EXAM CHEAT SHEET used for 2015

Metric variables- Metric variables are mostly variables that can be measured in terms of **NUMBERS** of something e.g. number of bowls of cereals eaten, number of hours in labour, number of flowers in the garden etc.

Categorical variables- You can identify them by the keywords "**Type of/mode of**" e.g. type of garbage, mode of transport, type of flowers in the garden or even the type of mobile phones students own.

- Gender is a categorical variable

What is a Z-score? A z-score allows any value in a normal distribution to be represented by the number of standard deviations it is from the mean. The z-score for the mean of a normal distribution will always be zero. For any other value in the distribution, a negative z-score indicates that it is lower than the mean, while a positive z-score indicates that it is greater than the mean. As 68% of values on a normal distribution are within 1 SD of the mean, most z-scores are between -1 and 1.

Calculation for Z score To calculate a z-score for any value in a normal distribution, you find the difference between that value and the mean of the distribution, and then divide this difference by the standard deviation (SD) of the distribution.

That is, z = (value - mean) / SD

Example: Suppose that in 2012 the average amount of time that an overseas visitor stays in Sydney is 7 days, with a standard deviation of 2 days. Josh visits Sydney and stays for 12 days. What is the z-score for this stay?

Answer: Z= (12-7)/2= 2.5

Histogram/a Bell-curve- It is used to look at the shape of the distribution of metric data.

Percentage Table-It only shows the percentage of outcomes in a categorical variable. Therefore, the main difference is the variable examined.

Describing/reporting distribution: it is done in terms of shape (include a histogram and say symmetrical/asymmetrical and positively or negatively skewed), center (mean for symmetric and median for skewed data), spread (middle 50% and SD if it's symmetrical), and outliers (any extremely high or low values? state figures).

Example: The distribution of house prices in a sample of 542 houses is displayed in figure 1 (or the given figure). The distribution is positively skewed with 50% of houses priced at $430,000 or less. Typically, houses were priced between $390,000 and $550,000 with half of the houses priced within this range. Two houses had exceptionally high prices of over 1,000,000.

95% Confidence Interval (CI) example: We can be confident that the mean time spent watching TV for Australian university students is between 3.9hrs and 4.2hrs (lower and upper bound values.

Confidence interval is the interval within which the population proportion is likely to lie.

The significance figure (Sig.) is the p-value.

Giving t-test, provide the following 3 items: the t value, the degrees of freedom (df), and the p value e.g. t(443)= 11.67, p<.001.

Any p-value other than .000 is reported exactly as it is given- to 3 decimal places

Types of studies: Observational- the experimenter does not manipulate the independent variable (IV). S/he just observes the relationship between the IV and the DV.

Experimental- where the IV is manipulated, and any changes in the DV are observed.

Types of Variables:Nuisance variables- other variables that increase the variation on the DV, making it more difficult to detect the relationship between the ID and the DV.

- Nuisance variables are always present in a study, but only nuisance variables that become confounding factors undermine the logic of the study.

Confounding factors- It is a nuisance variable that differs systematically between the two groups/levels of the IV.

How to deal with nuisance variables to make good experimental designs: ALWAYS keep the nuisance variables constant e.g. sampling people of the same age only.

Important tips of all the tests for the unit:

-One Sample t-test and the Binomial test are used to describe a single variable (metric and categorical variable respectively).

- Paired Samples t-test and the Independent t-test are used to explore the relationship between 2 variables (i.e. one metric and one categorical).

