

Workshop 3

More about normality assessment + Graphs

Task 2 from previous Workshop



- Access data file ICSR from the LMS.
- We want to have the descriptive statistics of the variable ICSR2.

First: Decide the type of measurement. Is it ordinal, nominal or scale? In our case it is scale.

Second: go to analyze in the menu bar>> Descriptive Statistics>>> Explore>>> Move the variable ICSR2 to the dependent list>>> Click on Statistics, and ensure Descriptive is chosen>>> Click on Plots, un-tick steam-and-leaf, then tick Normality plots with tests>>>Click continue>> then OK.



Interpreting the results

Number of observations

Number of observations regarded by SPSS

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
ICSR2	27	100.0%	0	0.0%	27	100.0%

The estimated difference between the mean of the population and the mean of the sample

Descriptives

		Statistic	Std. Error
ICSR2	Mean	23.5556	1.07990
	95% Confidence Interval for Mean	Lower Bound	21.3358
		Upper Bound	25.7753
	5% Trimmed Mean	23.5473	
	Median	24.0000	
	Variance	31.487	
	Std. Deviation	5.61134	
	Minimum	13.00	
	Maximum	34.00	
	Range	21.00	
	Interquartile Range	9.00	
	Skewness	-.064	.448
	Kurtosis	-.784	.872

What is this?

Where is the mode?

We are 95% confident that the mean is between those two bounds

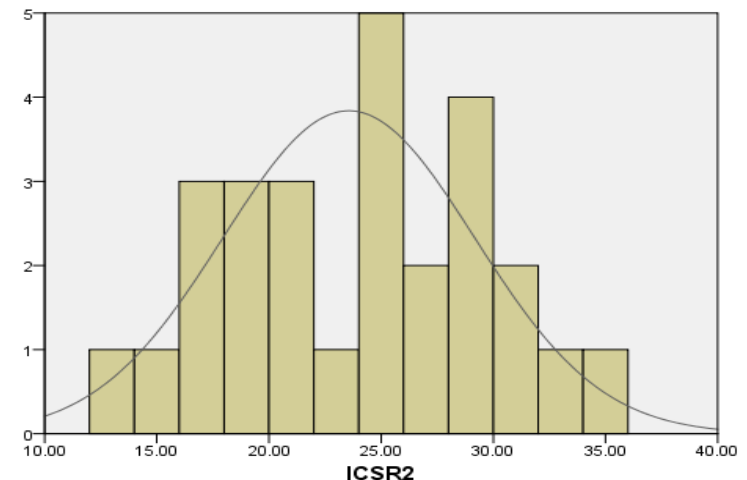
The mean without 5% of the observations in the upper and lower ends of the data

Interpreting the results + Histogram

Go to Graphs in the menu>>> select Graphboard Template Chooser>>> Select the same variable of the previous example>>>Make sure the Basic tab is selected>>>Select Histogram with normal distribution>>>OK.

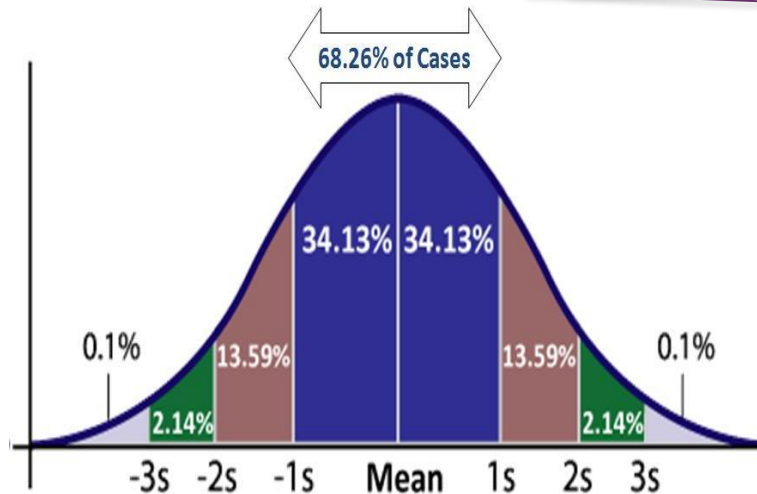
Descriptives

			Statistic	Std. Error
ICSR2	Mean		23.5556	1.07990
	95% Confidence Interval for Mean	Lower Bound	21.3358	
		Upper Bound	25.7753	
	5% Trimmed Mean		23.5473	
	Median		24.0000	
	Variance		31.487	
	Std. Deviation		5.61134	
	Minimum		13.00	
	Maximum		34.00	
	Range		21.00	
	Interquartile Range		9.00	
	Skewness		-.064	.448
	Kurtosis		-.784	.872



