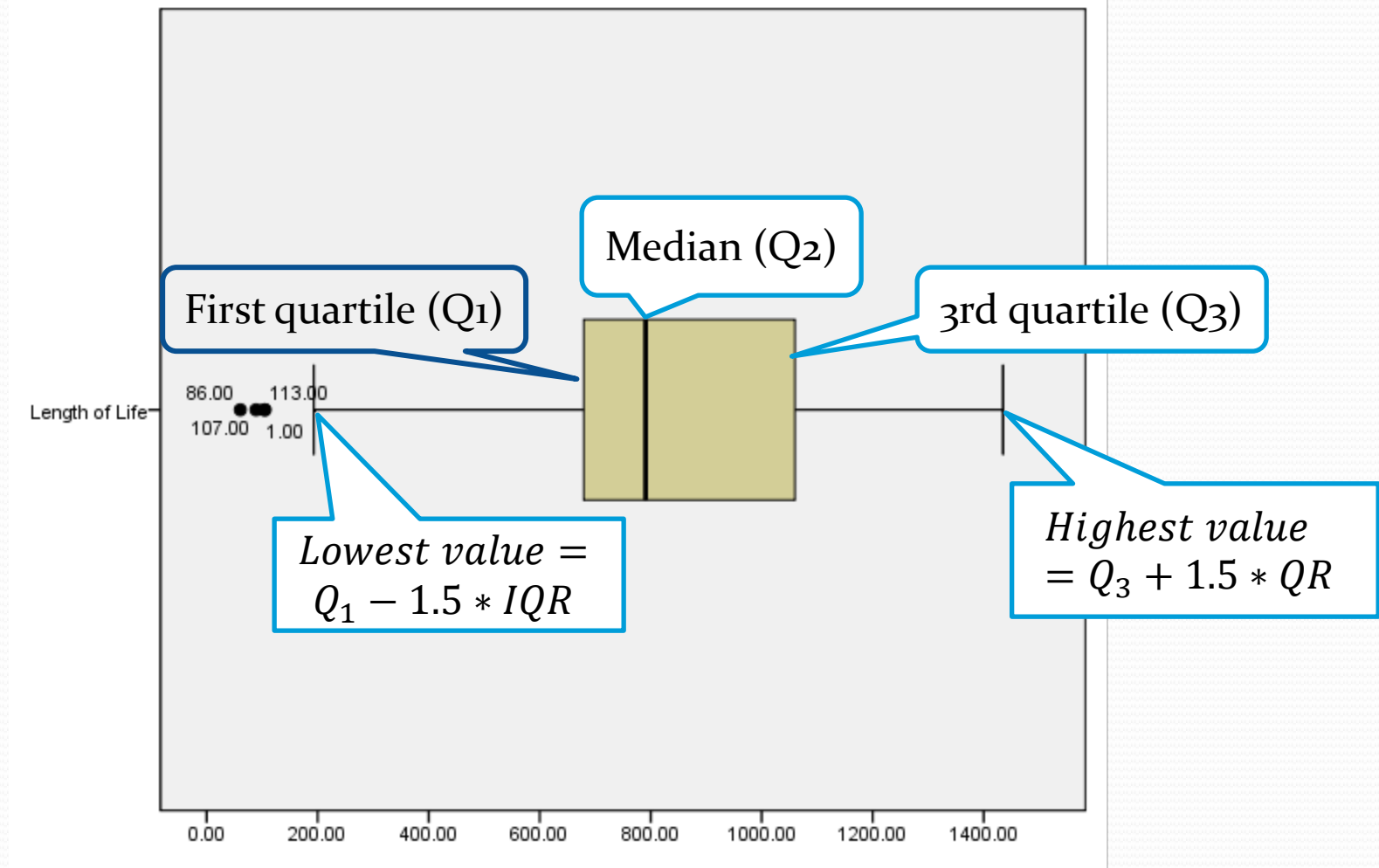


Manually : Graphs -> Legacy
dialogs -> Histogram