* Paired samples t-test is for repeated measures, while the independent samples t-test is for independent groups

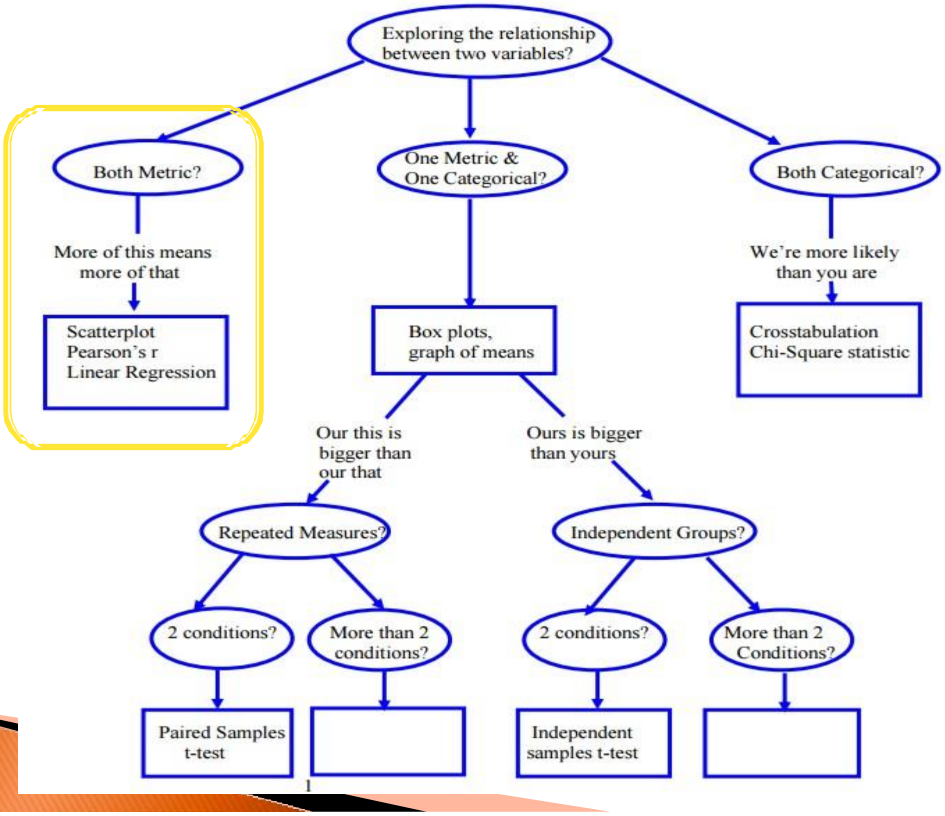
- Crosstabulation and Chi-Square statistic are also used in exploring the relationship between 2 variables (both categorical variables).

STA10003 Foundations of Statistics: Guide for formulas and symbols in a digital exam

If you want show that you are using a symbol or a formula in your answer, you don't need to use the exact symbols that you would hand write. Just express yourself as clearly as you can so that the marker can understand what you are doing. Here is a table with some common symbols and formulae and how you might type them in your answers.

$\mu = 54$	mu = 54 or
	mean = 54

$\sigma = 5$	sigma = 5 or standard deviation = 5 or std dev = 5
$r^2 = 23.3$	r2 = 23.3 or r^2 = 23.3 or r squared = 23.3 coefficient of determination = 23.3 or coeff of deter = 23.3
□ = .34	rho = .34
χ^2	Chi squared or chi^2
($\bar{X}$ = 10.80, σ = 2.30)	(x bar = 10.80, sigma = 2.30) or (xbar = 10.80, s = 2.30) or (mean = 10.80, std dev = 2.30)



Exploring the relationship between two variables?

Both Metric?

More of this means more of that

Scatterplot
Pearson's r
Linear Regression

One Metric & One Categorical?

Box plots, graph of means

Our this is bigger than our that

Repeated Measures?

2 conditions?

Paired Samples t-test

More than 2 conditions?

Ours is bigger than yours

Independent Groups?

2 conditions?

Independent samples t-test

More than 2 Conditions?

Both Categorical?

We're more likely than you are

Crosstabulation
Chi-Square statistic

Correlations Notes

-In Correlations, the reports are pretty similar to other reports, except the slope, which should also be reported.