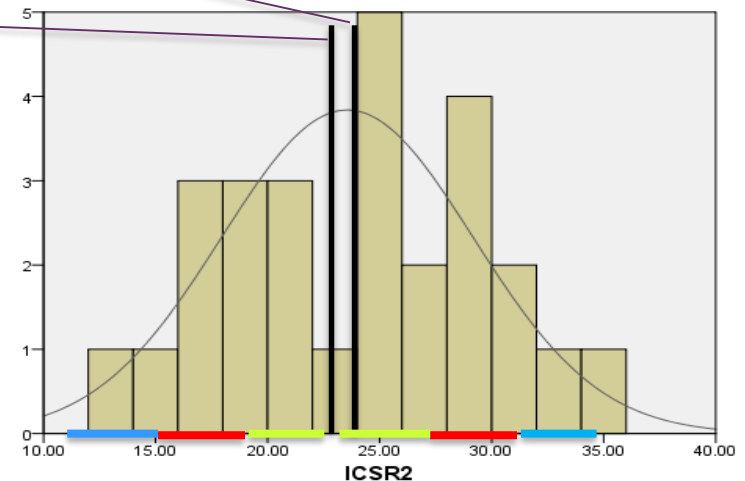


Interpreting the SD



The mean (23.6)

The median (24)



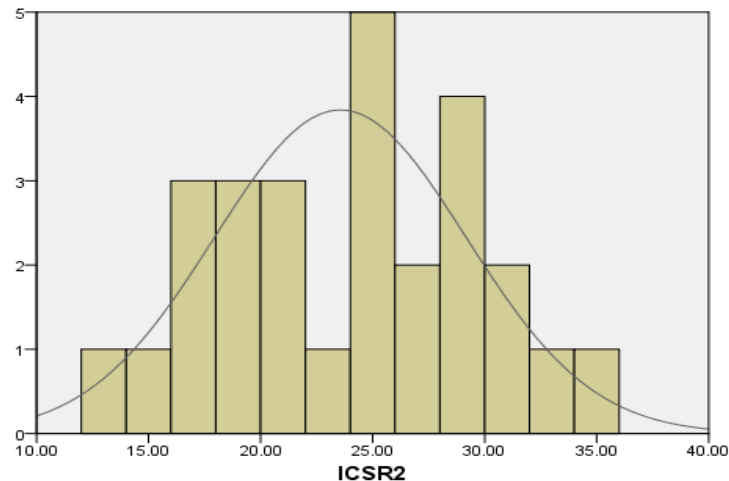
- 68% of the values occur within (-1s to 1s).
- 95% of the values occur within (-2s to 2s).
- 99% of the values occur within (-3s to 3s).



How to report it?

Descriptives

			Statistic	Std. Error
ICSR2	Mean		23.5556	1.07990
	95% Confidence Interval for Mean	Lower Bound	21.3358	
		Upper Bound	25.7753	
	5% Trimmed Mean		23.5473	
	Median		24.0000	
	Variance		31.487	
	Std. Deviation		5.61134	
	Minimum		13.00	
	Maximum		34.00	
	Range		21.00	
	Interquartile Range		9.00	
	Skewness		-.064	.448
	Kurtosis		-.784	.872



Descriptive statistics indicates that the scores of ICSR2 have a mean of 23.6 with 1.08 standard error from the mean of the population. A confidence interval of 95% of the mean makes it located between the upper bound of 25.8 and the lower bound of 21.3. The trimmed mean value is very closed to the mean value explaining the low influence of the extreme values. With a value of 24 that is close to the mean, the distribution of the scores is skewed slightly to the upper end of the scores with a value of $-.06$ as shown in the figure bellow. The standard deviation value is 5.6 can give more insights about the shape of the distribution when taking into consideration the minimum and maximum scores of 13 and 34 respectively.