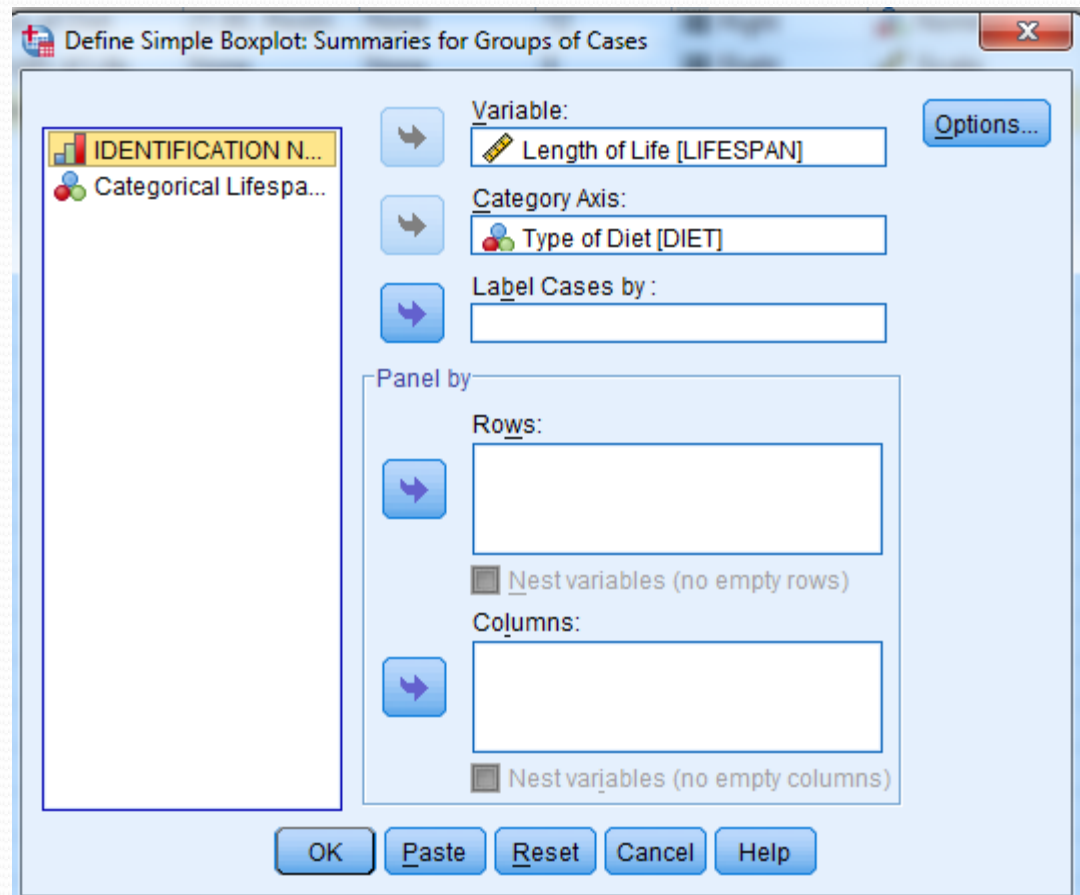
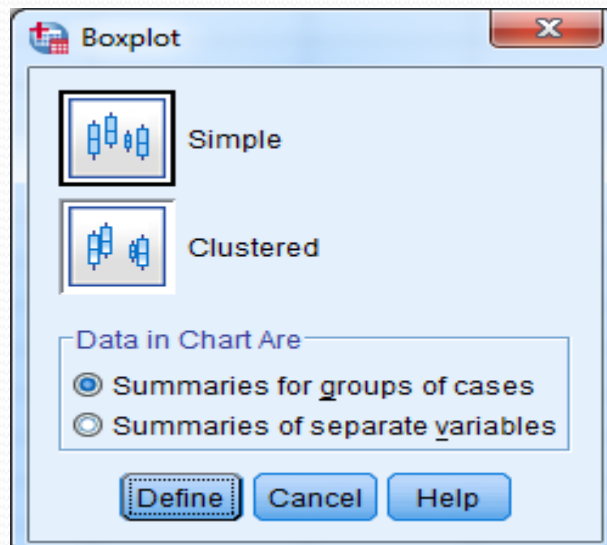
Syntax:
GRAPH
/HISTOGRAM(NORMAL)=LIFESPAN.