- A formal report must have a **conclusion**, which starts in one of the following 3 ways:

*As expected

*Contrary to the expectations

*There was insufficient evidence

- ***When making the conclusion of a report, one should take into account:***

*the result of the statistical test

*whether it agrees or disagrees with the hypothesis

The 95% CI for Correlations is stated as, for example: The 95% confidence interval for Person's correlation indicates that the relationship is between rho (*p*) (indicate the lower boundary value) and rho (*p*) (indicate the upper boundary value).

Important report format:

Remember the report should address:

*Research question or hypothesis

*Description of the sample, including descriptive statistics

*Test used, and results of the test

*Where appropriate, the interpretation of the slope (this is something that is only in Correlation reports)

*Interpretation of the 95% CI

*Conclusion which addresses the hypothesis

Correlation Strength, direction, and form

If r is:

*0.8 or above, report as a STRONG correlation

*0.5 to 0.7 as a MODERATE

*0.3 to 0.4 as a WEAK

*below 0.3 is EXTREMELY WEAK

- It can be negative or positive correlation (shown by sign - if negative).

- It can also be linear or non-linear

Those three elements are important i.e. strength, direction, and form

CROSS Tabulation and Chi-Square

Dependent variable (DV) is the variable that is affected if the other variable is manipulated.

Independent variable (IV) is the variable whose manipulation causes change in DV.

A relationship/test/statistic is said to be significant if the p value is less than .05 (if the CI is 95%). The test is not significant if it is more than .05.

Significance (p value): The p value is said to be significant if it is less than.05 or equal to .05. Any p value more than.05 is not significant.

NB: If the *p* value is .000, then it is reported as *p*<.001, BUT NOT *p*=.000.

- When reading the p value for a Chi-square, only take the value that is on the same line where the Pearson Chi-Square is reported (i.e. the column written Asymp. Sig. 2-sided).

- df means the Degrees of Freedom. You report the value given in the df column.

REMEMBER:

- If the p value is significant (less than .05), then there is a relationship between the variables
- If the p value is more than .05, then the p value is not significant AND there is INSUFFICIENT EVIDENCE TO CONCLUDE that there is a relationship between the variables.
- If the value is significant, but the hypothesis is contradicted, then you conclude that THERE IS NO RELATIONSHIP

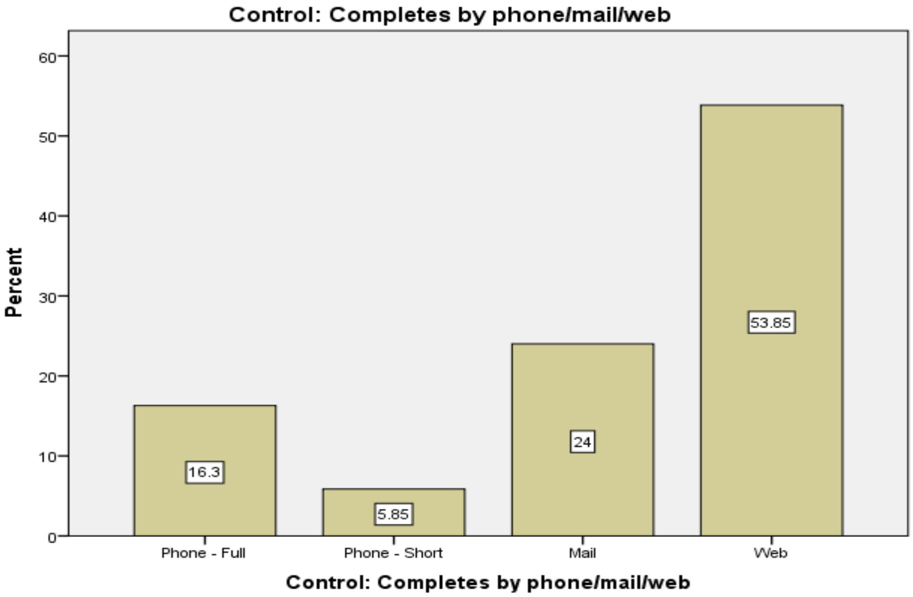
The above tips under (REMEMBER) apply for ALL tests.

