Feedback for Assignment 01 of 2015

Question 1

The term 'inference' in psychological research refers to - - - - -

1. describing information in a precise way
2. making a prediction or generalization based on existing information
3. the procedures for making a construct visible so that a measurement can be made
4. the development of a hypothesis as a relationship among variables

➔ ***Answer:*** Option 2 is correct. An inference is when you use limited information to predict the state of affairs in a larger context, or when you use your specific information to say something about a more general state of affairs. Option 1 is not valid. Note that Option 3 refers to operationalization and Option 4 to the process of setting up a formal statistical hypothesis.

Question 2

Which of the options below provides the best description of the main purpose of quantitative research in psychology? Its purpose is to - - - - -.

1. develop theories that explain the relationships among observed aspects of human behaviour and mental processes
2. develop predictions about human behaviour of which we can be applied with absolute certainty
3. describe and classify aspects of humans and human behaviour
4. develop hypotheses about relationships that may exist among various constructs

➔ ***Answer:*** Option 1 is correct. The ultimate aim of research is to explain the phenomena we see, and this is done by way of theories. These may be used in prediction but not with 'absolute certainty' as claimed in Option 2. Research begins with description and classification (Option 3) but this is not its ultimate purpose. Hypotheses are developed in order to test the probability of certain explanations (as stated in Option 4), but this is a step in the research process, not its purpose.

Question 3

Empirical knowledge is defined as - - - - -.

1. theories which explain why facts appear as they are observed to be
2. knowledge based on creative insights
3. direct experiences as they are caught in the moment of conscious awareness
4. information derived from careful observation and description of objects and events

➔ ***Answer:*** Option 4 is correct.

Empirical knowledge is the knowledge that you have obtained because you observed what is happening or you have the personal experience of the event. Option 1 is wrong because empirical knowledge derive from experience rather than theory. Option 2 and 3 also wrong because neither of these fit to the definition of 'empirical' knowledge.

Question 4

Operationalizing a construct means to - - - - -

1. find an explanation for the construct to explain why it appears as it is
2. make an educated guess on how it relates to other constructs
3. determine the correct level at which it should be measured
4. devise a systematic procedure to make the construct observable, in such a way that we can measure it.

➔ ***Answer:*** The correct answer is option 4.

A construct is some event or entity which the researcher regards as being of importance in an explanation of the phenomenon under investigation. In quantitative research, where data are treated in numeric form, the problem is how to measure it, and the procedure used to achieve this is referred to as operationalization. For example, one may suppose that persons with concentration problems are affected by anxiety. To research this, some way will have to be found to measure each of the constructs 'ability to concentrate' and 'anxiety', perhaps by developing suitable tests or observation methods.

Option 1 refers to the process of developing theories and Option 2 to the process of developing hypotheses. Option 3 is fairly ambivalent but is not part of a definition of the process called 'operationalizing'

Question 5

A theory is - - - - -.

1. an explanation of observed facts and relationships among them
2. an inspired guess or supposition about the relationships among constructs
3. a careful description of the facts based on observations of objects or events
4. an explanation of human behaviour based on literature

➔ ***Answer:*** Option 1 is correct.

The word 'theory' is used in scientific text to indicate an explanation of why facts are as they are, or how they are connected (see Section 1.2.2 in the PYC3704 Guide).

Option 2 is wrong because this is actually a definition of a hypothesis. A hypothesis is a guess used to explain something, whereas a theory will be based on the result of testing one or more hypotheses and establishing their validity. Option 3 is wrong because while all scientific research should begin with a careful description of phenomena based on observation to establish the facts (and this may include careful measurements), a theory goes beyond this to explain why things are as they seem to be, or how and why they are connected in certain ways. Theories are reported in the literature so that people can read about it, but the theory as such is based on research, not on literature, so Option 4 is wrong.

Question 6

"The mental age of child number one is eight years". In this statement "mental age" is a(n) - - - - -, whereas "eight years" is a(n) - - - - -

1. variable; specific value of that variable
2. construct; variable
3. independent variable; dependant variable
4. hidden variable; descriptive statistic

➔ ***Answer:*** Option 1 is correct.

Option 1 gives the only set of answers that make sense. 'Mental age' could be a variable (a measurement of a construct which can take a variety of values) with 'eight years' one possible instance of it (a specific measurement).

While 'mental age' could refer to a construct as implied in Option 2, 'eight years' cannot be a variable, so this option is not valid. The terms "mental age" and "eight years" also clearly do not refer to different types of variable, so Option 3 is irrelevant. "Eight years" is not a type of descriptive statistic, which invalidates Option 4.

Question 7

A measurement that summarizes an aspect of a population is called a - - - - - while a measurement that describes the same aspect of a sample is called - - - - -

1. construct; variable
2. parameter; statistic
3. statistic; parameter
4. variable; construct

➔ ***Answer:*** The correct choice would be Option 2.

This option makes use of the definitions of 'parameter' (for a population) and 'statistic' for a sample. See Section 1.4.3 of the PYC3704 Guide (especially the part that begins on the bottom of p. 12). The other options are incorrect.

Question 8

A - - - - - is a speculative statement about the relationship among - - - - -, based on observations or expectations

1. theory; constructs
2. hypothesis; statistics
3. theory; variables
4. hypothesis; constructs

➔ ***Answer:*** Option 4 is correct.

A theory is an explanation for the relationships observed among constructs (or variables) which has been tested and is taken to be true, not a 'speculative statement', so both options 1 and 3 are false. Such a speculative statement is usually referred to as a hypothesis, so either Option 2 or 3 can be true. A hypothesis does not however refer to relationships among 'statistics', as this term refers to specific values (such as means or standard deviations) which are used to summarize some overall characteristics of a variable. Therefore Option 2 can be eliminated.

Note that if there was an option 'hypothesis' and 'variables', this could also be correct. A variable is just a construct which has been measured or can be regarded as measurable, and the two words are often used as synonyms in statistics books, where it is assumed that data are quantified.

Question 9

Rose is interested in studying creativity among fine arts students. She randomly selects 100 students who are attending a school for fine arts and tests each of them on a creativity test. In this example, the entire collection of fine arts students which Rose is interested in is referred to as the - - - - -, while the 100 cases which she requires to perform the creativity test are the - - - - -,

1. population; sample
2. range; cases
3. sample; population
4. data; parameters

➔ ***Answer:*** Option 1 is correct. The population is the general group of people who are being studied, and the sample is a smaller group of participants who are selected to represent the population. (See section 1.4.3 of the PYC3704 Guide).

Question 10

Suppose the height of military recruits is distributed normally with a mean of 1750 mm and a standard deviation of 50 mm. Drawing repeated samples of 25 recruits each we expect the standard deviation of the sample means to be about - - - - - mm

1. 2
2. 10
3. 50
4. 25

➔ ***Answer:*** The correct answer is option 2.
If repeated samples of recruits are drawn, the mean for each of these samples can be calculated. The standard deviation of the sample means is the way in which these calculated means vary across the repeated samples. It is an indication of how accurately these sample means reflect the true population mean. This measurement of the variability of the sample means is referred to as the *standard error* of the mean (see pp. 58 – 62 in the PYC3704 Guide). It can be estimated from the population standard deviation and the size of the samples using the formula on p. 61 of the Guide, as follows:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{25}} = \frac{50}{5} = 10$$

Question 11

Which statement best represents an application of the law of large numbers? If I flip a coin 1000 times it will fall heads up - - - - - 500 times.

1. approximately
2. exactly
3. at least
4. either much more or much less than

➔ ***Answer:*** Option 1 is correct.

The law of large numbers states that the larger the sample, the closer the observed outcome is to the actual state of affairs that can be expected in the sample. A coin will fall heads up about half of the time and the more trails there are, the closer to this number it would be, so one would expect around 500/1000 to be 'heads'. This is always an approximation, slightly above or below the real outcome, which is why 'exactly' or 'at least (Options 2 and 3) do not apply. Option 4 expresses exactly the opposite of this law.

Question 12

Suppose that over the years 10 000 students wrote the examinations in PYC3704 and that 6000 of them passed, of which 300 obtained exactly 50%. This means that for randomly selected students the probability of obtaining exactly 50% is - - - - - while the probability of obtaining 50% or more is - - - - -

1. 0.50; 0.03
2. 0.03; 0.60
3. 0.60; 0.03
4. 0.30; 0.60

➔ ***Answer:*** Option 2 is correct.

The probabilities can be estimated from the frequency data (see the discussion in Section 2.1.1.2 starting on p. 30 of the PYC3704 Guide, where this is referred to as the 'relative frequency approach'). You use this approach when you do not have complete data regarding the sample space (that is to say, information regarding the distribution of all possible outcomes) but only some sample data to base your calculations on. The two calculations which are needed to answer this question are as follows:

300 out of 10000 students got exactly 50%, and 300/10000 = 3/100 = 0.03

6000 out of 10000 students passed (getting 50% or more) and 6000/10000 = 6/10 = 0.60

Question 13

The scale along the horizontal axis of the *standard normal distribution* indicates - - - - -

1. probabilities
2. the mean of the distribution
3. the number of standard deviations below and above the mean
4. the p-values

➔ ***Answer:*** The correct answer is given in option 3.

The standard normal distribution (also called the **z**-distribution) is the distribution which is obtained when a normally distributed variable is transformed into a z-value, which is a variable with a mean of zero and a standard deviation of 1. The horizontal axis of this distribution therefore has a mean of zero, and is divided into units of 1 standard deviation each to the left and the right of the mid-point (i.e. zero) of the distribution. See figure 2.7 in the PYC3704 Study Guide.

Note that p-values are probability values, and these are indicated by areas under the curve of the graph; not on the horizontal but on the vertical axis, so Options 1 and 2 are both wrong. The mean of the z-distribution is always zero, but this is a point in the middle of the horizontal axis, not the scale of the entire horizontal axis, so Option 2 is not correct.

Question 14

A researcher want to check that a pair of dice are fair (i.e., evenly balanced). What would the probability be of rolling two sixes with two throws of a pair of fair dice?

1. 0.167
2. 0.333
3. 0.667
4. 0.028

➔ ***Answer:*** Option 4 is correct.

The table below shows all possible outcomes when two six-sided die are thrown.

1; 1	1; 2	1; 3	1; 4	1; 5	1; 6
2; 1	2; 2	2; 3	2; 4	2; 5	2; 6
3; 1	3; 2	3; 3	3; 4	3; 5	3; 6
4; 1	4; 2	4; 3	4; 4	4; 5	4; 6
5; 1	5; 2	5; 3	5; 4	5; 5	5; 6
6; 1	6; 2	6; 3	6; 4	6; 5	**6; 6**

There are 6 X 6 = 36 possible outcomes, with a probability of 1/36 for each specific outcome. There is only one case where both outcomes are 6 (in the bottom right hand cell) so the probability of this is 1/36. And 1/36 = 0.02777778, which can be rounded to 0.028

Note that another way to get the result is to make use of the multiplicative rule (see p. 35 in the PYC3704 Guide), since we are interested in probability both dies getting 6:

p(die number 1 is 6) <u>**and**</u> p(die number 2 is 6) = 1/6 X 1/6 = 1/36, which comes to the same result as the one calculated above.

The other options are all wrong.

Question 15

What is the principal advantage of z scores? They enable one to - - - - -.

1. determine whether scores are normally distributed around the mean
2. transform a person's scores on tests with different means and the same standard deviations into comparable percentages
3. compare a person's scores on tests with different means and standard deviations
4. determine frequency distributions for tests with different means

➔ ***Answer:*** Option 3 is correct.

Transforming two scores into z-scores transform them into a scale with a common mean (zero) and a common standard deviation (one), which makes it possible to compare scores which originally may represent measurements on different scales. The z-score will however still have the same overall distribution as an original score (which is why Option 1 is false) and is therefore not relevant to the determination of frequency distribution (which invalidates Option 4), as the distribution does not change. The scores are not transformed into percentages, so Option 2 is also not true.

Question 16

A bowl contains 6 red, 5 green, 8 blue and 3 yellow marbles. If a single marble is chosen at random from the jar, what is the probability that the marble would be green OR yellow?

1. 0.363
2. 0.031
3. 0.533
4. 0.045

➔ ***Answer:*** Option 1 is correct

Since these are mutually exclusive possibilities, we can apply the additive rule directly (see section 2.1.4.1, starting on p. 34 of the PYC3704 Guide).

$$p(green\ OR\ yellow) = p(green) + p(yellow)$$

$$= \frac{number\ of\ green\ marbles}{total\ number\ of\ marbles} + \frac{number\ of\ yellow\ marbles}{total\ number\ of\ marbles}$$

$$= \frac{5}{22} + \frac{3}{22} = \frac{5+3}{22} = \frac{8}{22} = 0.363$$

The other options are incorrect.

Question 17

Suppose we have stated H_0: $\mu = 10$, and H_1: $\mu < 10$, and find that the sample mean corresponds to a z-score of -3. This means that the corresponding p-value - - - - -

1. need not be found to reach a decision
2. is 0.0026
3. is 0.0013
4. is 0.9987

➔ ***Answer:*** Option 3 is correct.

The z-tables (Appendix D in the Guide) should be used to find the p-value for z = –3, which refers to the area in the tail of the distribution, to the far left side of a standard normal distribution (therefore it will be indicated as the 'smaller portion' in the tables). According to the tables, this p-value = 0.0013. (When using the tables, keep in mind that the distribution is symmetrical around the mean of zero, so the right hand side of z = 3 would be the same as the left-hand side of z = - 3; see the examples on p. 163 in the PYC37-4 Guide).

Question 18

The *standard error* is a measurement of - - - - -.

1. how well a sample mean approximates a population mean
2. the extent to which a variable varies around its mean
3. the extent to which one variable changes as another one changes
4. the size of the error being made when you fail to reject a null hypothesis which is actually false

➔ ***Answer:*** The correct definition of a standard error is given in Option 1. If one takes repeated samples and calculates the mean of each, these calculated means will vary around the 'true' mean (the mean of all the sample means) and the standard error shows the extent of this variance. (This is explained in Sections 2.4.1 and 2.4.2 of the PYC3704 Guide. See also the feedback to Question 10, above).