Criteria to decide on normality

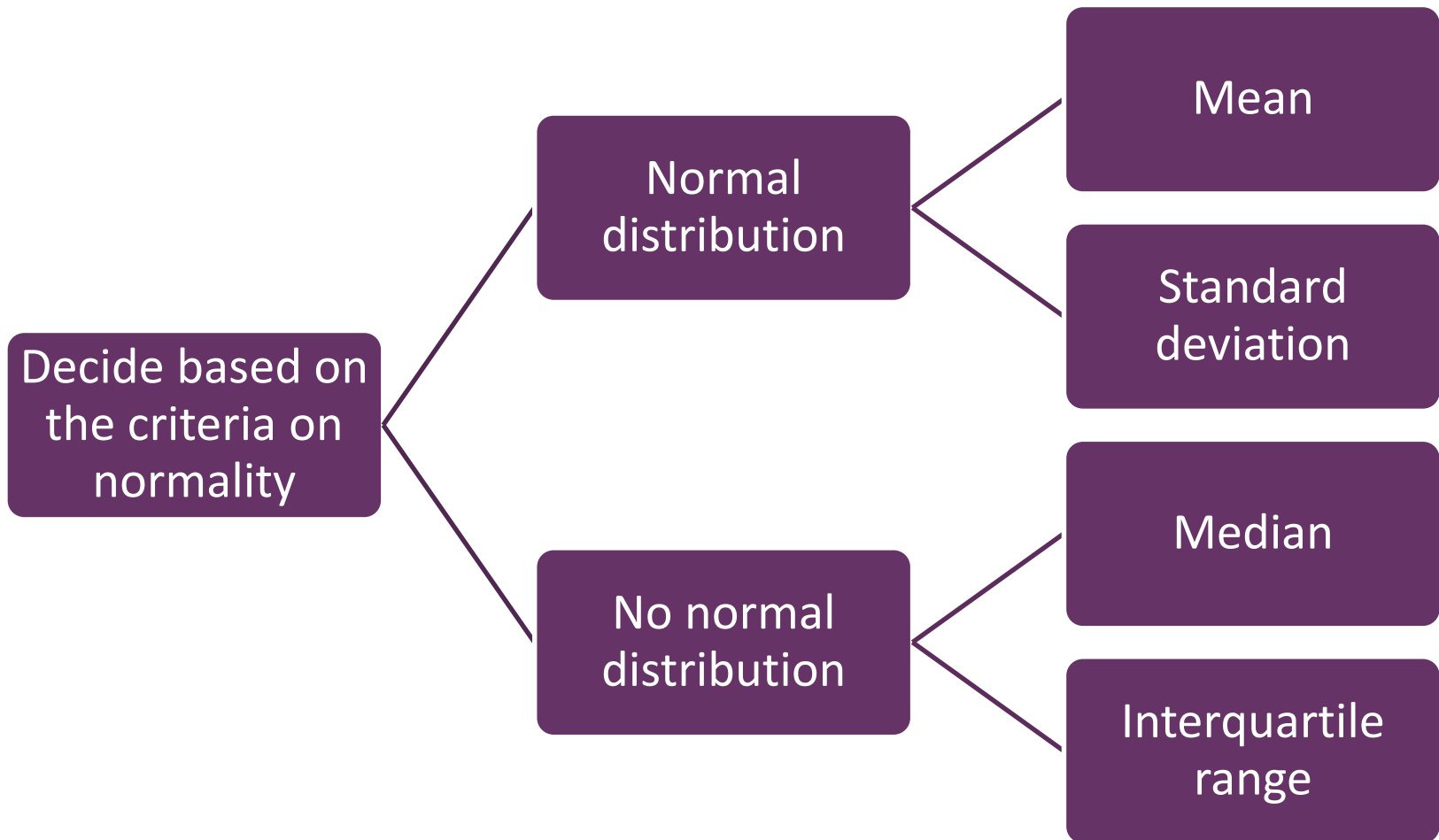
- Histogram: the distribution of the data according to the normal distribution curve.
- Q-Q Plot: the distribution of the data according to the line.
- Skewness value: There is no cut point; however, some statisticians may consider the value of ± 0.8 as a factor to consider when compromising all the criteria mentioned in this slide.
- The difference between the mean and the median. This should be considered along with the minimum and maximum scores of the variable to decide on how large is that difference between the mean and the median.
- Kolmogorov-Smirnov and Shapiro-Wilk tests are not always reliable. Thus, their results must be compromised with the above criteria.
- **Deciding on normality is a bit subjective**



What to report

- Nominal variables: mode – frequencies – Pie or bar chart - a paragraph.
- Ordinal variables: mode – median – frequencies – Bar chart- a paragraph.
- Scale variables: Descriptive Table> Mean, lower and upper bounds of the mean, trimmed mean, median, skewness, kurtoses, histogram>> Q_Q Plot, Kolmogorov-Smirnov and Shapiro-Wilk tests>>>standard deviation>>Interquartile range.

How to describe a 'scale' variable





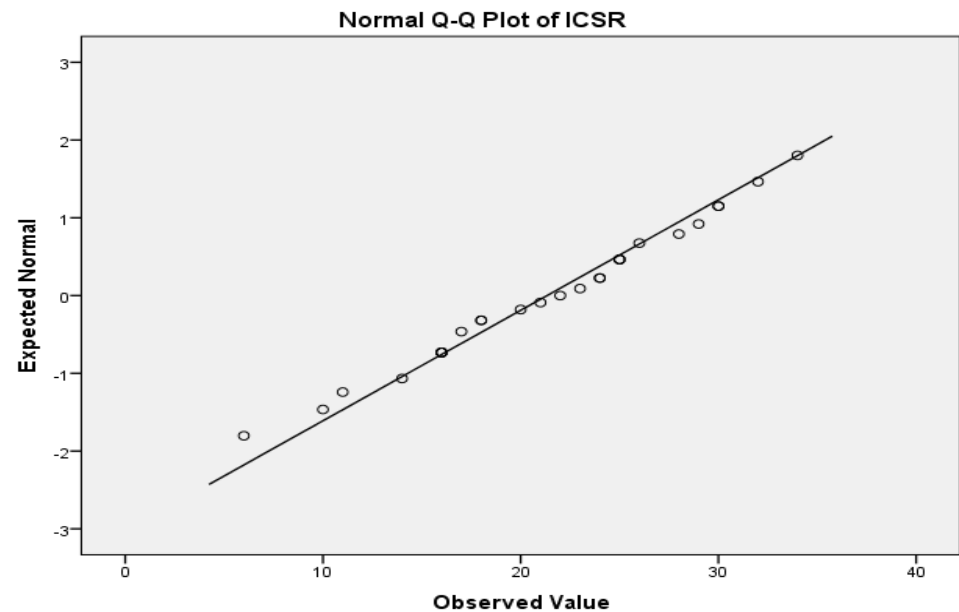
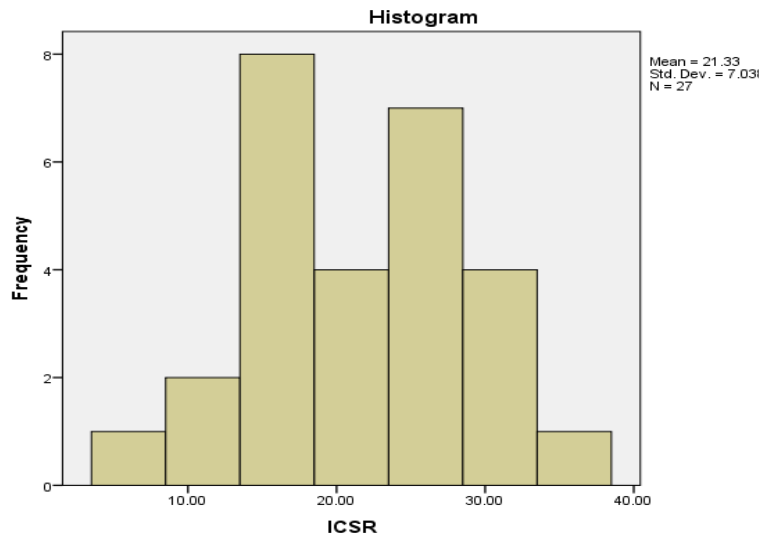
More about normality