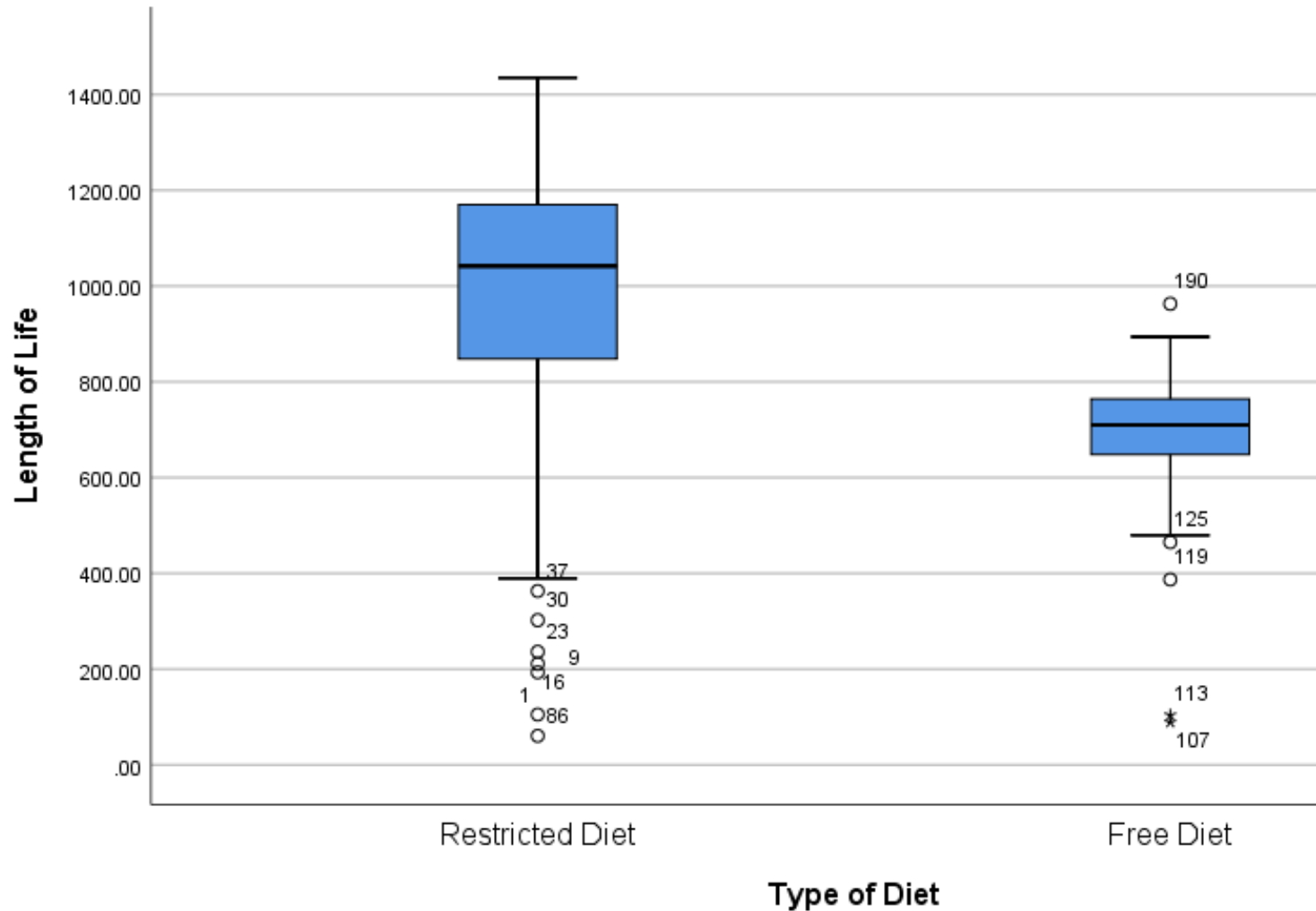
Box-plot



Side by side Box-plot



Side by side Box-plot



- ▷ How can you calculate **Geometric Mean** and **Harmonic Mean** for a continuous variable?
- ▷ Available in **Compare Means** options. But its require a categorical independent variable to compare. But we get total and sub-total. So, we can easily get the result of GM and HM for our interested dependent variable.

Descriptive Statistics



- ▶ Let's run the **Compare Means** command using LIFESPAN as dependent variable and diet as DIET as independent variable.