Option 2 refers to a variance or standard deviation, Option 3 refers to a correlation coefficient and Option 4 has to do with an error of Type II.

Use the scenario below to answer Questions 19 and 20

An educational psychologist measures the general knowledge of 100 learners on a questionnaire with 50 questions. She finds the mean and standard deviation of the scores are 20 and 8 respectively for the sample of data.

Question 19

What is the z-score corresponding to a test score of 14 on the general knowledge test?

1. 1.33
2. 0.75
3. – 0.75
4. – 1.33

➔ ***Answer:*** Option 3 gives the correct answer.

It is necessary to transform the general knowledge score of 14 to its equivalent value terms of z-score. Use the formula on p. 55 in the PYC3704 Guide, but make use of the sample mean and standard deviation (as indicated in the last paragraph on this page in the Guide).

The calculation is as follows:

$$z = \frac{x - \bar{x}}{s} = \frac{14 - 20}{8} = \frac{-6}{8} = -0.75$$

Question 20

What is the probability that a specific learner will obtain a score of 14 or less on the questionnaire in the scenario above?

1. 0.2266
2. 0.7734
3. 0.0918
4. 0.9082

➔ ***Answer:*** Option 1 is correct.

The relevant z-value was calculated in the previous question as z = -0.75. A score of equal to or smaller than 14 is therefore equivalent to one equal to or smaller than -0.75 in the z-distribution. This probability can be found in the z-tables (Appendix D in the PYC3704 Guide). The probability will fall in the section in the far left of the distribution (marked as the 'Smaller Portion' in the tables since it is the section in the one tail of the distribution), and according to the tables it is 0.2266. This could be written in symbolic form as $p(x<14) = p(z<-0.75) = 0.2266$.

Question 21

The size of the level of significance depends on - - - - -.

1. a choice made by the researcher
2. calculating it from the appropriate formula
3. the size of the test statistic
3. the p-value under H_0

➔ ***Answer:*** The correct answer is option 1.
The researcher chooses the greatest risk of making an error of Type I (rejecting the null hypothesis in error) that she is willing to make, and compares the p-value with this to see whether this p-value is lower than the chosen level of significance (α). Values of 0.01 and 0.05 are often chosen but this is a convention.

Question 22

What does it mean to say "the difference between the means of groups A and B is statistically significant"?

1. It is unlikely that the alternative hypothesis will be true
2. The sample result is more probable under the alternative hypothesis
3. The null hypothesis explains the sample result
4. The alternative hypothesis should be rejected

➔ ***Answer:*** Option 2 is correct.

A *statistically significant* result implies that the null hypothesis can be rejected. That is to say, given the results calculated from the sample it is probable that the alternative hypothesis is correct. The other options all imply that the result was not statistically significant and consequently that the null hypothesis was not rejected.

Question 23

Which of the hypotheses below is an example of a directional hypothesis which would require a one-tailed test? It is hypothesised that - - - - - for children of six to ten years.

1. there will be a significant difference in vocabulary between those spending less than ten hours a week playing, and those spending more than ten hours a week
2. as the amount of hours spent playing with peers increases, so will general vocabulary increase
3. the amount of hours spent playing with peers per week will be related to their vocabulary
4. their vocabulary will not be related to the amount of hours spent playing per week with peers

➔ ***Answer:*** The correct answer is given in Option 2.

Option 2 specifies a positive correlation but not a negative one, so it is directional, and this implies that a one-tailed test is required. (See p. 137 in the PYC3704 Guide.) Option 1 refers to a difference without specifying the direction. Options 3 and 4 refer to a relationship but do not specify the direction of the relationship, because they do not give information about the type of relationship; i.e. whether a positive or a negative correlation between the variables is implied.

Question 24

Complete the following sentence. A null hypothesis is - - - - -.

1. the assumption that there is no relationship or difference between parameters of the variables you are testing
2. the assumption that there is a relationship or difference between parameters of the variables you are testing
3. based on the assumption that a significant result is unlikely
4. the specification of the expected relationship between parameters of the variables you are testing.

➔ ***Answer:*** Option 1 is correct.

A null hypothesis is the hypothesis that there is no (statistically significant) effect; in other words, the hypothesis that that the one variable does not affect the other variable. An example of an effect would be difference between means or relationships between variables that are greater than chance. (See pp. 73 – 74 in the PYC3704 Guide).

Option 2 refers to an alternative hypothesis. There is always a null hypothesis, so your expectation regarding the outcome (as implied in Option 3) is irrelevant. Option 4 (' the specification of the expected relationship') can refer to any hypothesis being tested, but this is usually the alternative hypothesis.

Question 25

It is hypothesised that higher levels of depression are related to higher levels of anxiety. Considering this research hypothesis, what would the null hypothesis be?

1. There is no relationship between depression and anxiety and any observed relationship is the result of chance.
2. Individuals with lower levels of depression will have higher levels of anxiety while any result otherwise observed is the product of chance.
3. Individuals with higher levels of depression will have lower levels of anxiety while any relationships otherwise observed are the result of chance or measurement error.
4. There will not be a significant difference between those individuals who score high on depression in comparison to those individuals who score high on anxiety. Any observed difference is the result of chance alone.

➔ ***Answer:*** Option 1 is correct.

As explained in the feedback to the previous question, a null hypothesis implies that there is no relationship between variables. Both Option 2 and Option 3 imply some kind of relationship. Option 4 refers to difference between individuals, not relationships among their scores, so it is poorly formulated. In any case, the data would be examined for a relationship between level of depression and level of anxiety overall, not just high scores in one variable compared with high scores on the other.

Question 26

A p-value generally tells us the - - - - -.

1. probability of obtaining the pattern of results due to common variance if there is in fact no relationship between our variables in the population.
2. likelihood of our obtaining the pattern of results due to sampling error if there is a population.
3. likelihood of our obtaining the pattern of results due to sampling error if there is in fact no relationship between our variables in the population.
4. maximum likelihood of our obtaining a significant result due to a relationship observed between the variables in the population.

➔ ***Answer:*** Option 3 is correct.

The p-value in fact gives the probability that the results we calculate on the basis of our data from a sample would be obtained if the null hypothesis is actually true (i.e. if in reality there is no effect). This is why we would like it to be small before we would consider rejecting the null hypothesis. Any relationships among variables we see in the sample data can in such a case be attributed to sampling error. (See p. 81 in the PYC3704 Guide).

Question 27

How can p<.05 be interpreted? Select the correct option below.

(a) There is a less than 1 in 20 probability of the result occurring by chance alone if the null hypothesis were true.
(b) The probability of obtaining the data if the null hypothesis were true is less than 5%.
(c) There is a 5% chance of making a type one error if the null hypothesis is rejected.

1. Only (a) and (b) are correct
2. Only (c) is correct
3 Only (b) and (c) are correct
4. Statements (a), (b) and (c) are all correct.

➔ ***Answer:*** The correct choice would be Option 4, as all three statements are valid.

Note that a 1 in 20 probability is 1/20 = 0.05, and this is equal to 5%.

Note also that (b) and (c) are just different ways of interpreting a p-value (see p. 81 and p. 85 in the PYC3704 Guide).

Question 28

The relative importance of a relationship which may be found between variables, irrespective of the size of the sample, can be determined by finding the - - - - - for the statistical test.

1. power
2. effect size
3. level of significance
4. standard error

➔ ***Answer:*** Option 2 is correct.

The question refers to the *effect size,* which is usually calculated to get an idea of how large a statistical effect is, irrespective of the sample size. This is often a useful measure to calculate because statistical tests become more sensitive the larger the sample is (see pp. 86 – 88 in the PYC3704 Guide).

Power (Option 1) is the overall sensitivity of the test; the extent to which it is able to pick up significant effects, but this is related to sample size (see p. 85 in the Guide).

Level of significance (Option 3) refers to the maximum risk that one is willing to take to reject H_0 in error (see p. 83 in the Guide).

Standard error (Option 4) refers to the extent to which the mean of a number of samples from the same population is expected to vary across repeated samples. (See p. 61 in the Guide as well as the feedback given above to Question 10).

Question 29

Which of the following is a description of an error of Type I?

1. A type one error refers to a non-significant result which is obtained due to the effect of sampling error alone.
2. A type one error refers to a significant result which is obtained due to the effect of sampling error alone when the null hypothesis is actually true.
3. A type one error is when the null hypothesis is not rejected due to random factors when the alternative hypothesis is actually true.
4. A type one error occurs when the null hypothesis is not rejected even though the p-value exceeds the level of significance

➔ ***Answer:*** Option 2 is the correct answer.

A Type I error occurs when the null hypothesis is rejected when in fact it should not be rejected. This eliminates Options 3 and 4 which both state that the null hypothesis is *not* rejected (Option 3 is in fact an error of Type II). Keep in mind that we base our calculations on sample data to make conclusions on the probable state of affairs in the population, so it is always possible that the conclusion is invalid. The point of statistical tests is to make the probability of this happening reasonably low.

An invalid rejection of the null hypothesis can happen when the statistical test, calculated from data obtained in a particular sample, seems to show a significant effect because the sample does not provide a good reflection of the information in the population, due to problems like sample bias or measurement error. The researcher therefore comes to the wrong conclusion, and rejects the null hypothesis even though the null hypothesis should actually not be rejected for the particular population. Type I errors is linked to the rejection of the null hypothesis (in error) and a non-significant result would not lead to a rejection of the null hypothesis (whether in error or not), so Option 1 is not a valid definition of this error.

Question 30

Which of the following should you include in your report to indicate that the significance test results are not misleading?

(a) The sample size(s).
(b) The exact p-value as it was calculated.
(c) The effect size.
(d) The value of the calculated test statistic.

1. (b) and (c) are sufficient
2. (a), (b) and (d) are sufficient
3. (b) and (d) are sufficient
4. All of (a), (b), (c) and (d) should be reported

➔ ***Answer:*** Option 4 is correct. All of the information in the statements (a), (b), (c) and (d) is of importance in making sense of a significance test.

Feedback to Assignment 02 for semester 2

Question 1

Why do we calculate a test statistic?

1. To determine whether or not we can accept that the null hypothesis is true
2. To determine how far the observed measurements deviate from what we may expect by chance
3. To get a measurement by which we can calculate the level of significance
4. To determine whether or not we can reject the alternative hypothesis

➔ ***Answer:*** Option 2 is correct.

Calculating the test statistic is the first step in a process of comparing the observed data with what may be expected by chance (that is, if in fact the null hypothesis is true and any effects observed in the sample of data are due to random errors).

Option 1 is not really appropriate because the emphasis is wrong here. The test statistic is calculated to determine whether the effect is large enough to reject the null hypothesis and not to try to accept it. For the same reason, Option 4 is not really correct (rejecting the alternative hypothesis would follow only as a consequence of the null hypothesis not being rejected). The level of significance is not calculated but chosen, so Option 3 is false.

Question 2

A researcher wants to test the hypothesis that the mean depression score on a depression scale for patients diagnosed with clinical depression is greater than 120. The statistical hypothesis to be tested is:

H_0: $\mu = 120$
H_1: $\mu > 120$

She uses a random sample of n=64 drawn from the population of diagnosed patients and finds that $\overline{x} = 127$ and s = 24. Which of the values below is the closest to the correct value of $s_{\overline{x}}$?

1. 0.37
2. 3.0
3. 0.61
4. $s_{\overline{x}}$ cannot be calculated from the information that was provided

➔ ***Answer:*** Option 2 gives the correct answer.

This question refers to the standard error of the mean, which is a measurement of how well a sample mean approximates a population mean (see p. 61 in the PYC3704 Guide).

This can be calculated from the data which is provided by substituting the (unknown) population parameters with measurements from the sample (see p. 105 in the Guide where a similar substitution is done), as follows:

$$s_x = \frac{s}{\sqrt{n}} = \frac{24}{\sqrt{64}} = \frac{24}{8} = 3$$

Question 3

Suppose the alternative hypothesis states that $\mu > 60$. The researcher should test H_0 against H_1 if the - - - - -.

1. sample mean is larger than 60
2. sample mean is smaller than 60
3. sample mean differs from 60, irrespective of the direction of the difference
4. p-value is smaller than the level of significance

➔ ***Answer:*** Option 1 is correct.

This is a one-tailed test (see p. 81 in the PYC3704 Guide) where only sample means of larger than 60 are relevant. If the sample mean is calculated and a value of $\overline{x} \leq 60$ is found, the null hypothesis can obviously not be rejected, since there in zero probability that a value of 60 or smaller is also larger than 60. (See p. 77 of the Guide for the discussion of a situation where the observed value of a sample statistic lies in the wrong direction).

Question 4

When applying a t-test for the difference between the means of two independent samples, the probability of obtaining the calculated t-statistic under the null hypothesis is compared to the - - - - - to reach a decision.

1. level of significance
2. degrees of freedom
3. two-tailed probability
4. effect size

➔ ***Answer:*** The correct answer is Option 1.

'The probability of obtaining the calculated t-statistic under the null hypothesis' refers to the p-value, and this calculated p-value is to be compared to α, the level of significance chosen as a cut-off point by the researcher. If the p-value < α, it implies that there is a low probability that any difference between the means which is observed in the sample data is due only to random factors. If this is the case, the null hypothesis can be rejected in favour of the alternative hypothesis.

Question 5

A social psychologist wants to test how long people will wait before responding to cries of help from an unknown person. The psychologist wants to confirm his suspicion that people will take less time to react when they hear a female voice than when they hear a male voice. He tests this on a sample of n=15 people who are told (one at a time) to sit in a waiting room to be called for an interview. While they wait, each participant hears a call for help from a male or female voice, which is actually a recording. The dependent variable is the number of seconds that each participant waits until they go to investigate or tried to find help.