- Assessing normality
 - Many of the statistical techniques assumes that the data are normally distributed (normality)
 - Normality concerns the shape of the distribution of the data (bell shape).
 - Procedure- Analyze-Descriptive statistics-explore-select ICSR2 and move it to the independent list-Statistics section>> select descriptive>>> Plot section>> Select histograms, untick Stem and Leaf, and select normality plots with tests>>>Options section, tick exclude case pairwise>>> continue and OK



Normality assessment

- Assessing normality
 - Practice: provide normality assessment for the variable ICSR.

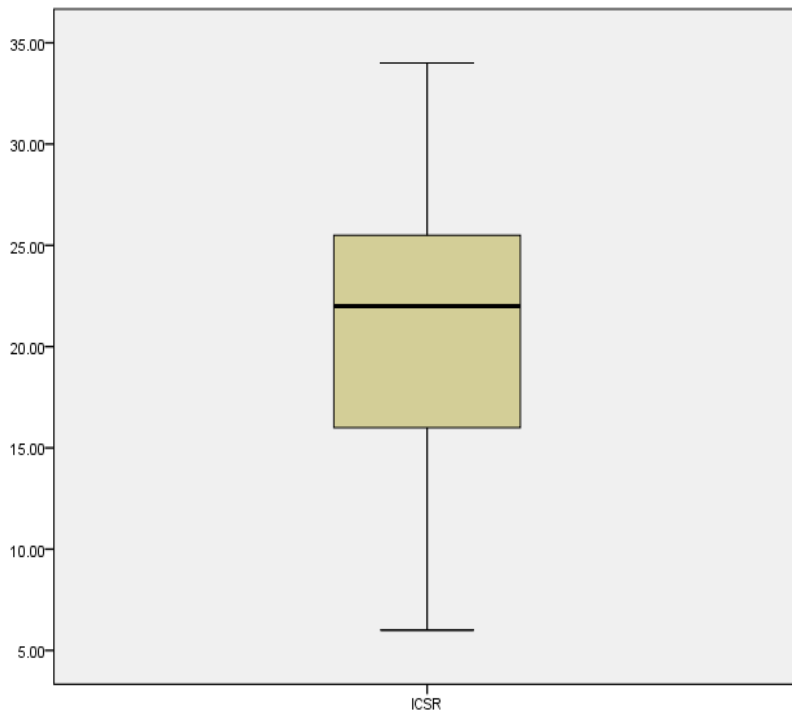


Histogram : shows that scores have been close to normal distribution.

Normal Q-Q Plots: The observed value of each score is plotted against the expected value from the normal distribution. A reasonably straight line suggests a normal distribution.

Normality assessment

- Assessing normality
 - Practice: provide normality assessment for the variable ICSR.



Boxplot: the rectangle represents 50% of the cases, the line going up to the largest and the line going down to the smallest.

This test is good for outliers. In our case we do not have outliers, but when they are present, SPSS will show them as small circles above the line of the largest cases and below the line of the smallest cases.