Manually:

Analysis> Compare Means > Means> Select Dependent Variable> Select Independent Variable> Click on Options Button> Choose the options> Continue>Ok / Paste

Syntax:

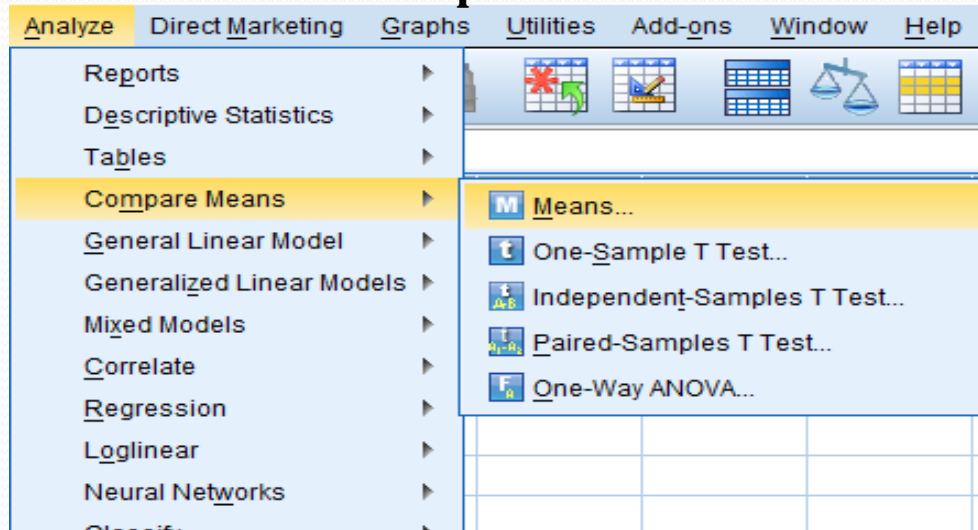
MEANS TABLES=LIFESPAN BY DIET

/CELLS MEAN COUNT STDDEV MEDIAN GMEDIAN SEMEAN
SUM MIN MAX RANGE FIRST LAST VAR KURT SEKURT SKEW
SESKEW **HARMONIC GEOMETRIC** SPCT NPCT