The sample following sample statistics are calculated from the results:

Male voice: $\overline{x}_1$ = 11.9 seconds; s_1 = 3.5

Female voice: $\overline{x}_2$ = 15.3 seconds; s_2 = 4.1

Given these findings, what type of statistical test will the psychologist have to do to confirm the relevant statistical hypothesis?

1. A one-tailed statistical test
2. A two-tailed statistical test
3. A test for independent samples
4. No statistical test is necessary

➔ ***Answer:*** Option 4 is correct.

The hypothesis that it will take less time to react when people hear a female voice than when they hear a male voice can be stated as follows:

H_0: $\bar{x}_1 = \bar{x}_2$ and H_1: $\bar{x}_2 < \bar{x}_1$

The sample results however show that people take *less* time when the voice is male than when it is female; i.e. $\bar{x}_2 = 15.3 > \bar{x}_1 = 11.9$. There is no probability that this finding can confirm the alternative hypothesis, so no further statistical testing is really necessary. (See p. 77 of the PYC3704 Guide as well as the feedback on Question 3, above).

Question 6

A researcher wants to test the following hypotheses:

$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

On the basis of data provided, the output from a computer program indicates that a t-value of $t = 1.72$ was found, with the p-value for a two-tailed test given as $p = 0.056$. What should the researcher do to evaluate this result at a level of significance of $\alpha = 0.05$?

1. Divide the p-value by 2 before comparing it with α
2. Multiply the p-value by 2 before comparing it with α
3. Divide α by 2 before comparing p to α
4. Compare the p-value as given with α

➔ ***Answer:*** The correct answer is given by Option 4.

The test is non-directional or two-tailed, so the two-tailed p-value should be used without any adjustment. It is only if the test was one-tailed that an adjustment to the given (two-tailed) p-value should be made (see p. 81 in the PYC3704 Guide).

Question 7

A matched-pair t test should be used when you are - - - - -.

1. testing a two-tailed hypothesis
2. comparing means on a measurement from before and after a specific event
3. comparing two variables which come from the same group
4. comparing two means on a variable where the data were drawn from the same population

➔ ***Answer:*** Option 2 is correct.

A measurement from before and after a specific event (for example a therapeutic intervention) is an example of matched-pairs test for dependent variables (see pp. 117 – 118 in the PYC3704 Guide). The type of hypothesis being tested (one- or two-tailed) is not relevant to the used of matched pairs, so Option 1 is false. The reference to 'two variables which come from the same group' is not very clear, but this could imply measurements on totally different ranges of values in which case a matched-pair t test could not be used, so Option 3 is not applicable. Two means from samples from the same population) need not be dependent or matched, which eliminates Option 4.

Question 8

Which of the symbolic expressions below are appropriate for symbolizing a population variance?

1. σ
2 σ^2
3. s
4. s^2

➔ ***Answer:*** Option 2 is correct.

By convention, a population standard deviation is indicated by the symbol σ and the variance would be the square of this. Options 3 and 4 are wrong because they refer to the sample standard deviation and variance, respectively (see p. 160 in the PYC3704 Guide).

Base your answers to Questions 9 and 10 on the following scenario:

To test the efficacy of a workshop aimed at improving people's interpersonal skills, a researcher applies a scale which rates the interpersonal skills of 20 participants before and after they participate in the workshop. Scores on his rating scale among the general population have a mean of 5 and a standard deviation of 1.5.

Question 9

Which of the following is the most appropriate way to express the null hypothesis for an analysis of the results?

1. $H_0: \mu = 5$
2. $H_0: \mu_1 = \mu_2$
3. $H_0: \bar{D} = 0$ (where $\bar{D}$ is the population mean of the differences scores)
4. $H_0: \mu_1 \neq \mu_2$

➔ ***Answer:*** Option 3 provides the most appropriate way to express the hypothesis.

Testing before and after an event or intervention implies that a matched pairs design with dependent measurements is used, and Option 3 is the conventional way to express this (see PYC4704 Guide, pp. 117-120).

Option 2 could actually also be considered as correct, since it so happens that $\bar{D} = \mu_2 - \mu_1 = 0$, which actually implies $\mu_1 = \mu_2$. Because of this possibility of confusion, we will delete this item from the assignment when we calculate the assignment mark.

Question 10

Which is the appropriate test statistic to calculate to test the hypothesis in the previous question?

1. The **z**-statistic for the mean of a sample
2. The t-statistic for the difference between the means of two dependent samples
3. The t-statistic for the difference between the means of two independent samples
4. The t-statistic for the mean of a single sample

➔ ***Answer:*** Of the options given, the best would be Option 2.

These are two repeated tests on the same sample of participants, which is a matched pair design (as mentioned in the feedback to Question 7, above). Therefor a t-test for two dependent samples is to be used, and a independent sample test (Option 3) would not be appropriate

Options 1 and 4 are wrong because two sets of measurements are compared, which is to be regarded as two samples of (matched-pairs) data, even if it refers to the same population.

Question 11

In which circumstances can the **z** test for comparing two independent means NOT be used?

1. The population parameters are available to a researcher.
2. The population standard deviations for the two groups are unknown.
3. The population means for the two groups are unknown.
4. The sample standard deviations for the two groups are unknown

➔ ***Answer:*** Option 2 is correct.

For a **z**-test, the population standard deviations for the any means being compared should be known (see Section 4.1.2 beginning on p. 100 in the PYC3704 Guide). This rule of requiring the population standard deviation when using a **z**-test can be extended to the comparison of two sample means also. See also p. 114 in the PYC3704 Guide, where the case when these parameters are unknown is discussed for independent samples.

If the population parameters are known as stated in Option 1, this would include the population standard deviations, so there is a **z**-test which could be used, and this option is therefore false. However, it is the *population standard deviations* that are required, so whether or not the population *means* are known (Option 3) is irrelevant. The null hypothesis will assume that they must be equal, but their actual sizes are not important, since it is the probability that they differ significantly which is being investigated. It can be assumed that the sample standard deviations will be known when means are to be compared, since these can be calculated from the sample data of each group, so Option 4 is not valid.

Question 12

When studying correlations in research, one investigates the relation between - - - - -

1. the mean of a single sample of subjects and a population mean
2. two dependent groups of subjects, with respect to a single variable
3. two variables measured on the same group of subjects
4. two independent groups of subjects, with respect to a single variable

➔ ***Answer:*** Option 3 is correct. A correlation coefficient compares the relationship between two variables in a single group. An example would be the relationship between anxiety and depression in a sample of persons undergoing therapy, with each participant being measured on both of these variables.

The means are not relevant to correlation, only the way in which the sets of measurements vary together or fail to do so, so Option 1 is incorrect. There must be at least two variables being compared, so Option 2 is not really appropriate. If the two groups are independent, it follows that the measurements of a particular variable cannot be paired off or matched at all, so Option 4 is not valid.

Question 13

A positive correlation between variables X and Y implies that persons scoring low on X will generally score - - - - - on Y.

1. high
2. low
3. either high or low
4. in an indeterminate way

➔ ***Answer:*** Option 2 is correct.

A positive correlation implies the variables vary together: a low score on one variable is likely to produce a low score on the other, and a high score on one variable would imply a high score on the other. Option 1 implies a high *negative* correlation was found (with a low score on one variable implying a high score on the other). Both Options 3 and 4 imply there is no correlation at all.

Question 14

A number of psychiatric patients are classified by gender (male or female) and into one of four categories as schizophrenic, severely depressed, bipolar disorder and others. Which of the following is suitable for representing counts or frequencies of persons which fall into each possible subcategory?

1. A contingency table
2. A scatter plot
3. A histogram
4. A spreadsheet

➔ ***Answer:*** Option 1 is the correct answer.

Both these measurements (gender and diagnosis category) are *nominal scale* measurements (see Appendix B in the PYC3704 Guide), so the best way to cross classify the data would be by way of a contingence table (see p. 142 in the Guide). A scatterplot (Option 1) cannot be used since a more powerful level of measurement would be required and a histogram (Option 3) refers to the distribution of a continuous variable of stronger than nominal scale. A spreadsheet is just a two-way table with raw data and is not used to represent frequency counts, but rather the raw data of each case, where the frequency counts would come from.

Question 15

Two samples may be regarded as independent when - - - - -.

1. there is no systematic relationship between the composition of one sample and the other
2. they were drawn at different occasions
3. each measurement in one sample can be matched with a measurement in the other sample
4. they are both totally random

➔ ***Answer:*** Option 1 is correct.

Data that were drawn from repeated measurements of the same persons at different occasions would be dependent, so Option 2 is not necessarily wrong. Option 3 would be true of *dependent* samples. Whether the samples are random (Option 4) is not really relevant to their independence.

Question 16

A researcher wants to establish whether a relationship exists between people's religious affiliation and whether they are in favour of or against the death penalty (yes or no). Which of the following would be the most appropriate test to use?

1. The t-test for two independent samples
2. The chi-square (χ^2) test statistic
3. Pearson's correlation test statistic
4. The t-test for two dependent samples

➔ ***Answer:*** Option 2 is correct.

Two nominal-scale measurements are compared. Information on 'religious affiliation' can be classified into groups or categories, but it is not really a measurement (where one person can score more or less than another person). 'Opinion on the death penalty' is to be answered as 'yes' or 'no', so it is also a nominal scale measurement. The chi-square test is the only one of the options which could be used.

Question 17

The lower we set the level of significance, the greater the probability of - - - - -

1. rejecting the null hypothesis
2. a type I error
3. a type II error
4. a significant result

➔ ***Answer:*** Option 3 is correct.

Keep in mind that a Type I error is the error of rejecting the null hypothesis when it is in fact true (and therefore should not be rejected), and this is controlled for by setting a low level of significance (α) in advance. The lower we set this level of significance, the *less* the chance of making an error of Type I, so option 2 is not correct. This also makes the chance of rejecting the null hypothesis and obtaining a significant result *smaller*, so options 1 and 4 are also false.

The probability of a Type II error (β) works in the opposite direction: it refers to the case where the null hypothesis is not rejected when it is false and should in fact be rejected. This is not deliberately controlled in advance by the researcher. However, the lower α is set (that is, the lower the probability of a type I error), the greater the probability of a type II error (β). You could reduce the chances of an error of Type I greatly by choosing an extremely small value for α, but that would increase the chances of *not* rejecting H_0 *when in fact you should reject it*. (See section 3.3.2 on p. 84 in the Guide for PYC3704).

Question 18

To test the efficacy of psychotherapy aimed at relieving depression, a researcher applies a depression scale to 50 depressed patients at the start and again at the end of their treatment, predicting that the latter scores will be lower (reflecting less depression). Scores on his depression scale among the general population have a mean of 30 and a standard deviation of 10. Which research design is appropriate to test the research hypothesis?

1. A two-sample groups design with independent groups
2. A two-sample groups design with dependent groups
3. A one-sample groups design
4. A design where the correlation between two variables is tested

➔ ***Answer:*** Option 2 is correct.

This is a repeated measures design, with the same people being measured twice, before and after an event. From the point of view of formal research design, this is to be regarded as two samples with highly dependent measurements. (See section 5.1.1 on p. 112 and Study Unit 5.2, beginning on p. 117 in the Guide for PYC3704).

Question 19

Suppose you compare a group mean with a particular population mean by using a t-test, and you find that the t-test statistic calculated for your research results is zero. Which conclusion is appropriate?

1. The p-value will be extremely small.
2. The null hypothesis is likely to be true.
3. The alternative hypothesis is likely to be true.
4. The null hypothesis can probably be rejected.

➔ ***Answer:*** The correct alternative is option 2.

Look at the formula for the t-test for independent samples (given on p. 114 of the Guide):

$$t_c = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The only time that the value of t_c can be zero, is when there is no difference between the means (i.e., when $_1 - \bar{x}_2 = 0$), which would imply that the null hypothesis must be true. A similar argument works for the other forms of the t-test as well: if the t-test statistic comes to zero, the means being compared do not differ, so the null hypothesis is supported.

Question 20

Suppose you calculate the correlation between two variables, X and Y, and you find a very high correlation of greater than 0.9. Why cannot you infer from this that one variable is the cause of the other one?

1. The direction of the effect of the one variable on the other is not known.
2. The direction of the causation can work in both ways.
3. You first have to establish which variable is dependent on which
4. The effect may be the consequence of hidden variables affecting both X and Y

➔ ***Answer:*** The correct answer is Option 4.

Correlation just tells you that as one variable becomes greater, so does the other, but you cannot see the reasons for this. There may be one or more hidden variables that affect both of the variables that you are observing in indirect ways. Note that while the issue of dependence (which variable depends on which) is relevant when we suspect causality, causality is not necessarily implied by it, so the answer in option 3 is not sufficient. (See Section 6.1.4.2 on p. 140 of the Guide).

Question 21

A contingency table is used to summarise the relationship between two variables measured on a(n) - - - - - scale.

1. nominal
2. ordinal
3. interval
4. ratio

➔ ***Answer:*** Option 1 is the correct answer.

Contingency tables are used to represent frequency counts of data that have been classified in terms of 2 nominal variables (for example, gender and occupational category). It is possible to fit ordinal, interval or ratio scale measurements into such a table, but they would first have to be transformed into a classification system; that is, the data have to be treated as if they represent nominal scale measurements. (See section 6.2.2 on p. 142 of the Guide for PYC3704, and also Appendix B, where measurement levels are discussed).

Question 22

In a scatter diagram, the more tightly clustered the data points are around a straight line, the - - - - - the correlation is between the two variables.

1. lower
2. higher
3. closer to zero
4. closer to 1

➔ ***Answer:*** The answer we were looking for is given in Option 2.

A high correlation implies that the points on a scatter plot will fall in a straight line. A Pearson's correlation coefficient, conventionally indicated by r, is a measurement of how closely a plot of two sets of data is to a straight line that represents the best fit (this is referred to as the regression line). A high correlation can however be either positive or negative, leading to a value of r that is close to either +1 or -1 respectively, so option 4 is not always true. A value of close to –1 is actually an indication of a high negative correlation. It is the *absolute value* of r (that is, the value of $|r|$) which indicates whether it is a high correlation or not. (See Appendix E in the Guide to PYC3704 for the use of '| |' to indicate an absolute value).