Normality assessment

Check the significance

Check the significance

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ICSR2	.088	27	.200 [*]	.976	27	.761

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Descriptive statistics indicates that the scores of ICSR2 have a mean of 23.6 with 1.08 standard error from the mean of the population. A confidence interval of 95% of the mean makes it located between the upper bound of 25.8 and the lower bound of 21.3. The trimmed mean value is very closed to the mean value explaining the low influence of the extreme values. With a value of 24 that is close to the mean, the distribution of the scores is skewed slightly to the upper end of the scores with a value of -.06 as shown in the figure bellow. The standard deviation value is 5.6 can give more insights about the shape of the distribution when taking into consideration the minimum and maximum scores of 13 and 34 respectively. **With non-significant results of Kolmogorov-Samirnov and Shapiro-Wilk tests, the data is almost normally distributed.**

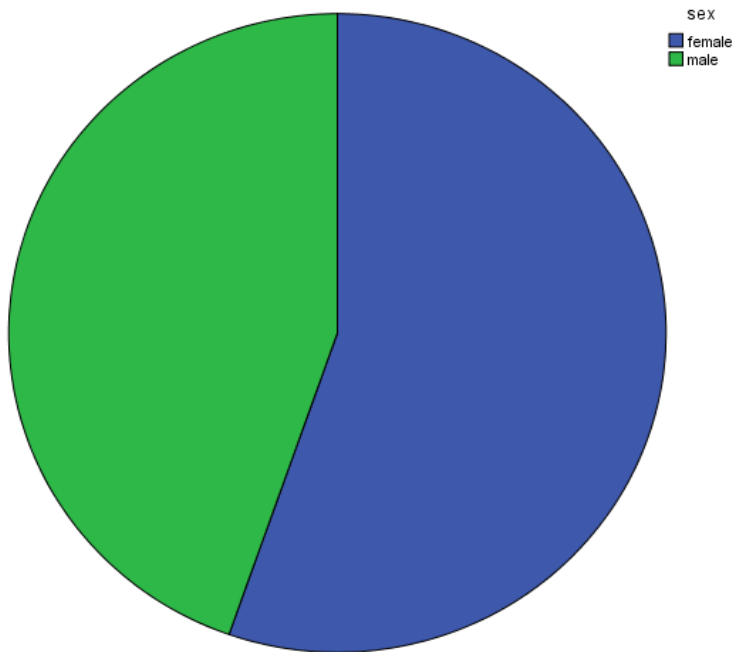
In addition to all that we has explained about normality assessment. The most important tests are Kolmogorov-Samirnov and Shapiro-Wilk tests.

In this regard, we need to check the significance of the two tests. If they are less than 0.05 then the data is NOT normally distributed. Otherwise, the data is normally distributed.

Graphical presentation using SPSS.

Single Variable: The Pie Chart

Data at **nominal** and **ordinal** level for a single variable can be presented in a pie chart which illustrates the proportion of the total falling into each category –they work best when there are few categories.



Create a pie chart for the variable “Sex” in the data file.

Procedure: Graphs>>>Legacy

Dialogues>>> Pie chart>>>Summaries

for groups of cases>>> Define>>>

move the variable “Sex” to the Define slices by>>>Click OK.