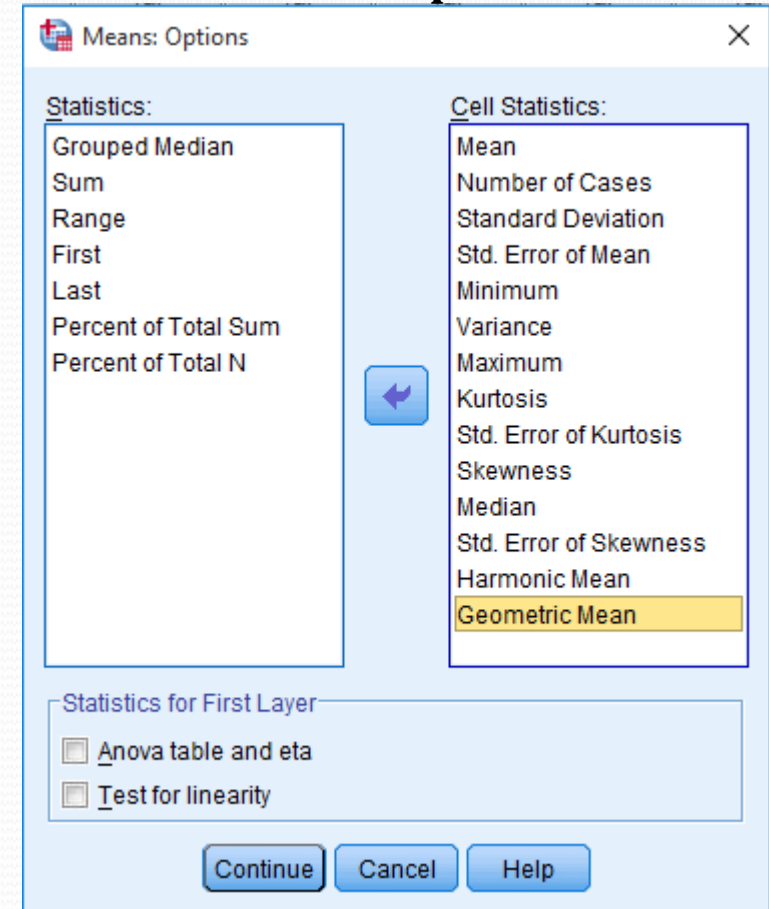
Descriptive Statistics



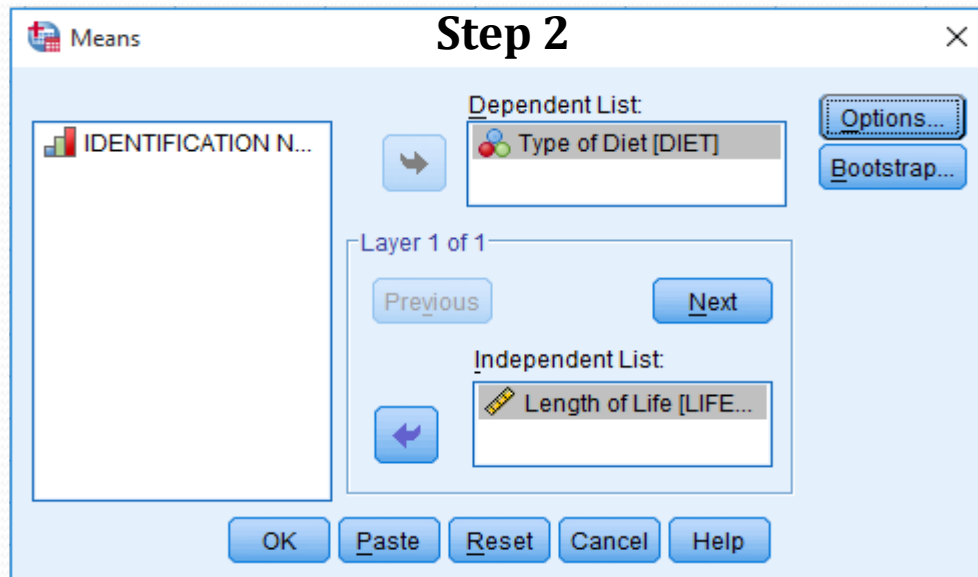
Step 1



Step 3



Step 2



Output



Report

Length of Life

Type of Diet	Mean	Std. Deviation	Median	Range	First	Last	Variance	Kurtosis	Std. Error of Kurtosis	Skewness	Std. Error of Skewness	Harmonic Mean	Geometric Mean
Restricted Diet	964.2972	296.26632	1042.0000	1374.50	105.00	1393.00	87773.732	1.199	.465	-1.244	.235	702.4770	882.7262
Free Diet	684.0112	134.08423	710.0000	874.00	89.00	788.00	17978.579	7.027	.506	-2.010	.255	598.8583	659.7834
Total	836.3718	274.32259	791.0000	1374.50	105.00	788.00	75252.883	.046	.346	-.260	.174	651.0618	772.9022

These are the
required HM and
GM for Total

► Basics on Cross-tabulation

- The cross-tabulation analysis is the basic technique for examining the relationship between two or more **categorical** (nominal or ordinal) variables (attribute), possibly controlling for additional layering variables.
- The **Crosstabs** procedure offers tests of independence and measures of association for nominal and ordinal data.
- Additionally, you can obtain estimates of the relative risk of an event given the presence or absence of a particular characteristic.