Question 23

A variable that can take only one of two possible values is called - - - - -

1. binomial
2. dichotomous
3. nominal
4. categorical

➔ ***Answer:*** Option 2 is correct.

The word 'dichotomous' means 'divided into two'. The word 'binomial' (Option 1) implies something that has two names (in the case of the *binomial distribution*, it refers to a probability event which has two possible outcomes; see p. 41 in the PYC3704 Guide). Dichotomous variables are in fact also nominal and categorical (Options 3 and 4), but the term 'dichotomous' refers specifically to the fact that there are only two categories, which is what makes Option 2 the most appropriate answer. (See Appendix B on p. 156).

Question 24

Employees of a large organisation can be classified into one of three groups, technical, clerical and managerial workers. Based on a survey, these workers are divided into workers that are satisfied with their working conditions, and those that are dissatisfied. A researcher wants to determine whether there are differences between the different types of worker as far as their satisfaction is concerned. Which would be the most suitable statistical test to do?

1. The test statistic based on the correlation of the type of work with the level of satisfaction.
2. Use a t-test to determine for independent samples whether differences exist between those that are satisfied and those that are not.
3. Use the chi-square test to see if the distribution of levels of satisfaction differs for the different job category
4. Use a t-test to determine for independent samples whether differences exist between those that are satisfied and those that are not.

➔ ***Answer:*** Option 3 specifies an appropriate test because two nominal-scale or categorical variables are being compared. To correlate variables (for the r-test statistic) or to work out means (for a t test) would require a stronger level of measurement, so options 1, 2 or 4 are not appropriate.

Base your answers to Questions 25 to 28 on the following scenario:

A group of middle managers in a company are required to undergo a training programme aimed at improving their leadership abilities. A researcher wants to establish the effectiveness of the training programme. She tests each participant before the training commences on a questionnaire that measures the leadership ability of each participant as well as on their score on a scale that measures self-esteem. After the training programme is completed, each participant is again tested for their leadership ability. The table below shows the results of the measurements.

Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Gender	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
SelfEst	5	4	3	1	3	4	5	3	2	2	3	3	4	2	2	3	3
Lead1	5	4	3	3	4	6	5	4	4	4	5	4	5	1	2	3	6
Lead2	6	6	3	4	6	7	4	5	4	6	7	6	5	3	4	5	5

Key:
*The variable **Gender** is coded with 1 = Male; 2 = Female.*
***SelfEst** is a variable indicating self-esteem and was measured on a 5-point scale (ranging from 1='low self-esteem' to 5='high self-esteem').*
*Leadership ability was measured twice on a 7-point scale (with 1='very low leadership ability' up to 7='very high leadership ability'). **Lead1** indicates the measurement before the training programme, while **Lead2** represents the measurement thereafter.*

Question 25

The researcher wants to establish whether a significant difference exists between male and female participants in the level of self-esteem of the participants before the training begins. He decides to do statistical significance testing to a level of significance of $\alpha = 0.05$.

Which of the following is the most appropriate way to express the hypotheses to be tested?

1. $H_0: \mu_M = \mu_F$ $\quad H_1: \mu_M > \mu_F$
2. $H_0: \bar{X}_M = \bar{X}_F$ $\quad H_1: \bar{X}_M \neq \bar{X}_F$
3. $H_0: \rho = 0$ $\quad H_1: \rho \neq 0$
4. $H_0: \mu_M = \mu_F$ $\quad H_1: \mu_M \neq \mu_F$

➔ ***Answer:*** Option 4 is correct.

Two means are being compared (for the male and female groups), but since the question does not indicate which mean is expected to be greater than which, the alternative hypothesis would be a two-tailed or non-directional hypothesis (see p. 75 of the PYC3704 Guide). This eliminates option 1. In the formal statistical hypotheses, the population parameters are specified, which eliminates option 2. One would not use a correlation coefficient to compare two different groups of persons, which makes Option 3 untrue.

Question 26

Assuming that the data is normally distributed and the variances do not differ significantly, which of the following tests are appropriate to test the hypotheses in the previous question?

1. Pearson's correlation statistic (r)
2. The t_d test for two dependent groups
3. The t_c test for two independent groups
4. The chi-square (χ^2) test for two variables

➔ ***Answer:*** Option 3 is correct.

The t-test is required to compare two group means (which eliminates Options 1 and 4). In this case the two samples come from independent groups, because there is no reason to suppose a systematic relationship between the two groups (male and female). This eliminates Option 2. Note that the two samples are not even the same size, which would be required in a test for dependent groups, since it should be possible to match each individual measurement in the one group with a particular individual in the other group.

Question 27

The researcher calculates the value of the appropriate test statistic (in Question 26 above). In which of the four intervals below will the absolute value of the test statistic fall (i.e., ignoring a plus or minus sign)?

1. Between 0 to 1.0
2. Between 1.0 and 2.0
3. Between 2.0 and 3.0
4. Above 3.0

➔ ***Answer:*** Option 2 is correct.

To calculate the value of t_C, you will first have to calculate the means and standard deviations for the self-esteem score for each of the two groups (male and female).

You can see from the data in the table that the number of males in the sample is n_M=9 (counting the number of people with 'Gender' coded as '1') and the number of females is n_F = 8 (counting the number of people with 'Gender' coded as '2').

If you use the formulas for means and standard deviations for the male and female research participants respectively (see Appendix C in the Guide), you should find the following:

The calculations for mean and standard deviation for the group of 9 males are as follows:

$$\bar{x}_M = \frac{1}{n}\sum x = \frac{5+4+3+1+3+4+5+3+2}{9} = \frac{30}{9} = 3.3333$$

$$s_M = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{(5-3.3333)^2+(4-3.3333)^2+\cdots+(2-3.3333)^2}{9-1}}$$

$$= \sqrt{\frac{1.6667^2+0.6667^2+\cdots-1.3333^2}{8}} = \sqrt{\frac{2.7779+0.4445+\cdots 1.7777}{8}}$$

$$= \sqrt{\frac{14.0000}{8}} = \sqrt{1.7500} = 1.3229$$

Note that we use the sample standard deviation, not the one for populations (as explained in Appendix C in the Guide). This is because we presume that this problem refers to a sample of middle managers in the company who represents the population of all middle managers.

For the female group, we find the following: $\overline{x}_F$ = 2.7500 and s_F = 0.7071

We leave the calculation of these values to you.

To calculate the test statistic, we have to substitute these values into the formula for the t-test for independent samples (t_c as given on p. 115 of the Guide):

$$t_C = \frac{(\bar{x}_M - \bar{x}_F)}{\sqrt{\frac{s_M^2}{n_M}+\frac{s_F^2}{n_F}}} = \frac{(3.3333-2.7500)}{\sqrt{\frac{1.3229^2}{9}+\frac{0.7071^2}{8}}} = \frac{0.5833}{\sqrt{\frac{1.7501}{9}+\frac{0.5000}{8}}}$$

$$= \frac{0.5833}{\sqrt{0.1944+0.0625}} = \frac{0.5833}{\sqrt{0.2569}} = \frac{0.5833}{0.5068}$$

$$= 1.1509$$

Your result should be similar (there may be a small deviation because of rounding-off errors). The answer therefore falls within the interval specified by option 2 above.

Note that whether the sign of the calculated t-statistic is positive or negative will depend on which mean you subtract from which, so it does not affect the absolute size of the t-statistic. Therefore, since the hypothesis being tested is two-tailed or non-directional (see the answer to Question 25 above), you can ignore the sign when you determine significance (however, this would matter if you were doing one-sided or directional testing, when you have to consider whether the change is in the direction which you specified in your alternative hypothesis).

Question 28

Using appropriate statistical tables, the researcher finds that the test statistic calculated in Question 27 yields a p-value of $p = 0.1839$. What conclusion can be drawn from this? (Use the level of significance as given above in Question 25).

1. There is a significant difference in self-esteem level before the training programme between males and females
2. The self-esteem levels before the training programme do not differ significantly between male and female participants
3. The self-esteem before the training programme of females is significantly greater than that of males
4. The self-esteem before the training programme is significantly greater for males than that for females

➔ ***Answer:*** Option 2 is correct.

The p-value is not smaller than the level of significance (set in Question 25 to $\alpha = 0.05$), so the null hypothesis cannot be rejected. The p-value of 0.1839 actually implies that there is a better than 18% probability that the difference between the two means (2.67 and 3.25) is the result of chance, which, given our chosen level of significance, is too large to justify rejecting the null hypothesis.

Use the following scenario to answer Questions 29 and 30.

Lila wonders whether a relationships exists between a person's length and their leadership ability. She collect data from a sample of 95 people, classifying them as short or tall, and as leaders, followers and those she could not classify. From this, she creates the contingency table below.

Cross-classification	***Tall***	***Short***
Leader	*12*	*32*
Follower	*22*	*14*
Unclassifiable	*9*	*6*

Question 29

Consider the scenario given above. If frequency data is evenly distributed throughout the categories, with no proportional difference between tall and short people as far as leadership abilities go, what should the expected value of the frequency of *short leaders* be? (In other words, the number of people who can be classified as both 'short' and as ' leaders', if these variables have no effect on each other).

1. 15.8
2. 19.9
3. 24.08
4. 32

➔ ***Answer:*** The correct answer is given in option 3.

To answer the question, you need to calculate row and column totals. Expand the table as follows:

Cross-classification	Tall	Short	Row total
Leader	12	32	44
Follower	22	14	36
Unclassifiable	9	6	15
Column Total	43	52	95

To get the expected frequency (the way data would be distributed purely by chance) you need to multiply the total for the second column of frequencies (the total number of short people) with the total of the first row (total number of leaders) and divide by the overall total, using the general formula given on p. 144 of the PYC3704 Guide:

$$e_{ij} = (t_{row} \times t_{col}) / t_{all}$$

Since 'short leaders' refers to the expected frequency in row 1, column 2, we can call this e_{12} and calculate it like this: $e_{12} = (52 \times 44)/95 = 2288/95 = 24.0842 \approx 24.08$

Note that option 4 is incorrect because it gives the *observed* frequency for short leaders, while the question asks for the *expected* frequency, that is, the frequency which one would expect if the frequencies were distributed equally with no interaction between 'length' (or 'shortness') and 'leadership ability. This is what one would expect if the null hypothesis (a lack of interaction) was true. The whole point of the test is to compare the observed frequencies (as seen in the sample data) with what may be expected if the null hypothesis was true.

Question 30

Calculate the value of the appropriate test statistic which Lila would need to consider if she wants to determine whether a relationship exists between length and leadership ability. If we use 'Q' to indicate the result of this calculation, which of the following statements is true?

1. $Q < 0$
2. $0 \leq Q < 4$
3. $4 \leq Q < 8$
4. $Q \geq 8$

➔ ***Answer:*** Option 4 is correct.

To solve the problem, Lila must compare the observed frequencies (as given the scenario) with the frequency which can be expected in each category if the relevant null hypothesis is true (that is, the frequencies which can be expected if there is no interaction between 'length' and 'leadership ability'). The appropriate test statistic is therefore the chi-square (χ^2) test statistic, so this is the test statistic which 'Q' refers to.

The relevant equation or formula is given on p. 141 of the Guide:

$$\chi_p^2 = \sum \frac{(O-E)^2}{E}$$

To use this formula, it is necessary to work out the expected frequencies for each cell in the contingency table in the same way as was required for the case of the Short Leaders (in the previous question).

This would lead to the following contingency table (with expected frequencies added in brackets):

Cross-classification	**Tall**	**Short**	**Row total**
Leader	12 (19.92)	32 (24.08)	44
Follower	22 (16.29)	14 (19.70)	36
Unclassifiable	9 (6.79)	6 (8.21)	15
Column Total	43	52	95

It is useful to create a table like the one on p. 145 of the PYC3704 Guide if you have to do this calculation by hand. In the present case, the table will look like this:

O	E	O-E	$(O-E)^2$	$(O-E)^2/E$
12	19.92	-7.92	62.66	3.15
22	16.29	5.71	32.55	2.00
9	6.79	2.21	4.89	0.72
32	24.08	7.92	62.66	2.60
14	19.70	-5.70	32.55	1.65
6	8.21	-2.21	4.89	0.59
				10.71

The value of the chi-square statistic is the total of the numbers in the right-hand column:

$$\chi_p^2 = 10.71$$

Your own calculated value may differ slightly (due to rounding-off errors) but should be similar to this. (If you use tables or a computer program you will find that this value is significant, which would imply that a significant relationship exists between length and leadership ability).

To find the correct option, Q should be set to equal this calculated statistic, which would imply Q = 10.71. Option 4 is therefore correct because 'Q ≥ 8' means that Q is larger than or equal to 8, which is true.

The other options are all false. Option 1 is impossible, for a chi-square statistic can never be negative (smaller than 0). Option 2, '0 ≤ Q < 4' should be interpreted to mean 'Q is larger than or equal to zero and smaller than 4' which is of course not valid in this case. Option 3 would imply a value of larger or equal than 4 but smaller than 8, which is also false.

Feedback for Assignment 01 of 2016

Question 1

Which of the following are advantages of studying statistical methods in psychology?

1. It leads to a better understanding of the content of psychological theories.
2. It helps us to formulate research questions.
3. It helps us to evaluate whether the patterns we observe in data are due to more than the consequence of chance.
4. It gives insight into the nature of the scientific process.

➔ ***Answer:*** Option 3 is correct.

Statistical methods are required specifically to compare the relationships that we observe in data with what could be expected purely by chance, in order to decide whether the patterns we detect are probably not purely the result of random variations in the data (e.g., due to errors in our measurements). The other three options are more general aspects of research and go beyond the statistical analysis of the results.

Question 2

Which of the options below is a valid description of a 'construct' in psychological research? It is a(n) - - - - -.