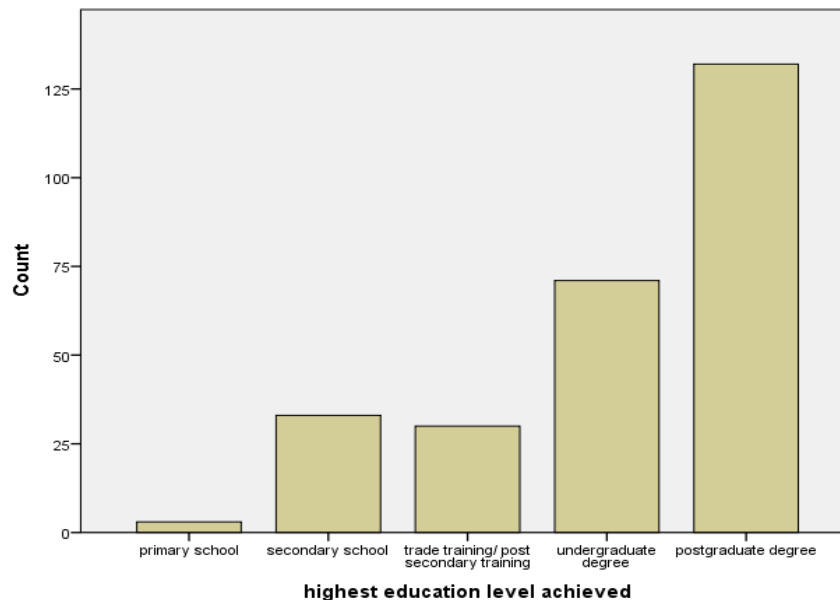
Single Variable: The Bar Chart



Murdoch
UNIVERSITY

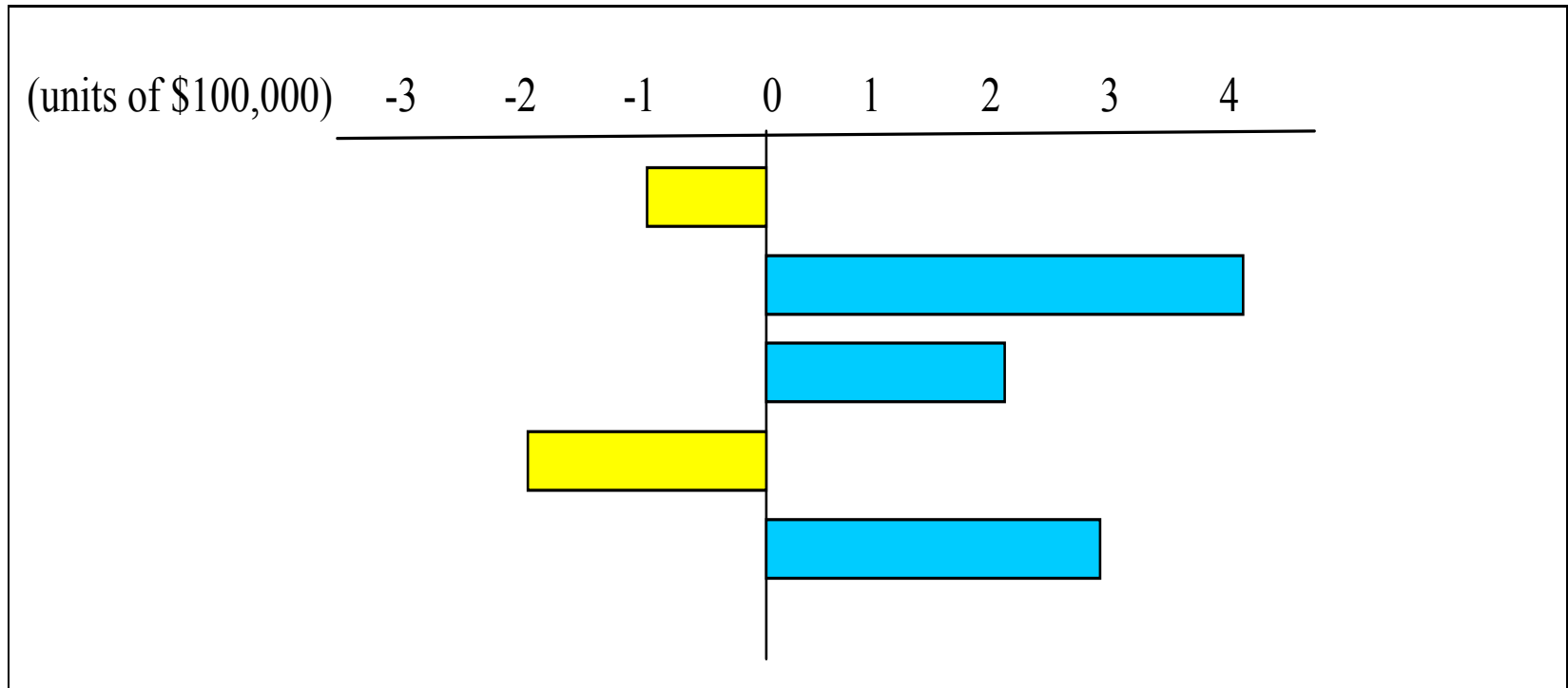
Data presented in **nominal** and **ordinal** scales for a single variable can be presented in a bar chart.

The frequency (or relative frequency) of each category is represented by a vertical bar.



Create a bar chart for the variable "edlevel".

Two Direction bar Chart



Single Variable Box Plots



Murdoch
UNIVERSITY

Box plots are used with **interval** and **ratio** scale data. Note SPSS usually prints them in a vertical dimension (next slide)