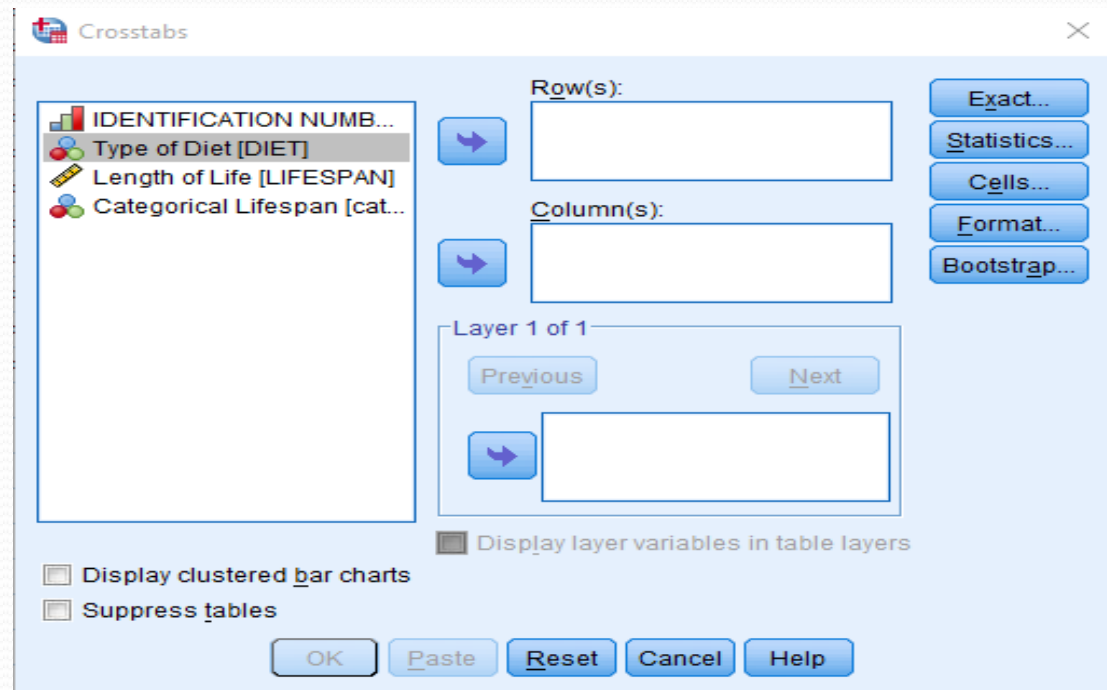
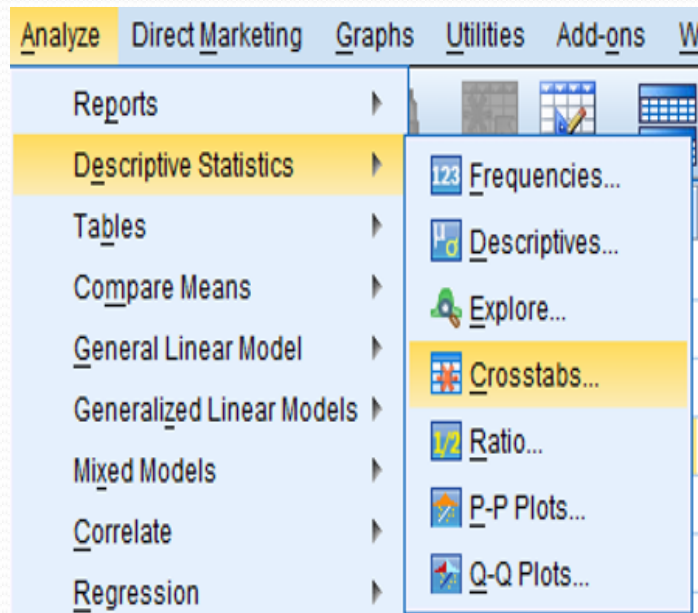
How do we examine the association?



Lifespan Group / Diet	Restricted	Free	Total
<500			
500-1000			
1000-1500			
Total			

Manually	Syntax
<i>Analyze> Descriptive Statistics >Crosstabs... >Select Column Variable > Select Row Variable> Choose other options> Continue > OK</i>	<i>CROSSTABS /TABLES=X3 BY X9 /FORMAT=AVALUE TABLES /STATISTICS=CHISQ /CELLS=COUNT ROW /COUNT ROUND CELL.</i>