1. measurement based on the careful observation of aspects of humans or human behaviour
2. observation of an aspect of humans or human behaviour which was operationalized in some way
3. hypothetical aspect of humans or human behaviour which we wish to investigate
4. explanation of empirical observations based on the measurement of certain variables

➔ ***Answer:*** The correct answer is given in Option 3.

Constructs are abstract aspects of humans such as 'intelligence' or 'depression' which can be best described as 'hypothetical' since we suppose they exist but they cannot be observed directly (see p. 3 of the Study Guide for PYC3704). We use 'operationalization' to make them visible (by developing some way of observing them) but this happens because we hypothesise that they exist in the first place (so Option 2 is wrong). This could entail measurement (Option 1) and may lead to an explanation for human behaviour (Option 4), but the construct is presupposed before any of this can happen.

Question 3

A psychologist believes that personality factors such as ability to get along with other people in a team are likely to have an influence their success in a team sport. To investigate this, she draws a sample of participants in competitive sport, comprising soccer players, hockey players and cricketers from various sports clubs. After evaluating their performance on an appropriate psychometric test, she divides players from the three types of sports into two groups: those who have good social skills and those who tend avoid social contact. She then uses evaluation forms, based on interviews with different sports experts, to assess the actual sports performance of the members of the sample. The dependent variable in the study is - - - - - and the independent variable is - - - - -.

1. measured sports performance; social skills
2. type of sports; social skills
3. social skills; measured sports performance
4. social skills; type of sports

➔ ***Answer:*** Option 1 is correct.

The researcher wants to determine whether sports performance (a measurement of how well a specific sports participant performs) depends on social skills. This implies that the construct 'social skills' is the independent variable which is varied to see how it affects the dependent variable, the measurement of 'sports performance.' The category of sport (soccer, hockey or cricket) gives us information about the population from which the sample was drawn (the types of sports participants that were involved), but in the context of this scenario this variable is not being investigated here. It would be possible to compare the three groups but this would not answer the research question.

Question 4

A psychologist studies the effect that age has on attitude towards Aids. She selects a sample of subjects divided into three different age groups, 20-30 years, 30-40 years and 40-50 years. The three age groups are - - - - -.

1. different variables for measuring 'age'
2. operationalisations of the variable 'age'
3. three different theoretical constructs
4. a way of measuring 'age'

➔ ***Answer:*** Option 4 is correct.

Measurement is a process where numbers are allocated to a construct according to a rule, and that is what happens here. It produces a single new variable (which could be called something like 'Age group'), so Option 1 is wrong. These groups are classifications of age, not the definition of how age should be measured, so Option 2 is wrong. The three values refer to the same underlying construct, to produce a variable, so Option 3 is also incorrect. One does not operationalise variables but operationalise constructs to be able to measure them, and this measurement is then referred to as a variable.

Question 5

The term 'inference' in psychological research refers to - - - - -.

1. describing information in a precise way
2. making a prediction or generalization based on existing information
3. the procedures for making a construct visible so that a measurement can be made
4. the development of a hypothesis as a relationship among variables

➔ ***Answer:*** Option 2 is correct.

The notion of an inference is described as a conclusion which is based on existing information, where the specific information is used to make statements about the type of phenomenon in general. This enables social scientists to make predictions which have a reasonable probability of being true. (See p. 2 in the Guide for PYC3704). Option 1 refers to descriptive statistics and Option 3 refers to operationalization. Such predictions can be used to create new hypotheses, but it is not the process of creating hypotheses as such (as suggested in Option 4).

Question 6

Consider the following statement: "The experience of strong emotion is accompanied by physiological reactions such as an increase in heart rate". This statement can be viewed as a research hypothesis because it - - - - -.

(a) makes a prediction that can be tested by observation
(b) describes a possible relationship between variables

1. (a) but not (b)
2. both (a) and (b)
3. (b) but not (a)
4. neither (a) nor (b)

➔ ***Answer:*** Option 2 is correct.

A psychological hypothesis formulates a testable empirical claim, that is, a prediction which can in principle be observed, and this usually involves postulating a relationship between two or more variables.

Question 7

A psychologist uses a psychometric test to study the intelligence of school children. Intelligence is the - - - - - variable and the psychometric test represents the - - - - - variable in this study.

1. independent; dependent
2 manifest; operational
3. dependent; independent
4. latent; manifest

➔ ***Answer:*** Option 4 is correct.

A hidden or underlying (latent) aspect of a construct has to be made visible (manifest) to be observed. 'Intelligence' is a latent variable because it cannot be observed directly. The results of the psychometric test can however be observed directly and can therefore be regarded as the manifest variable. Note that the test result is the measurement by which intelligence is made visible, these are manifestations of the same construct (or variable), while 'dependent' and 'independent' refer to relationships between two different variables, so Options 1 and 3 are both incorrect. The psychological test is the operation by which the variable 'intelligence' is made visible, so we cannot say that 'intelligence' is in itself 'manifest', as suggested in Option 2.

Question 8

In the process of psychological research, researchers try to - - - - -.

1. prove that a theory is true
2. account for theoretical assumptions
3. obtain empirical support for a theory
4. change hypotheses into theories

➔ ***Answer:*** Option 3 is correct.

Theories cannot really ever be 'proven' in an absolute sense (there is always room for doubt). In research, psychologists are instead trying to find empirical (observational) support for a theory. The research is not really intended to 'account' for the theory (Option 2), but is rather conducted to show that the theory is supported by the evidence. Hypotheses are developed based on observations (e.g. measurements) and on reasoning about the implications of the theory, and occasionally on inspired guesswork, which implies that a theory follows from the validation of a number of interrelated hypotheses. It is however not specific hypotheses that becomes a theory, as implied in Option 4, but rather a number of hypotheses which, if supported by research, leads to the theory being accepted (which is what is implied in Option 3).

Question 9

If a coin is flipped three times, the sample space of possible outcomes is:

1. HHH; TTT; THT; HTH; HHT; TTH; HTH
2. HHH; HHT; HTH; HTT; THH; THT; TTH; TTT
3. HTT; THT; HTH; HHH; TTH; TTT
4. HHH; HTT; HTH; TTT; HTT; THH; HHT; THT

➔ ***Answer:*** Option 2 is correct.

This question requires that you determine the sample space for three flips of a coin, which in turn means that you must determine all the possible outcomes if a coin is flipped for three times. Option 2 gives the correct answer, because this is the only option that lists all the outcomes correctly. Option 1 does not list THH, Option 3 does not list HHT and THH, and Option 4 does not list TTH while HTT is listed twice.

Question 10

An industrial psychologist wants to investigate the levels of assertiveness among different categories of workers in a large company. He draws a sample of 200 workers and then divides them into managers, clerical workers, technical workers and manual workers with the intention of getting each participant to do a test to determine their level of assertiveness. The division of workers into these four groups represents a measurement on a - - - - - level.

1. nominal
2. ordinal
3. interval
4. ratio

➔ ***Answer:*** Option 1 is correct.

When numbers are used to allocate people or objects to categories or groups with no implication of 'intensity' or 'ordering' related to the size of the number, it is referred to as a nominal level or nominal scale of measurement. Note that Option 2 is incorrect because, while one may presume that 'managers' have a higher rank than the rest, and perhaps 'manual workers' can be said to have a low rank, there is no obvious way to decide whether clerical workers have a lesser 'rank' than technical workers or the other way round. It would depend on the specific types of work that they do. A senior clerk can 'outrank' a junior technician, and *vice versa*. (Nominal and other scales are explained in Appendix B of the Guide).

Question 11

Suppose that 5 000 students wrote the examinations in PYC3704 and that 3000 of them passed, of which 200 obtained exactly 50%. This means that for randomly selected students the probability of obtaining exactly 50% is - - - - - while the probability of obtaining 50% or more is - - - - -.

1. 0.60; 0.04
2. 0.07; 0.50
3. 0.04; 0.60
4. 0.60; 0.05

➔ ***Answer:*** Option 3 is correct.

We have to base the calculations of probabilities on the relative frequencies approach here (see pp. 30-31 in the PYC3704 Guide), because we have only observations to work with, and no theoretical model which describes all possible outcomes. Since 200 students out of the 5000 who wrote the exam obtained exactly 50%, we can estimate that the probability of getting this specific mark is 200/5000 = 2/50 = 0.04. The students who passed are the ones who obtained 50 or more, and this occurred in 3000 out of 5000 cases, so the probability of this is 3000/5000 = 3/5 = 0.6.

Question 12

In response to a questionnaire 57 people answered "yes" to a particular question. Of these, 8 persons were male. There were 61 people that answered "no" to the question, and 5 of these were male. If one person is selected randomly from the group, what is the probability that the person answered "yes" ***or*** was male?

1. 0.14
2. 0.525
3. 0.593
4. 0.11

➔ ***Answer:*** Option 3 is correct.

Based on the information given, we know that a total of (57+61) = 118 people answered the questionnaire. Of these people (8+5) = 13 were male. Because we know that 57 people answered 'yes,' the probability of someone answering ‘yes’ in this group is given by 57/118. Likewise since there are 13 males in the group, the probability of someone being male in this group is 13/118. To determine the probability of any particular person drawn randomly from this group being either male or answering yes, we need to use the additive probability formula, *p(A or B) = p(A) + p(B).* (See pp. 34-5 in the PYC3704 Guide).

So in this case we have: p('Male' OR 'Answer Yes') = p('Male') + p('Answer Yes')

= 13/118 + 57/118

= 0.110 + 0.483 = 0.593

Question 13

A teacher is teaching a class about probabilities. She shows the learners a bowl which contains **3 Red** marbles, **5 Blue** marbles and **7 Yellow** marbles. One of the learners, Vusi, is asked by the teacher to select one marble out of the bowl at random and to hand it to her. She shows the class that Vusi chose a **Blue** marble and she demonstrates to the class how to calculate the probability of this outcome. While holding on to the first marble, she asks Vusi to select another marble in the same way. She now asks the class to calculate the probability that ***both*** the first ***and*** the second marble chosen by Vusi would happen to be ***Blue***.

Select the *best* estimate out of the options given below.

1. 0.1333
2. 0.1111
3. 0.6190
4. 0.0952

➔ ***Answer:*** Option 4 is the best choice.

There are altogether 3+5+7 = 15 marbles in the bowl.

Since there are **5 Blue** marbles, it is easy in the first case to apply the classical formula (given on p. 29 of the PYC3704 Guide):

$$p(First\ marble\ is\ BLUE) = \frac{Number\ of\ favourable\ outcomes}{Number\ of\ possible\ outcomes} = \frac{5}{15} = 0.3333$$

However, the outcome in the case of the second marble to be selected by Vusi is not independent of this. When the second marble is to be selected, the class knows that the first marble was blue, and the probability of this outcome can be calculated as above. This is actually an example of a *conditional probability* (see p. 36 in the PYC3704 Guide). Conditional probabilities can be used when additional information that was not previously known is available to improve the estimate.

We know that there are **4 Blue** marbles left out of **14** marbles (15 -1) in total (note that the scenario describes the teacher as "holding on to the first marble"). We can think of this as the probability that 'the Second marble is Blue' *given that* 'the First marble was Blue':

P('Second marble is Blue' | 'First marble is Blue') = 4/14 = 0.2857.

(Note the use of the vertical bar symbol to signify 'given that').

The overall probability that the *First marble was Blue* AND *the Second marble is Blue* can then be calculated as follows, using the multiplicative rule for combining conditional probabilities:

P(*First marble was Blue* AND *Second marble is Blue*)

= P(*First marble is Blue*) X *P(Second marble is Blue | First marble was Blue)*
= 0.3333 X 0.2857 = 0.0952.

Question 14

The **z** score corresponding to a raw score represents the number of - - - - - that the raw score differs from the mean of the raw score distribution.

1. points
2. standard deviations
3. percentiles
4. variances

➔ ***Answer:*** Option 2 is correct.

The **z**-scores are equivalent to standard deviations on the standard normal distribution, and each unit on the **z**-scale represent one standard deviation of the raw score. (See section 2.3.3 in the Guide for PYC3704). This implies that **z**=1 is equivalent to one standard deviation of x removed from the raw score mean (μ for a population or $\overline{x}$ for a sample) and **z**=2 is equivalent to two times the standard deviation of x removed from the mean, and so on.

Use the following summary of marks to answer Questions 15 and 16.

The marks in different subjects of Patrick, a high school pupil, are represented in the table below. His marks for different subjects are given along with the mean and standard deviation for each of these subjects, for all the learners in his class.

Subject	***Patrick's mark***	***Mean of class***	***Standard deviation of class***
Mathematics	56%	42%	6%
Science	54%	54%	4%
Geography	62%	50%	8%
History	68%	75%	5%

Use this information to answer the questions below.

Question 15

In which subject did Patrick do best, relative to his class?

1. Mathematics
2. Science
3. Geography
4. History

➔ ***Answer:*** Option 1 is correct.

The marks should first be converted to **z**-values, to make it possible to compare them across the different means and standard deviations. Using the **z** transformation (on p. 55 of the PYC3704 Guide, substituting population parameter symbols with symbols for sample statistics), you should find that the **z**-values which are equivalent to each of the marks in the table above are as follows:

$$z_M = \frac{x_M - \bar{x}_M}{s_M} = \frac{56 - 42}{6} = \frac{14}{6} = 2.3$$

$$z_S = \frac{x_S - \bar{x}_S}{s_S} = \frac{54 - 54}{4} = \frac{0}{4} = 0$$

$$z_G = \frac{x_G - \bar{x}_G}{s_G} = \frac{62 - 50}{8} = \frac{12}{8} = 1.5$$

$$z_H = \frac{x_H - \bar{x}_H}{s_H} = \frac{68 - 75}{5} = \frac{-7}{5} = -1.4$$

So it is clear that in the case of Mathematics (z_M), Patrick's mark is higher than any of the others, and it is in fact 2.3 standard deviations above the mean. In each of the other subjects his marks transforms into a smaller value relative to the z-distribution.