Statistics Exam Practice Questions 2015

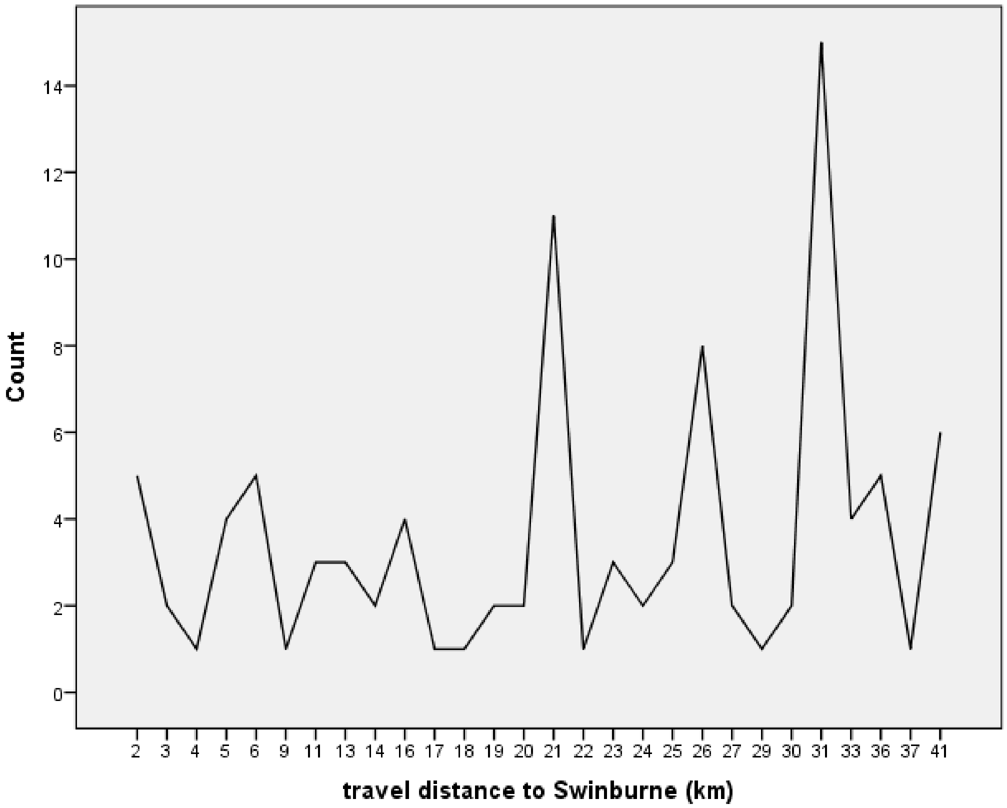
1. Different types of graphs

The SPSS program generates different types of graphs, including bar graphs, line graphs, box plots, scatter plots, pie charts, and histograms.