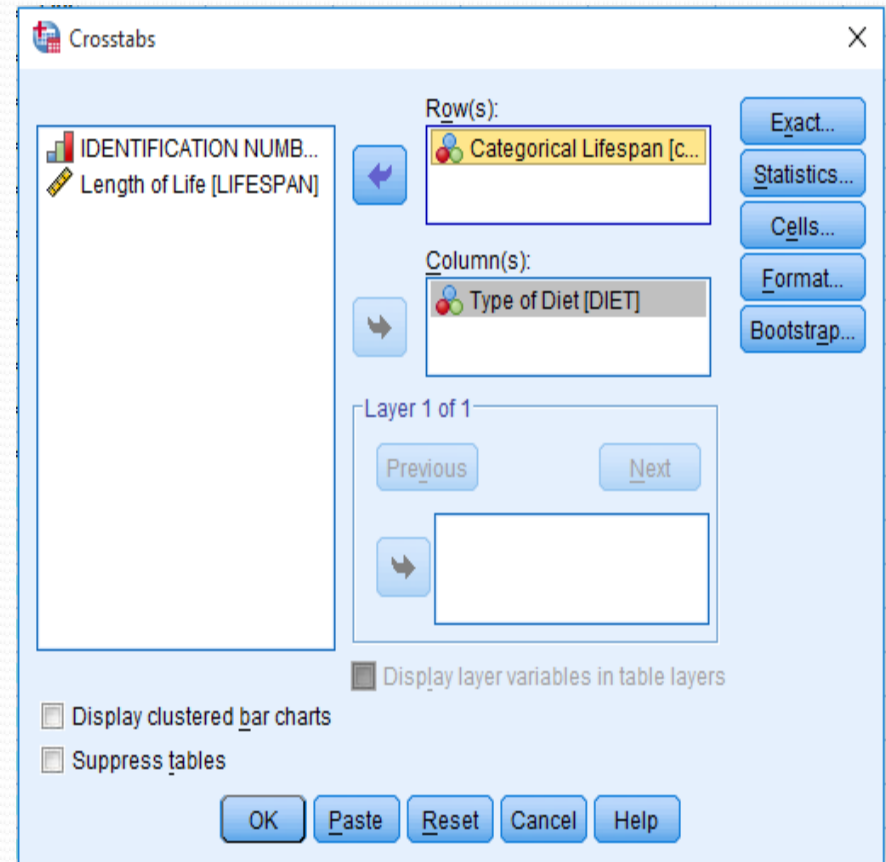
Cross-tabulation



Cross-tabulation



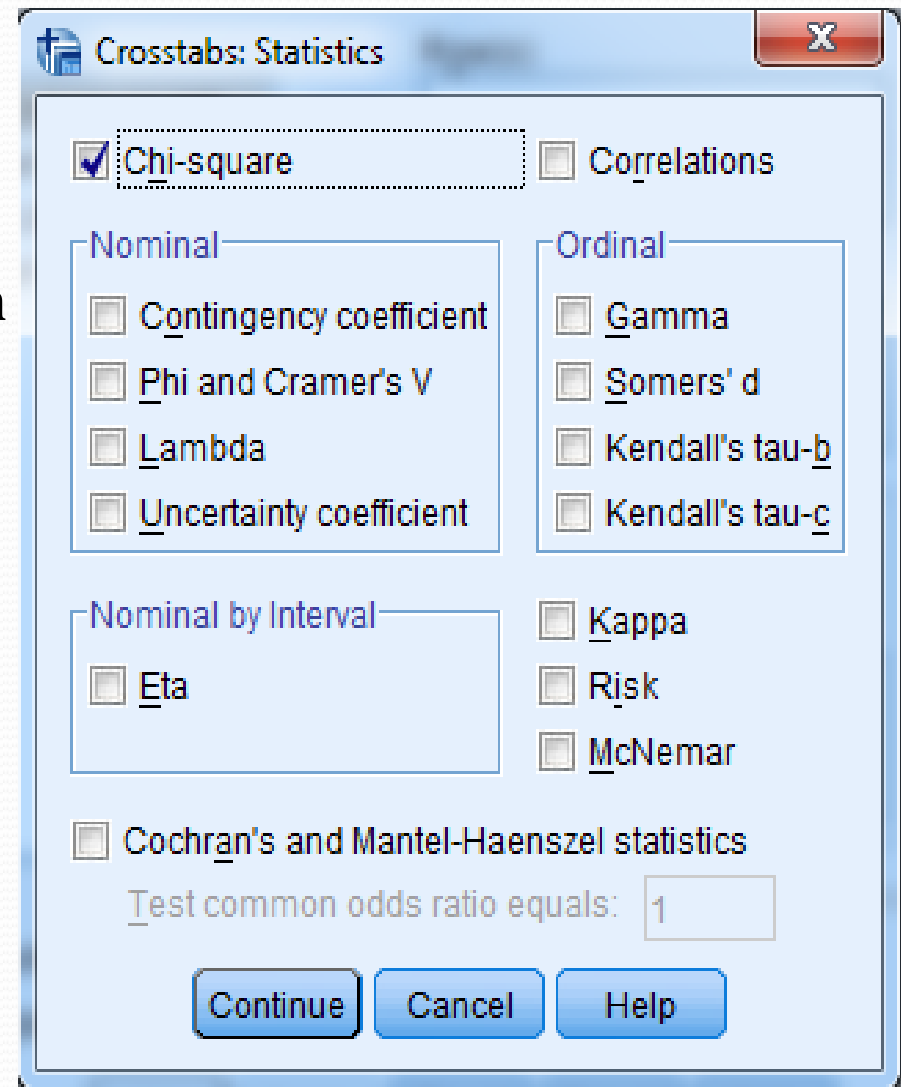
- ▶ Column variable: **Type of Diet**
- ▶ Row Variable: **Categorical Lifespan**
- ▶ We want to see association between “**Type of Diet**” and **Categorical Lifespan**
- ▶ Select the “**Categorical Lifespan**” and Click on the arrow (>) to transfer them to the **Row box**. Click on “**Type of Diet**” to transfer it to the **columns box**.