Question 16

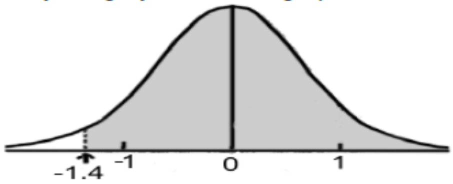
Find the subject where Patrick did *worst* relative to the rest of his class. What is the probability of any other one of Patrick's classmates getting a mark of equal to this or better for this particular subject? Choose the option closest to the correct answer.

1. 0.1
2. 0.3
3. 0.6
4. 0.9

➔ ***Answer:*** The correct answer is Option 4.

From the calculations for Question 15 above, you can see that Patrick's mark of 68% for History translates to z_H= -1.4 on the z distribution, which is lowest in comparison to any of the other scores. The fact that the z value is negative shows that it falls below the class mean.

Now you have to find what the probability is of a random classmate getting a mark which is better or equal to this for History. This is indicated by the grey area in the graph below.



You can see that the probability of getting a mark of 68 or greater is equivalent to the area from the point where z = -1.4 to the far right of this on the graph of the standard normal distribution. If you look this up in the tables for the z-distribution (Appendix D in the PYC3704 Guide), you will see that the probability of getting a mark greater or equal to 68% (which is the same as the probability of getting a value of greater than or equal to z = -1.4) is 0.9192. This can be written as $p(z \geq -0.14) = 0.9192$ and can be rounded off to 0.9.

Note how the graph helps you to see that it would be the 'larger portion' indicated in the tables that is relevant. It is always useful to draw a graph like this if you get a question where you have to look up the probability of a certain range of z-values.

Question 17

The *sampling distribution of means* refers to:

1. the normal distribution of a raw score distributed around the mean
2. the distribution of the different possible values of the sample means together with their respective probabilities of occurrence
3 the distribution of the values of the items in the population
4. the distribution of the means of all possible samples of a particular size randomly selected from the same population

➔ ***Answer:*** Option 4 is correct.

Read the discussion of the distribution of sample means in the PYC3704 Guide (pp. 60 - 61). Option 1 is false because it refers to the raw data, not the distribution of means, and Option 2 is incorrect because we typically do not know what the probability of occurrence of a particular sample mean is. Option 3 is too vague to be useful.

Question 18

The size of the p-value depends on - - - - -

1. a choice made by the researcher
2. conventional rules
3. the null hypothesis
4. the value of the test statistic

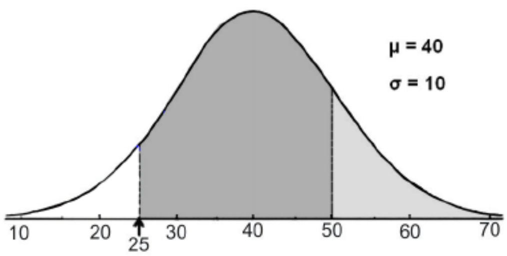
➔ ***Answer:*** Option 4 is correct

The researcher wants to establish that the effect which was observed (for example a relationship among variables in a sample of data) has a very small chance of being purely the result of chance variations in the data. He does this by first calculating a test statistic (such as a t-test or z-test statistic). The p-value (usually read from tables or calculated with the aid of a computer) is the probability that the test statistic would have this particular size if the null hypothesis was true. This is explained in section 3.2.2 in the Guide.

Note that this does *not* imply that option 3 can also be true: the p-value depends on the test statistic, which in turn compares the observed data with the result which could be expected if the null hypothesis was true, but the p-value does not depend on the null hypothesis as such. Options 1 and 2 refer rather to the *level of significance*, which is the *maximum* value of this p-value which the researcher would consider if the null hypothesis is to be rejected.

Question 19

Look at the graph based on the frequency distribution of a measurement of a variable x from normally distributed data which is reproduced below. Using the information provided in the graph, calculate the probability that a measurement made at random on this particular scale will fall in the area under the curve coloured in grey.



Select the answer closest to the calculated probability from the options given below:

1. 0.933
2. 0.775
3. 0.159
4. 0.5

➔ ***Answer:*** Option 2 is correct. [Note that to make this discussion easier to follow, we added a light grey section at the right hand side of the graph above]

We want to determine the probability of a value of x which falls within the dark grey area in the graph, that is, the probability of getting a value of x which falls *above* 25 but *below* 50. To find this, we first have to find the value on the z-distribution which is equivalent to an x-value of 25. In other words, the value of 25 must first be transformed into a z-score. The formula which should be used is given on p. 55 of the Guide (we can presume that we are working with sample data, so the population mean symbol μ can be substituted by the sample mean symbol $\bar{x}$ and the population standard deviation by the sample standard deviation symbol s).

$$z = \frac{x - \bar{x}}{s} = \frac{25 - 40}{10} = -\frac{15}{10} = -1.5$$

This tells us that getting a x-value of 25 or greater is equivalent to getting a z-value of -1.5 or greater. We can now look up the probability of getting this from the tables of the standardized normal distribution (the z distribution) in Appendix D in the Guide.

You should find that the probability of an x-value of 25 and greater is $p(x > 25) = p(z > -1.5) = 0.9332$. Note that you can see from the graph that you would be interested in the *larger portion* as indicated in the table, and that the negative sign can be ignored (since the normal distribution is symmetrical).

Since we are interested in the probability of z falling below 50 only, we have to subtract the area of a probability of a value *greater than* 50 from this result (the area is indicated in light grey in the graph).

To find out what the equivalent z-value is when x = 50, we again make use of the z-transformation formula:

$$z = \frac{x - \bar{x}}{s} = \frac{50 - 40}{10} = \frac{10}{10} = 1.0$$

The probability of getting a z-value of greater than 1 is $p(z > 1) = 0.1587$

Note that you can see from the graph that this time you are interested in the *smaller* region given in the table in Appendix D in the Guide. This probability now has to be subtracted from the probability of an x-value of 25 and greater which we calculated before. So the probability of a x-value of greater than or equal to 25 but less than or equal to 50 is equivalent to the probability of a z-value of greater than or equal to -.1.5 but less than or equal to 1:

$$p(-1.5 \leq z \leq 1) = p(z \geq -1.5) - p(z \geq 1) = 0.9332 - 0.1587$$
$$= 0.7745 \approx 0.775 \text{ (rounded off).}$$

Question 20

The lower we set the level of significance, the greater the probability of - - - - -.

1. rejecting the null hypothesis
2. a Type II error
3. a Type I error
4. accepting the alternative hypothesis

➔ ***Answer:*** Option 2 is correct.

Setting a very small level of significance makes it less likely that the null hypothesis will be rejected in error, but it makes it more probable that a researcher will *not* reject the null hypothesis when she should in fact reject it. This is an error of Type II. (See pages 84 – 86 in the PYC3704 Guide).

Answer Questions 21 to 25 on the basis of the following scenario:

Lebo claims that workers in large companies are less work motivated than workers in small companies. In a previous study involving all large companies in South Africa, it was found that the average workers' motivation score on a work motivation questionnaire (where a higher score indicates a higher level of work motivation) was 50 with a standard deviation of 15. Lebo plans to present the same questionnaire to a sample of workers from small companies to determine whether her suspicion is true.

Question 21

Which kind of hypothesis must be set up by Lebo for testing?

1. A directional hypothesis
2. A nondirectional hypothesis
3. Lebo can set up either a directional or a nondirectional hypothesis, depending on how she chooses to formulate her question.
4. Not enough information is given to decide

➔ ***Answer:*** Option 1 is the correct answer.

The claim is being made that the mean of the motivation score of large companies is less than that of small companies. Or conversely, it is implied that small companies will on average have a work motivation level greater than (>) 50, the known mean motivation level for large companies. This implies a directional or one-sided hypothesis.

Question 22

Which of the sets of statistical hypotheses below reflect the research hypothesis to be tested?

1. $H_0 : \mu = 50$ $\quad H_1 : \mu > \overline{x}$
2. $H_0 : \mu = 50$ $\quad H_1 : \mu \neq 50$
3. $H_0 : \mu = 50$ $\quad H_1 : \mu > 50$
4. $H_0 : \mu = 50$ $\quad H_1 : \mu < 50$

➔ ***Answer:*** The correct Option is 3.

As stated in the answer to the previous question, the hypothesis to be tested (the alternative hypothesis) is that small companies have a mean work motivation level of greater than 50. In other words, the population parameter for small companies (indicated by μ) should be greater than the population mean for large companies, which is given as 50 (it is therefore regarded as a constant value in a one-sample test). This is contrasted with the null hypothesis, which states that the population mean for small companies will not differ significantly from 50.

Note that formal hypotheses are always specified in terms of *population* parameters, even though calculations are made based on statistics computed from data in a *sample*. The purpose of the statistical procedure is to make inferences about populations, and the sample serves as a representation of the state of affairs in the populations.

Question 23

After Lebo collected the responses on the work motivation questionnaire for an appropriate sample of 100 workers from small companies, she calculates the following sample parameters:

Sample size: n = 100
Mean motivation score for small companies: $\overline{x}$ = 55.2
Standard deviation of the motivation score for small companies: s = 12.5

Since the mean and standard deviation for large companies are given (as 50 and 15 respectively), Lebo decides to use a single-sample z-test (i.e. the $z_{\bar{x}}$ test statistic) to test whether the mean motivation scores for small companies differ significantly from the mean motivation scores for large companies.

Which of the following are closest to the calculated value of this statistic?

1. 0.42
2. 41.6
3. 4.16
4. 3.47

➔ ***Answer:*** The correct answer is Option 4.

On p. 80 of the PYC3704 Guide (and again on p. 102), the equation for $z_{\bar{x}}$ test statistic is given as:

$$z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

Here, $\overline{x}$ is the mean we are testing for, $\mu_{\bar{x}}$ is the mean of the distribution of these means, and $\sigma_{\bar{x}}$ is the standard deviation of the distribution of the means (also referred to as the standard error).

We want to test is the mean of the distribution of means differs from the population mean for large companies, which is given as $\mu_{\bar{x}}$ = 50. You have to use the formula for the standard error to calculate $\sigma_{\bar{x}}$ (see p. 61 in the PYC3704 Guide, and pp. 79 - 80 for an example of how it is used in this test):

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1.5$$

Now, from the formula given above, the $z_{\bar{x}}$ test statistic can be calculated as follows:

$$z_{\bar{x}} = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{55.2-50}{1.5} = \frac{5.2}{1.5} = 3.4667$$ (or 3.47 when rounded off to two places).

Question 24

If Lebo tests this hypothesis against a significance level of α=0.01, what can she conclude about the motivation level of small companies?

1. The null hypothesis can be rejected, so it follows that the motivation level is higher in small companies than in larger ones
2. The null hypothesis cannot be rejected, so it follows that the motivation level is no higher in small companies than in larger ones.
3. The null hypothesis can be rejected, so it follows that the motivation level is higher in larger companies than in smaller ones.
4. The alternative hypothesis is not valid, so it follows that the motivation level is higher in small companies than in larger ones.

➔ ***Answer:*** Option 1 is correct.

The probability of a **z**-value of greater than 3.47 can be read off the **z**-tables in Appendix D, and from this it will be found that the p-value = 0.0002 (using an approximate value of **z** = 3.5).

Note that we are interested in the region where **z** > 3.47, which is the far rightmost portion of the area under the standard normal curve, and therefore the 'smaller portion' as indicated in the table. This value of 0.0002 is considerably smaller than the significance level of α=0.01, so the null hypothesis can safely be rejected and the alternative hypothesis is consequently supported. Referring back to the null hypothesis (see Question 22 above), we may conclude that the workers in smaller companies show a significantly higher level of motivation than those in larger companies.

Question 25

Lebo wants to determine whether the difference between the motivation level for small companies and those for large companies is reasonably large, in terms of its practical implication, irrespective of whether it is significant or not.

After making the appropriate calculations, what does she find?

The effective difference between the motivation level for small companies and those for large companies is - - - - -.

1. very small
2. between small and medium
3. between medium and large
4. very large

➔ ***Answer:*** Option 2 is correct.

The important point of the question here is that Lebo is interested in "... the practical implication, irrespective of whether it is significant or not." Significance is related to sample size, and what Lebo requires is some indication of how important her finding is without taking sample size into consideration.

She can do this by calculating the *effect size* of the finding (see pp. 86 – 88 in the PYC3704 Guide).

To do this, she can make use of Cohen's d, as follows:

$$Cohen's\ d = \frac{mean\ difference}{standard\ deviation} = \frac{\bar{x} - \mu}{\sigma} = \frac{55.2 - 50}{15} = \frac{5.2}{15} = 0.347$$

An effect of between 0.2 and 0.5 is between small and medium (see p. 88 in the PYC3704 Guide for the interpretation of effect sizes).

So even though our result is highly significant (fairly high above what we would expect by chance) the effect is not very large in terms of its practical significance. The motivation in smaller companies is greater than in large companies, but not to a very great extent.

Feedback to Assignment 02

Question 1

Why would a researcher calculate a test statistic? It is a calculation - - - -.

1. which determines how far the observed measurements deviate from what we may expect by chance
2. which shows directly whether or not we can accept that the null hypothesis is true
3. to get a measurement by which we can calculate the level of significance
4. to determine whether or not we can reject the alternative hypothesis

➔ ***Answer:*** Option 1 is correct.

Calculating the test statistic is the first step in a process of comparing the observed data with what may be expected by chance (i.e., if the null hypothesis were true).

Option 2 is not really appropriate because the emphasis is wrong here. The test statistic is calculated to see how far the effect which was observed in the data (our measurements) deviate from what we may expect by chance *if the null hypothesis was true*. Not rejecting the null hypothesis would imply the observation is too close to the situation predicted by the null hypothesis for a researcher to feel confident in rejecting the null hypothesis. So even if we were hoping to find no effect (for example, no difference between groups or no relationship among variables), the formal aim of calculating the test statistic is still to determine whether it is safe to reject the null hypothesis or not.

For the same reason, Option 4 is not really correct (accepting the alternative hypothesis would follow only as a consequence of the null hypothesis being rejected).