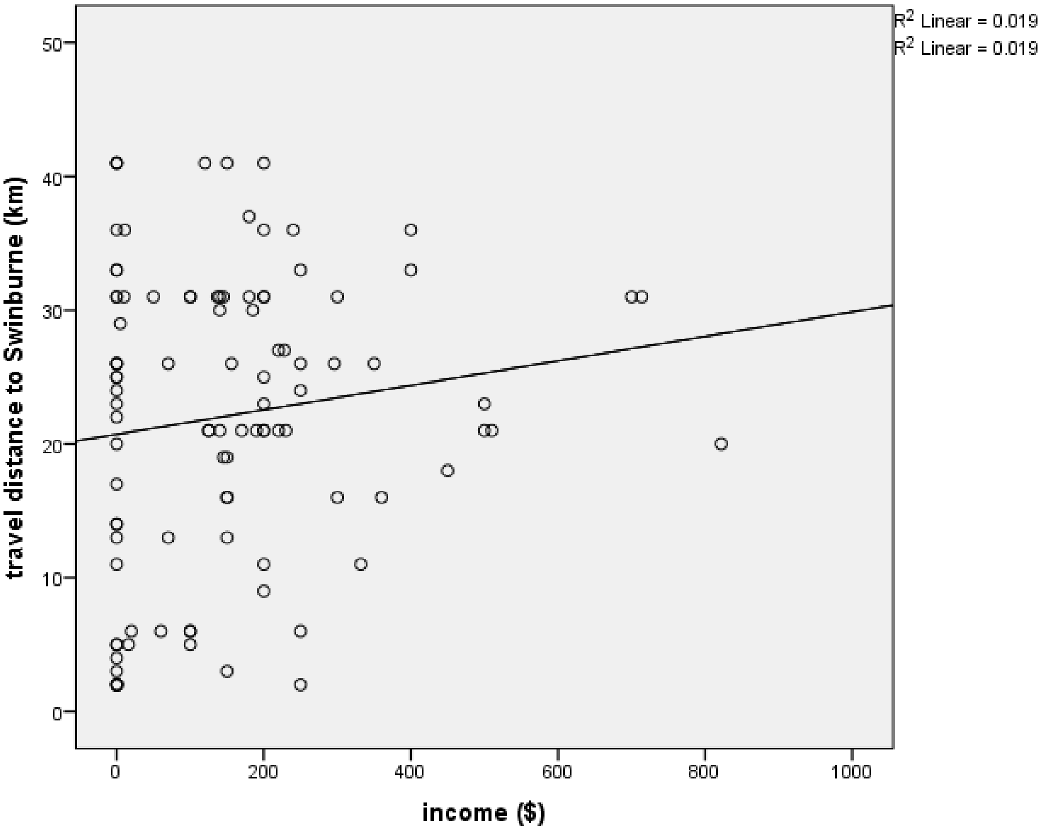
The bar graphs help in displaying the frequency of variables that are nominal. Below is an example of a bar graph:



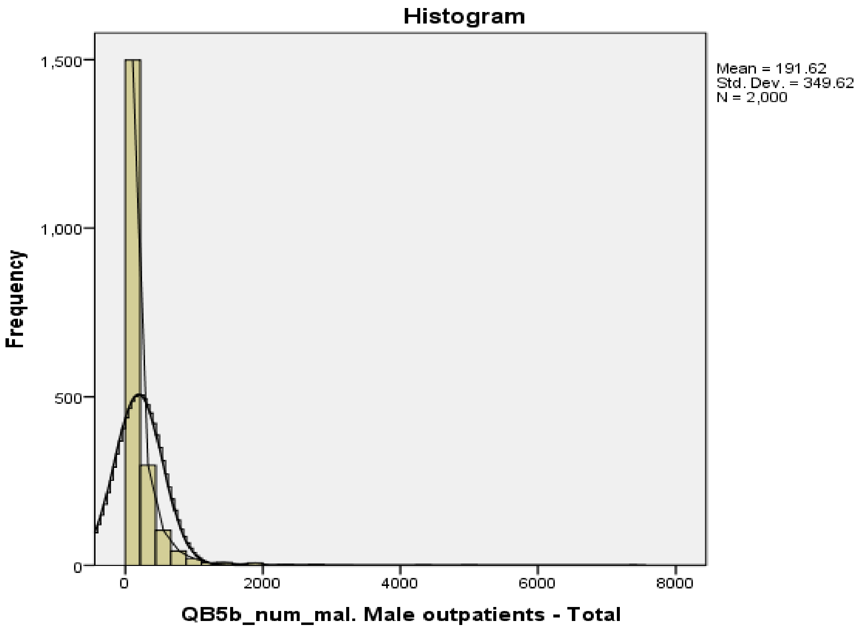
Line graphs serve the same purpose as the bar graphs, but line graphs are more appropriate for displaying continuous variable.