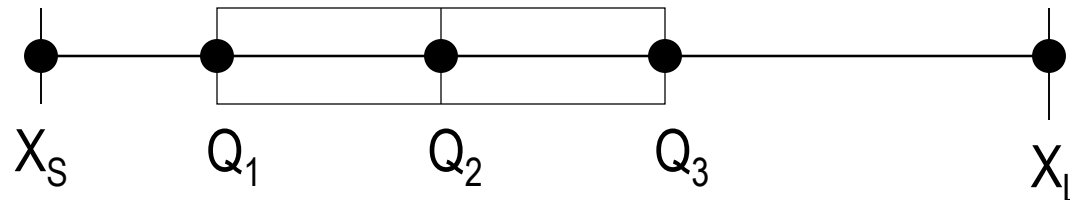
X_S - The smallest observation

Q_1 - The lower quartile

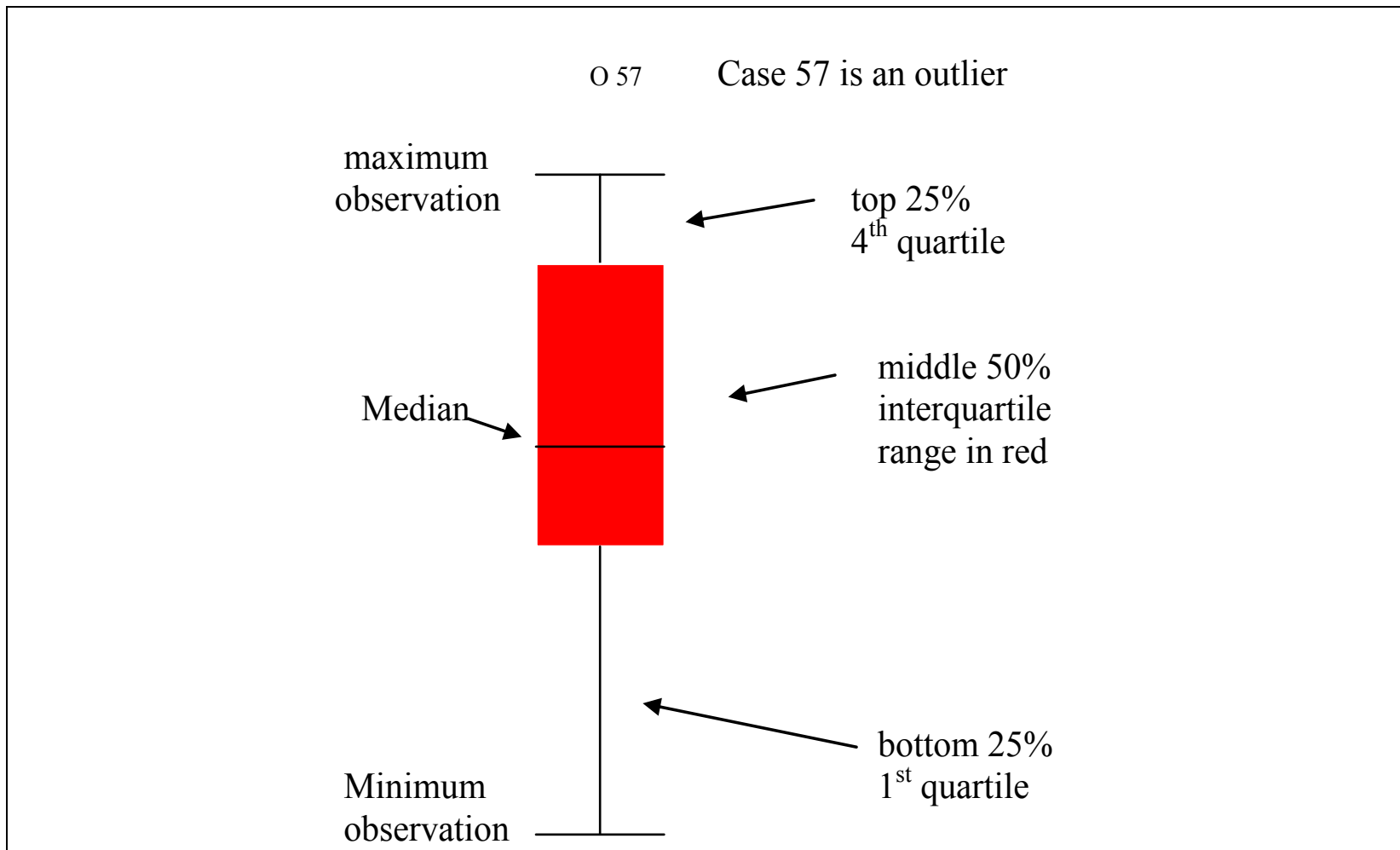
Q_2 - The median

Q_3 - The upper quartile

X_L - The largest observation



Box Plot

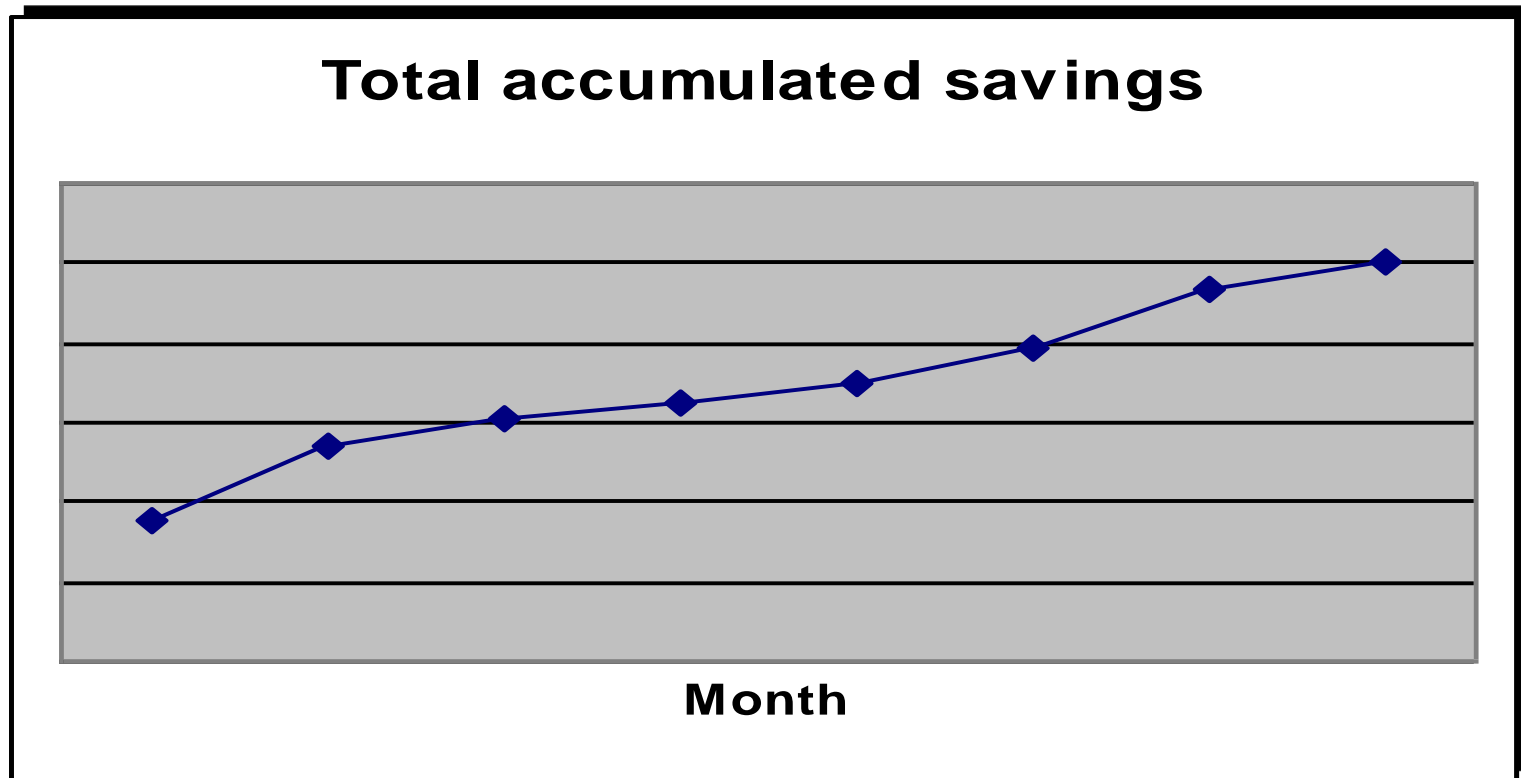


Single Variable: The cumulative line graph



Murdoch
UNIVERSITY

A cumulative line graph) provides a running total of a single variable measured at several points in time. Usually used with **interval** or **ratio** scale data.

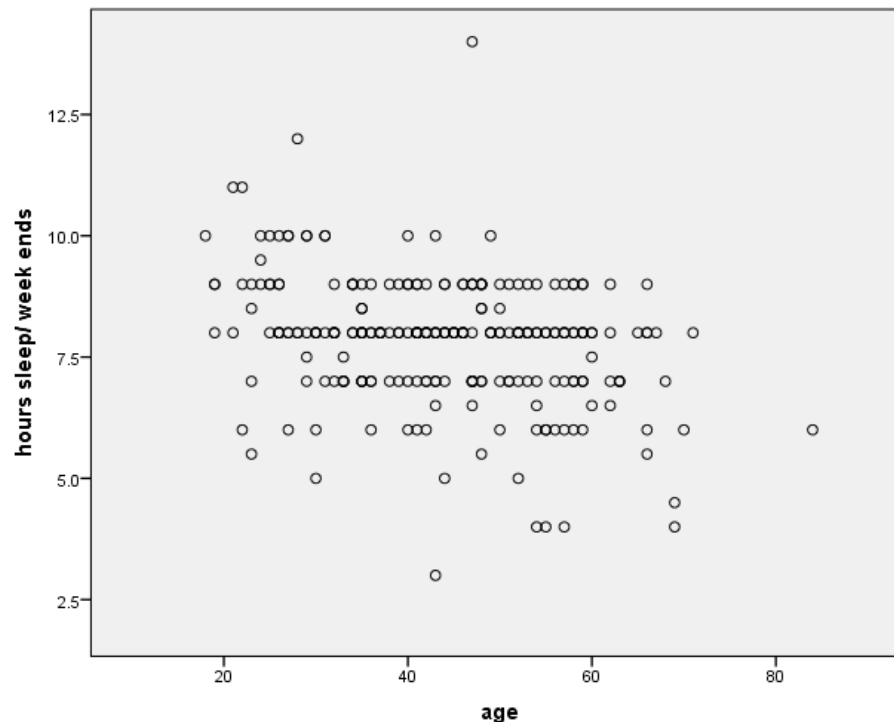


Two Variables: Scatter Plots



Murdoch
UNIVERSITY

A scatter plot usually used to show the relationship between two variables when both are on **interval** or **ratio** scales. Each point on the scatterplot represents a single case.



Create a scatter
plot for the
variables age
and "hourwnit".



Crosstabulation

- How my data is distributed according to categories of specific variables.
- Analyze>>>Descriptive Statistics>>>Crostats>>> Move "Smoke" to the rows>>> Move "Sex" to the columns>>> Click OK.

sex * do you smoke Crosstabulation

Count

		do you smoke		
		yes	no	Total
sex	female	19	131	150
	male	15	105	120
Total		34	236	270

In one paragraph, describe this table.

Thank you

Please refer to the LMS and complete
the tasks after the workshop