The level of significance (α) is not calculated but chosen by the researcher as the maximum risk of rejecting the null hypothesis in error that he or she is willing to take, so Option 3 is false.

Question 2

Two samples may be regarded as dependent when - - - - - -.

1. each measurement in one sample is correlated with a measurement in the other sample
2. they were drawn from the same population
3. there is a systematic relationship between the measurements in one sample and the other
4. they are both totally random

➔ ***Answer:*** Option 3 is correct and is a definition of dependent samples, where the measurements form matched pairs.

Option 1 is not really valid as a definition of dependence. Correlations are usually calculated from two measurements (that is, two variables) in the same sample, not from different samples. It is however sometimes possible to correlate measurements from two samples, and finding a correlation would probably imply dependence of some kind. In such a case one first would have to pair off the specific measurements in the one sample with those in the other before you would be able to calculate the correlation coefficient. Examples would be husband and wife pairs, or mother and infant pairs etc., compared on some characteristic (construct) such as ‘empathy.’ This implies that you *must already suspect* that a dependence exists (based on the principle in Option 3) to be able to match pairs of measurements, and the correlation is calculated to measure the *extent* of this correlation for some variable under investigation. A significant correlation would imply that as one specific measurement changes, a corresponding change takes place in the other measurement in the pair, above a level one may expect purely by chance.

Being drawn from the same population does not imply dependence, so Option 2 is not correct. Randomness of the samples is not relevant, and is in any case likely to imply a *lack* of dependence, so Option 4 is also not correct.

Base your answers to Questions 3 to 5 on the following scenario:

Sally, an educational psychologist, predicts that the mean IQ score of a group of 50 children in a special school for gifted children will be higher than the expected population average of 100.

Question 3

Indicate which *null hypothesis* Sally should specify from the options below:

1. $H_0: \bar{x} = 100$
2. $H_0: \mu = 0$
3. $H_0: \bar{x} > 100$
4. $H_0: \mu = 100$

➔ ***Answer:*** The correct answer is option 4.

Sally must compare the mean of her sample of 50 children with a population mean, and she knows this is equal to the expected population average of 100. A null hypothesis is a hypothesis which states that there is no effect, and in this case it would mean no difference between the observed mean (calculated after testing the sample of children) and the population mean of 100. Option 1 is wrong because hypotheses are always stated in terms of *population parameters* (not sample statistics), which in the case of a mean is symbolised by μ. This is also an error in Option 3, but that error is made worse by the fact that 'greater than' is not how we would formulate a null hypothesis, because 'greater than' implies an effect, not the lack of an effect which is implied in a null hypothesis. Option 2 is wrong because the mean IQ in the population used in the comparison is 100, not 0.

Question 4

From reading the psychometric test manual, Sally knows that the IQ test was standardized on a mean of 100 and a standard deviation of 15. Which of the options below would be the most appropriate statistical test which she could use to test the hypotheses?

1. t_c
2. $\mathbf{z}_{\bar{x}}$
3. $t_{\bar{x}}$
4. t_d

➔ ***Answer:*** Option 2 gives the most appropriate test statistic to use.

Sally needs to compare a sample mean with a constant population mean which is already known to be 100. This implies a single sample test for comparing a mean with a given value (100 in this case). It would therefore be either a single sample **z**-test or a single sample t-test. Because the population standard deviation is known (it is specified in the question above as σ = 15) the most appropriate test here would be the single sample **z**-test, which is written $\mathbf{z}_{\bar{x}}$. (In such a case the **z**-test is in fact more *powerful* than a t-test; i.e. it is more sensitive to statistical effects).

Option 1 and Option 4 both refer to two-sample t-tests which are to be used to compare two sample means and not one sample mean with a constant population mean.

Question 5

Sally decides to perform her statistical test at a level of significance of $\alpha = 0.05$. Based on the data from the gifted children the computer calculates a test statistic and reports a two-tailed p-value of $p = 0.082$. What can Sally conclude with regard to the hypotheses?

1. She should reject H_0, and can conclude that the IQs of the gifted children is significantly higher than the population average
2. She cannot reject H_0 and so she cannot conclude that the IQs of the gifted children is significantly higher than the population average
3. She should reject H_1 and can therefore conclude that the IQs of the gifted children is not higher than the population average
4. She cannot make a conclusion because the exact value of the calculated test statistic is not provided

➔ ***Answer:*** The correct answer is option 1.

To test whether the null hypothesis should be rejected, the calculated p-value should be found to be smaller than the level of significance α, which was set by the researcher at 0.05. This would imply that the probability of making an error if we reject the null hypothesis is less than α, which is why we would be willing to reject it. Note however that the p-value given above is for a two-tailed test, whereas Sally requires a one-tailed test, because the alternative hypothesis being tested is one-sided. This is implied in the scenario above: Sally wants to know whether the mean IQ score of children in the school for gifted children is *higher than* the population average of 100. She is not interested in the outcome if the mean IQ score is significantly less than this population mean. The two-tailed p-value should therefore be divided by 2 to calculate a one-sided p-value (see p. 81 of the PYC3704 Guide):

One-tailed p-value = (two-tailed p-value)/2 = 0.082/2 = 0.041

This one-tailed p-value is indeed smaller than the chosen level of significance (p-value = $0.041 < \alpha = 0.05$), so the null hypothesis can be rejected. This implies that the mean IQ of the gifted children is significantly higher than the population average of 100.

Note that the reason why we calculate a test statistic is to use it to determine what the p-value would be. In this case the p-value was given (even though we needed to adjust it), so Option 4 is not valid.

Question 6

A researcher wants to test the efficacy of psychotherapy aimed at relieving anxiety. She applies a test which measures level of anxiety to 30 patients before their therapy begins, and again three months later, predicting that the latter scores will be lower (reflecting less anxiety). Which description of the research design below is the most appropriate?

1. A two-sample groups design with independent groups
2. A two-sample groups design with dependent groups
3. A one-sample groups design
4. A design where the correlation between two variables is tested

➔ ***Answer:*** Option 2 is correct.

This is a repeated measures design, with the same people being measured twice, before and after an event. From the point of view of formal research design, this is to be regarded as two samples with highly dependent measurements. (See the top of p. 118 in the Guide for PYC3704).

Question 7

Suppose you are comparing two means with a t-test for independent samples, and you find that the value of the statistic calculated for your research results is zero. Which conclusion is appropriate?

1. The null hypothesis is likely to be true
2. The probability of a Type I error is zero
3. The alternative hypothesis is likely to be true
4. A calculation error was made because a t-statistic value can never be zero

➔ ***Answer:*** The correct alternative is option 1.

Look at the formula for the t-test for two independent samples:

$$t_c = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The only time that the value of t_c can be zero, is when there is no difference between the means, which implies that the null hypothesis must be true. A similar argument works for any of the other forms of the t-test.

This also implies that Option 3 must be false. If the means are the same the t-statistic will be zero, so Option 4 is also not true. Option 2 is false because the probability of a Type I error will actually be very large: if there is no difference between the means you are sure to make an error if you reject the null hypothesis in spite of this. The whole point of the test is to determine whether the means differ significantly. If they do not differ at all there is actually no need to do a statistical test.

Question 8

The *probability under the null hypothesis* of obtaining a t-value of 2.5 or higher in the case of a one-tailed test is - - - - - that for a two-tailed test.

1. the same as
2. twice
3. half
4. unrelated to

➔ ***Answer:*** The correct Option is 3.

Note that the question actually relates not to the t-statistic value as such, but to the *probability* of getting that particular value for the t-statistic; in other words, to the p-value. A one-tailed p-value (used in the case of a directional hypothesis) is half the size of a two-tailed probability. Conversely, a two-tailed p-value (used in the case of a *non-directional* hypothesis) is twice the size of a one-tailed p-value. (See page 81 of the PYC3704 Study Guide).

Base your answers to Questions 9 to 11 on the following scenario:

An educational psychologist believes that the performance of pupils in mathematics can be improved by teaching them to play chess. She tests this on two samples of children. She gives chess lessons to a treatment group of 20 pupils while a control group of 24 pupils are given their normal classes. The psychologist intends to test her hypothesis at a significance level of $\alpha = 0.05$.

Question 9

Which is the most appropriate research hypothesis for the teacher to test?

1. The mean mathematics score for the treatment group will be greater than that of the control group
2. The mean mathematics difference score (difference scores of treatment group minus control group) will differ significantly from zero
3. The mean mathematics score for the control group will be greater than that of the treatment group
4. There will be a significant difference in the mean scores for mathematics between the treatment and control groups.

➔ ***Answer:*** Option 1 is the most appropriate of the options.

Given the fact that the psychologist expects to find an improvement in the mathematics score, one would expect a greater score for the treatment group (which were exposed to the chess lessons) than for the control group (which was not). Option 3 is the wrong way round and Option 4 indicates two-sided hypotheses, where the expected direction of the change (an improvement in mathematics scores) is not taken into account. Difference scores can only be calculated for matched pairs of data, so Option 2 is not appropriate.

Question 10

Which is the appropriate test statistic to be calculated when testing this hypothesis (as specified in Question 9)?

1. The t-statistic for the difference between the means of two independent samples
2. The t-statistic for the difference between the means of two dependent samples
3. The t-statistic for the mean difference score of a single sample
4. The r-statistic for the correlation between two variables

➔ ***Answer:*** Option 1 is correct.

There are two samples (treatment and control groups) which can be regarded as independent (there is no particular systematic relationship between them). Since two independent samples of different sizes are being compared, rather than two variables from a single sample, the correlation coefficient is not really relevant, so Option 4 is also incorrect.

Question 11

When calculating the test statistic referred to above, the psychologist uses a computer program to calculate the p-value. The computer presents her with a two-sided p-value of $p = 0.054$. What can she conclude?

1. The outcome is significant, so she can conclude that the alternative hypothesis can be rejected and that learning to play chess does have a positive effect on the mathematics ability of the pupils
2. The outcome is significant, so she can conclude that the null hypothesis can be rejected and that learning to play chess does have a positive effect on the mathematics ability of the pupils
3. The outcome is not significant, so she can conclude that the null hypothesis cannot be rejected and that chess has no effect on the mathematics ability of the pupils
4. The outcome is not significant, so she can conclude that the null hypothesis cannot be rejected and that chess has a negative effect on the mathematics ability of the pupils

➔ ***Answer:*** Option 2 is correct.

Note that we need to do a *directional* test (as shown in the answer to Question 9 above). The two-sided p-value given by the computer must therefore be divided by 2 before it can be compared to the level of significance.

This gives us a one-tailed p-value of 0.054/2 = 0.027. This is smaller than the level of significance which was given in the scenario above as $\alpha = 0.05$, and the outcome is therefore significant. For this reason, Options 3 and 4 can be eliminated. A significant result means that there is only a small probability that the null hypothesis is true, so it is relatively safe to reject this null hypothesis. This in turn implies that we can accept the alternative hypothesis, that is, we can conclude that learning to play chess does have a positive effect on the mathematics ability of the pupils.

Note that Option 1 is false because significance implies that the null hypothesis can be rejected, not the alternative hypothesis. In any case, hypothesis testing is always concerned with whether the null hypothesis can be rejected or not. If the p-value was larger than the level of significance, we would have concluded that the probability that the null hypothesis is true is too large for us to reject it (in other words, the improvement in score for the pupils who play math is not greater than what we would expect purely as a result of chance). By convention, this not referred to as 'rejecting the alternative hypothesis' even though this is what is implied (see Section 3.3.1 in the PYC3704 Guide for an explanation of this rule).

Question 12

A scatter plot is a graphical representation of - - - - -.

1. the relationship between two variables measured on a nominal scale within a single group
2. the frequency distribution of a sample of measurements
3. relationship between two groups of subjects with regard to a single variable measured on an interval or ratio scale
4. the relationship between two variables measured on a ratio or interval scale within a single group

➔ ***Answer:*** Option 4 is correct.

This type of graph is described on pp. 130 – 133 of the PYC3704 Study Guide. It would make sense to use such a graph only if the data represented actual measurements and the distances between measurements are equal, that is to say, for ratio or interval scale measurements (see Appendix B in the Guide). Such a graph can only be used when there are matched pairs of data, which implies two variables relating to a single group, which is why Option 3 is not appropriate. Option 2 is just false, since scatter plots are not relevant to frequency distributions.

Question 13

Which of the following does **not** represent a valid value for Pearson's product-moment correlation coefficient?

1. -0.72
2. 0.00
3. -1.01
4. 1.00

➔ ***Answer:*** Option 3 is the correct answer, since a number smaller than -1 is **not** a possible value for Pearson's product-moment correlation coefficient r. Pearson's correlation coefficient can range from -1 to 1, so Options 1, 2 and 4 all represent possible values. (See p. 133 in the PYC3704 Guide).

Question 14

The - - - - - of the correlation coefficient r for two variables indicates the direction of the relationship, while the - - - - - indicates the strength of the relationship.

1. value; sign
2. linearity; magnitude
3. sign; value
4. size; slope

➔ ***Answer:*** The correct alternative is Option 3.

A correlation coefficient is a measurement of the linear relationship between two variables and it can vary between -1 and +1. The absolute *value* of the number (ignoring the sign) indicates the *strength* of the relationship, while the *sign* (+ or -) shows the *direction* of the relationship. If the sign is positive the variables vary in the same direction, but if it is negative it implies that as one variable gets bigger, the other tends to become smaller (See pp. 130 - 133 of the PYC3704 Guide).

Question 15

A contingency table represents - - - - -

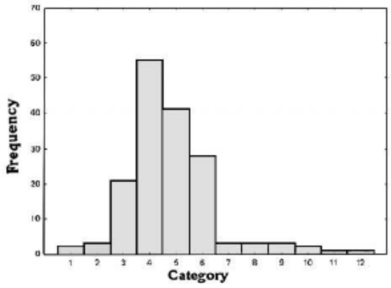
1. the distribution of the data as measured by a variable for a sample
2. the cross classification of frequency counts for two nominal-scale variables
3. the plot of the relationship between two variables
4. the probability distribution of outcomes of a random experiment

➔ ***Answer:*** Option 2 is correct.