Count
0
2
4
6
8
10
12
14
2
3
4
5
6
9
11
13
14
16
17
18
19
20
21
22
23
24
25
26
27
29
30
31
33
36
37
41
travel distance to Swinburne (km)

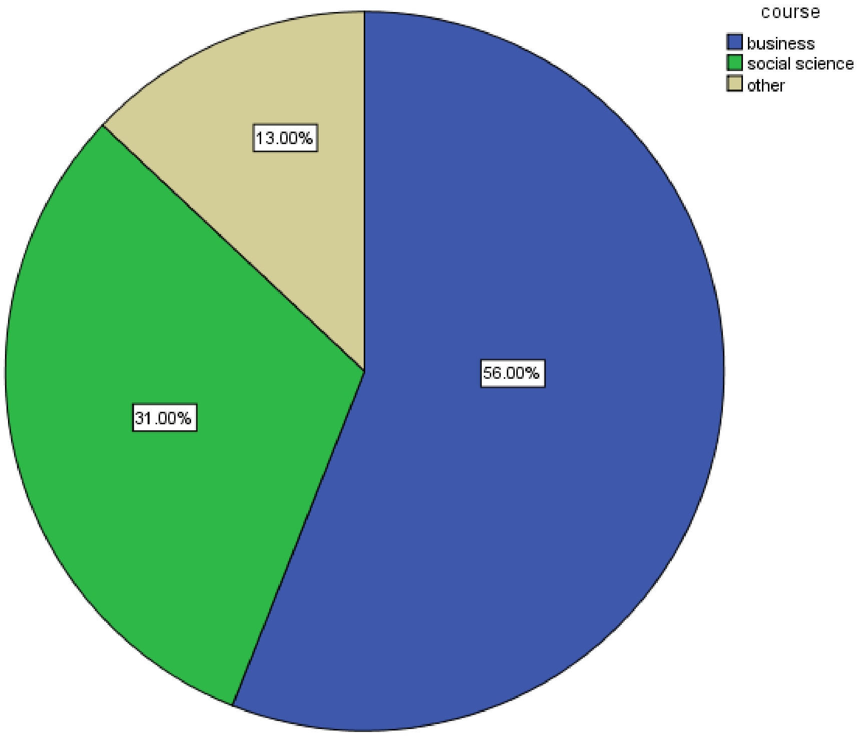
Scatter plots are useful in displaying the relationship between two variables along an X and Y axis. Scatter plots are commonly used in correlation and regression analyses. Below is an example of a scatter plot.



Histograms are important in displaying the frequency of observations. An example of a histogram is shown in the figure below:

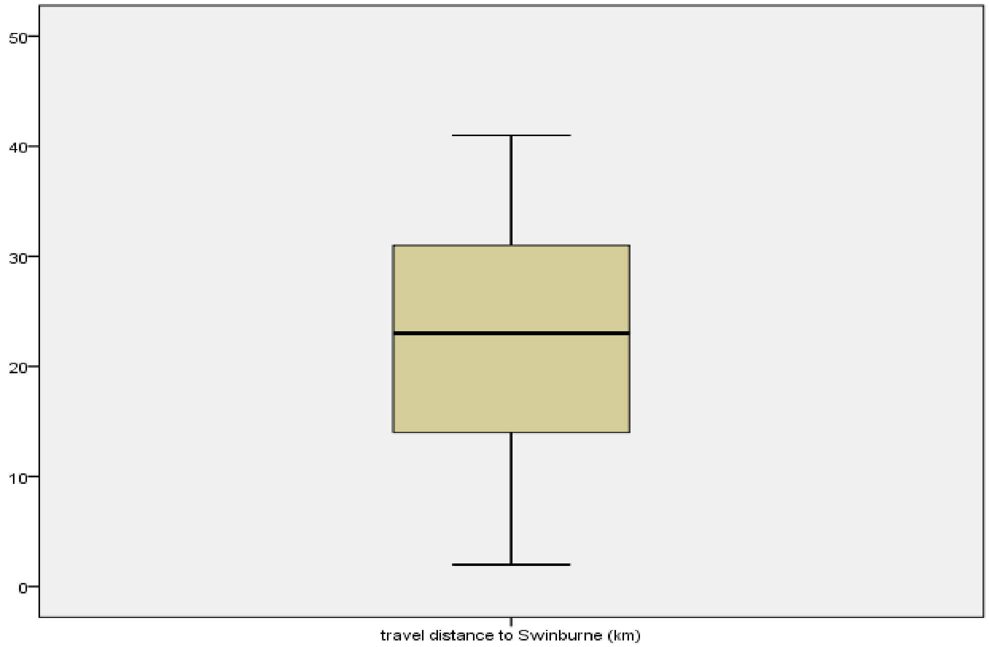


A pie chart shows the share of each categorical variable as shown in the example below:



2. Box plots

Box plots help in showing the distribution of the data in terms of percentiles (25th, 50th, and 75th percentiles). They also show the range (minimum and maximum) of values, beyond which values are considered to be outliers.



3. The different types of statistical tests

There are numerous statistical tests that can be conducted using SPSS. The choice of the test is determined by the type of variables to be analyzed. The tests include the Binomial test, which is a non-parametric test that tests a hypothesis by using two categorical variables. The one-sample t-test is used to test a hypothesis by examining whether means of the same sample differ significantly. The third common test is the independent samples t-test, which determines whether means of different samples differ significantly. The Chi-square test is another statistical test that is commonly used in determining whether a relationship exists between two variables that are categorical. The one-way analysis of variance (ANOVA) test checks whether the means of dependent variables against an independent variable. The paired samples t-test is another statistical test that determines whether the means of two related samples with normal distribution differ significantly. A correlation test determines the relationship between variables that have normal distribution. Other tests include regression (simple linear, and multiple, and logistic regression).

4. How to distinguish between dependent and independent variables

A dependent variable is a variable that is influenced by other factors. In other words, a dependent variable changes if another factor is adjusted. It is also known as the outcome variable because the manipulation of the independent variable causes an effect on the dependent variable.

An independent variable, on the other hand, is a standalone variable, meaning that other variables do not influence it. It is also commonly referred to as the experimental variable or the predictor variable. Most tests that determine the relationship between variable involve determining whether an independent variable causes a statistically significant change in the dependent variable. Therefore, most tests involve the manipulation of the independent variable to cause an effect on the dependent variable (Johnson & Kuby, 2008). For instance, in a study that hypothesized that babies prefer the color red over the color green, the independent variable is color while the dependent variable is the babies' preference. Varying the color of influences whether the child likes the color or not.

In summary, the independent variable influences the dependent variable to change. On the contrary, the dependent variable cannot lead to a change in the independent variable.

5. What does sampling actually mean (sampling theory)?

A sample is an exact number of objects obtained from the population under study. The theory of sampling is mainly based on the concept of random sampling, whereby the study subjects are selected from the population in a way that allows each object in the population an equal chance of being selected. Resultantly, one obtains a sample that is a true representation of the population.

The idea behind sampling is to obtain information that is ideally representative of the population under study. A subset of the population is selected for study because obtaining data from the whole population and cumbersome and costly.

The two main terms used in sampling theory are "statistic" and "parameter", both of which are numerical values. A statistic is a value that changes with samples e.g. sample mean and standard deviation. On the other hand, a parameter is a fixed value that is associated with a given population e.g. population mean (Johnson & Kuby, 2008).

What I needed to know for 2015 Exam

- Population of interest:
- How to determine which type of test to use, given a scenario
- How to identify the population of interest
- How to interpret SPSS output
- How to interpret sampling distribution

- Using the Z-score formula
- Types of study - experimental or observational, and their differences
- How to write a research hypothesis
- Interpret correlation and coefficient of determination
- Experimental designs - Independent Groups, Matched Pairs, Repeated Measures,
  Interpreting confidence intervals
- Identifying bias, nuisance variables, confounding factors
- What a significance level actually means
- Report writing for the different types of tests we have done this unit
- I would definitely recommend having example of how to write data report on your cheat sheet, something about sampling distribution and SPSS output