Cross-tabulation



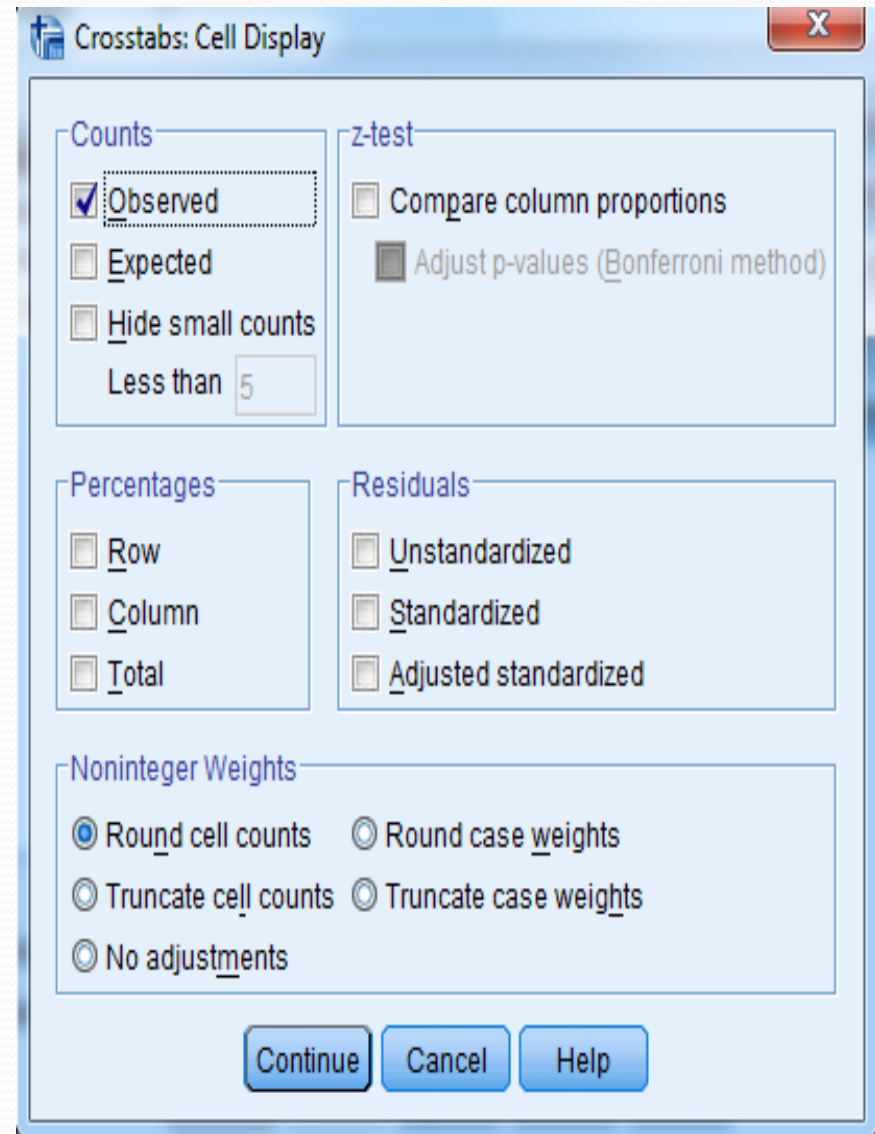
- ▶ All we need to do now is to select some options and run the procedure.
- ▶ Click on the **Statistics** button and its dialogue box will be loaded on the screen.
- ▶ Select **Chi-square** to test the goodness of fit and other options as you need.



Cross-tabulation



- ▶ From **Cell Display** dialog box select **observed** as counts.
- ▶ You can select **Expected** as well. In percentages select any of the **Row**, **Column** or **Total**.
- ▶ You also can get some other statistics as you require.
- ▶ We here select **observed** as count and **Row** as percentages.



Type of Diet vs. Categorical Lifespan



Categorical Lifespan * Type of Diet Crosstabulation

			Type of Diet		Total
			Restricted Diet	Free Diet	
Categorical Lifespan <=500	Count		11	7	18
	% within Categorical Lifespan		61.1%	38.9%	100.0%
501-1000	Count		34	82	116
	% within Categorical Lifespan		29.3%	70.7%	100.0%
1001-1500	Count		61	0	61
	% within Categorical Lifespan		100.0%	0.0%	100.0%
Total	Count		106	89	195
	% within Categorical Lifespan		54.4%	45.6%	100.0%

Test of Independence



Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	80.884 ^a	2	.000
Likelihood Ratio	104.448	2	.000
Linear-by-Linear Association	40.893	1	.000
N of Valid Cases	195		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.22.

Thank You!