A contingency table is a two dimensional table used to represent the cross classification, or cross tabulation, of the responses relating to two nominal or categorical variables. It is basically a way to display and record the relationship between the two variables. The frequency counts of one variable are presented in the rows of the table and the frequency counts of the other variable in the columns, as shown in table 6.4 on p. 142 and Table 6.5 on p. 144 of the PYC3704 Guide. (See also the table in the scenario for Questions 24 and 25 below).

Question 16

The type of graph reproduced below is called a - - - - -?



1. scatterplot
2. histogram
3. normal curve
4. probability distribution

➔ ***Answer:*** Option 2 is correct.

A histogram is a graphical display which shows what proportion of cases fall into each of several *categories*. The categories are usually specified as non-overlapping intervals of some variable and they have to be adjacent to each other. (An example is given on p. 40 of the PYC3704 Guide). A scatter plot (Option 1) is a display of data points for sets of variables (see the Guide, pp. 126-8). A normal curve (Option 3) is a graphical representation of the normal distribution (see p. 42 of the Guide).

Question 17

A researcher hypothesizes that a relationship should exist between spatial ability and general aptitude for mathematics. She collects the results of a sample of n = 100 school children for a mathematics test and measures the spatial ability of each with a test that represents a person's ability to rotate objects mentally on a 10-point scale.

Which of the following is the most appropriate way to express the null hypothesis for this research?

1. $r = 0$
2. $\mu = 0$
3. $\bar{x} = 0$
4. $\rho = 0$

➔ ***Answer:*** Option 4 is correct.

A null hypothesis implies that there is no effect, which in this case implies no relationship between the two variables *spatial ability* and *general aptitude for mathematics*. If the value of a Pearson's correlation coefficient is calculated for two variables with no relationship between them, one would expect it to be close to or equal to zero. Formal hypotheses are set in terms of population parameters, and the symbol for a population correlation is the Greek letter ' ρ' (or rho), not the Roman letter 'r' (see p. 137 in the PYC3704 Study Guide).

Base your answers to Questions 18 to 23 on the following scenario:

An industrial psychologist runs a series of workshops which provided assertiveness training to middle managers at a particular company. She wants to determine whether the workshops are effective and she is also interested to know whether there is any relationship between assertiveness and empathy. She tests each participant before the workshops commence on a questionnaire that measures the level of assertiveness of each participant as well as on their score on a scale that measures empathy. After the training programme is completed, each participant is again tested for their level of assertiveness. The table below shows the results of the measurements.

Case no.	1	2	3	4	5	6	7	8	9	10	11	12
Empathy	5	3	3	3	1	5	2	1	2	4	2	4
Assert1	1	4	2	3	4	2	3	5	3	1	4	1
Assert2	3	5	3	5	5	3	5	4	5	4	5	4

Key:
Empathy and *Assertiveness* were tested on 5-point scales (ranging from 1 equalling 'a small amount' to 5 equalling 'a large amount'). For Assertiveness, *Assert1* indicates the test result before the workshop, and *Assert2* is the test result thereafter.

Question 18

The researcher wants to determine whether the workshop is effective by comparing the measurements on the assertiveness test *before* the workshop with the measurements on this test *after* the workshop. Which of the following is an appropriate test statistic to calculate?

1. Pearson's correlation statistic (r)
2. The t_c test statistic
3. The t_d test statistic
4. The chi-square (χ^2) test for two variables

➔ ***Answer:*** Option 3 is the correct answer.

The problem implies that you would like to compare the two mean values for assertiveness, for matched pairs of measurements, before and after the workshop. Therefore independent group testing (Option 2) is eliminated.

Do not let the fact that the samples being compared refer to the same people being tested twice confuse you (see the top of p. 118 in the Study Guide for PYC3704, where the use of the t_d test for test-retest situations is explicitly mentioned). Calculating Pearson's correlation statistic r (as suggested in Option 1) is technically possible but is not a suitable choice for this particular problem. The correlation coefficient would tell you whether people in the same group who had a high assertiveness score before the workshop still have a high score thereafter. A high correlation could however occur even if the workshop was not effective at all (for example, if each individual somehow obtained nearly the same score as before; or even if their assertiveness became systematically *less* rather than *more*, as the researcher is likely to hope for). This is also why a directional hypothesis would be advisable here.

The variables being compared are quantitative measurements, not categories or nominal level measurements, so the chi-square test (Option 4) is not applicable here.

Question 19

The researcher calculates the value of the appropriate test statistic (in Question 23 above). In which of the four intervals below will the absolute value of the test statistic fall (i.e., ignoring a plus or minus sign)?

1. Between 0 and1.0
2. Between 1.0 and 2.0
3. Between 2.0 and 3.0
4. Above 3.0

➔ ***Answer:*** Option 4 is the correct answer.

In the previous question, we have established that the statistic that should be calculated is t_d. To calculate it, you first need to find the mean and standard deviation of the differences, which you can find from the table below, where d is each value of Assert1 subtracted from the corresponding value of Assert2 (see also the example on p. 113 of the Guide):

Assert1	1	4	2	3	4	2	3	5	3	1	4	1
Assert2	3	5	3	5	5	3	5	4	5	4	5	4
Difference (d)	2	1	1	2	1	1	2	-1	2	3	1	3

The mean and standard deviations of d can be calculated (using the formulas in Appendix C of the Guide, substituting x with d):

The mean of d will be:

$$\bar{d} = \frac{1}{n}\sum d = \frac{1}{12}(2+1+1+2+1+1+2-1+2+3+1+3) = \frac{1}{12}\times 18 = 18/12 = 1.5$$

Using this mean, the standard deviation of d (s_d) will be:

$$s_d = \sqrt{\frac{\sum(d-\bar{d})^2}{n-1}} = \sqrt{\frac{(2-1.5)^2+(1-1.5)^2+...+(3-1.5)^2}{12-1}} = \sqrt{\frac{0.5^2+(-0.5)^2+...+1.5^2}{11}}$$
$$= \sqrt{\frac{0.25+0.25+0.25+0.5+...+2.5+0.25+2.5}{11}} = \sqrt{\frac{13}{11}} = \sqrt{1.181818} = 1.087$$

Using these values in the formula on p. 119 of the Guide, the calculation of the t_d statistic proceeds as follows:

$$t_d = \frac{\bar{d}-\bar{D}}{\frac{s_d}{\sqrt{n}}} = \frac{1.5-0}{(\frac{1.087}{\sqrt{12}})} = \frac{1.5}{1.087/3.464} = \frac{1.5}{0.3138} = 4.780$$

Note that the symbol $\bar{D}$ in the formula refers to the expected value of the mean differences if the null hypothesis were true, which would be 0. If the null hypothesis is true, there would be no difference in measurements, so the mean difference score would be zero. (See p. 119 in the PYC3704 Guide).

Also keep in mind that whether the sign of the calculated t_d value is positive or negative will depend on which data set you subtract from which, and whether or not it matters will depend on whether you are testing a one- or two-tailed alternative hypothesis. We would expect leadership ability to increase (which implies directional testing), so the mean leadership ability *after* the workshop should be greater than mean leadership ability *before* the workshop if the alternative hypothesis is valid.

Question 20

The researcher finds that the test statistic (calculated in Question 19) yields a p-value of p = 0.000572 (as calculated by a computer). What conclusion can be drawn from this? (Use a level of significance of 1%).

1. There is a significant improvement in the level of assertiveness as measured before and after the training
2. There is no significant improvement in the level of assertiveness as measured before and after the training
3. The probability value is too low, so the null hypothesis cannot be rejected with confidence
4. There is a significant difference in the level of assertiveness as measured before and after the training, in favour of the former

➔ ***Answer:*** Option 1 is correct.

The p-value is smaller than the level of significance (that is, p-value $0.000572 < 0.01$). The small p-value implies that there is a very small probability that the difference between the two means (before and after the workshop) is the result of chance or random measurement error, so the null hypothesis (which would imply no change) can be rejected.

Option 2 is incorrect and Option 3 implies that a low p-value is associated with *not* rejecting the null hypothesis, which is the wrong way round. The smaller the p-value, the less the chance is that the null hypothesis is valid. The answer given in Option 4 is in the wrong direction: it implies assertiveness is greater before than after the assertiveness training.

Question 21

The researcher is also interested to know whether the difference between the two assertiveness scores before and after the workshop is of *practical* significance; that is to say, irrespective of its statistical significance, whether the effect can be regarded as big or small. Calculate the relevant measurement of the size of the effect to answer the question below.

The size of the effect was found to be - - - - -

1. small
2. medium
3. large
4. impossible to determine from this data

➔ ***Answer:*** Option 3 is correct.

To get an indication of the practical usefulness of the result the *effect size* should be calculated. This would give an indication of the absolute size of the statistical effect, independent of the sizes of the samples that were used (see Section 3.3.3 in the PYC3704 Study guide). In the case where two means are being compared, this effect size can be found by calculating Cohen's d. The general form of this measurement is to subtract one mean from the other and then to divide this by the overall standard deviation for all the data (see p. 87 in the PYC3704 Study Guide). The actual way in which Cohen's d is to be calculated can however differ depending on which mean is to be compared with which (i.e., it differs slightly depending on which statistical test was used for this data).

The specific form of Cohen's d which a researcher should use when those means being compared comes from two dependent samples is based on the mean and standard deviation of the difference score ***d***. (The fact that 'd' can refer to a difference score and that Cohen chose 'd' for his effect size index is just a coincidence; do not confuse them). The difference score values which were calculated before (in Question 19) can be inserted directly into the formula for Cohen's d.

The version of the formula which is specifically designed for dependent means is given on p. 120 of the PYC3704 Guide:

$$Cohen's\ d = \frac{\bar{d} - \bar{D}}{s_{\bar{d}}} = \frac{1.5}{1.087} = 1.38$$

(The explanation on why we can assume $\bar{D} = 0$ is the same as that given in the feedback to Question 19, above).

The rule of thumb given on p. 88 of the PYC3704 Guide can be used to judge the relative practical importance of the result. An effect of greater than 0.8 can be regarded as 'large'.

Question 22

The researcher is also interested to determine whether a relationship exists between *empathy* and *assertiveness* before the workshop commences. Which of the following is the most appropriate test statistic to use?

1. The chi square (χ^2) test statistic
2. The Pearson's correlation (r) test statistic
3. The t_c test statistic for independent variables
4. The t_d test statistic for dependent samples

➔ ***Answer:*** The correct choice would be Option 2.

You want to establish whether a relationship exists between two variables (*Assert1* and *Empathy*), neither of which are categorical (nominal scale measurements), and which represents measurements on the same group of people. Therefore, a test based on Pearson's correlation is the most appropriate. If the variables were categorical (nominal scale measurements), the chi-square statistic (Option 1) would have been appropriate. A variation of the t-test (Options 3 and 4) would be appropriate if you were comparing the mean values of the two measurements, not the relationship between them.

Question 23

Which of the following is the result if this test statistic (in Question 22) is calculated? The value of the test statistic will fall - - - - -.

1. below -0.4
2. between -0.4 and 0
3. between 0 and 0.4
4. above 0.4

➔ ***Answer:*** Option 1 is correct.

To answer the question you will have to calculate r. The best way to do this would be to first create a table with a list of all the variables and their derivatives which you need (similar to the table on p. 135 of the Guide).

Such a table is presented below:

Case	Empathy	Assert1	Empathy^2	Assert1^2	Empathy X Assert1
1	5	1	25	1	5
2	3	4	9	16	12
3	3	2	9	4	6
4	3	3	9	9	9
5	1	4	1	16	4
6	5	2	25	4	10
7	2	3	4	9	6
8	1	5	1	25	5
9	2	3	4	9	6
10	4	1	16	1	4
11	2	4	4	16	8
12	4	1	16	1	4
Sum(Σ)	35	33	123	111	79

The formula on p. 134 of the Guide is then used with information in this table substituted in the correct places to calculate r, as follows:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = \frac{(12\times 79)-(35\times 33)}{\sqrt{[(12\times 123)-35^2]\times[(12\times 111)-33^2]}}$$

$$= \frac{948-1155}{\sqrt{[1476-1225]\times[1332-1089]}} = \frac{-207}{\sqrt{251\times 243}} = \frac{-207}{\sqrt{60993}} = \frac{-207}{246{,}968} = -0.838$$

Note the difference between ∑xy and ∑x∑y. In the first case you multiply each x and y-value, then add the results up afterwards, while in the second case you add all x-values, add all y values, then multiply the two results.

The high negative correlation implies that the more assertive a person (in this sample of data) is, the less empathy they are likely to have. Since this value of r = –0.838 is smaller than –0.4, the result fits into the category given in Option 1.

Use the following scenario to answer Questions 24 and 25.

An industrial psychologist is conducting research to see whether a person's gender affects the type of post they occupy within a particular organization. She collects data from a 45 employees, classifying them as Male or Female, and according to the type of post which they occupy. From this, she creates the table below.

Cross-classification	***Male***	***Female***
Management	*4*	*2*
Clerical	*7*	*14*
Technical	*6*	*2*
Human resources	*3*	*7*

Question 24

What is the kind of table that is given in the scenario above referred to as?

1. A frequency distribution
2. A correlation matrix
3. A contingency table
4. A cross tabulation table

➔ ***Answer:*** The correct answer is Option 3.

This kind of table is used to cross-classify frequency data which are distributed in terms of two classification variables (these variables are often nominal level measurements but they need not be). For example, in the table above, there are 4 people in the sample of data who can be classified as both 'Male' and members of 'Management'. (See p. 142 in the PYC3704 Guide and also Question 15 above).

Question 25

If gender has no effect on the posts that employees occupy in the organization, how many males would one expect to find in the human resources department?

1. 5.62
2. 20
3. 4.44
4. 3

➔ ***Answer:*** The correct answer is given in Option 3.

To answer the question, you need to calculate row and column totals. Expand the table as follows:

Cross-classification	*Male*	*Female*	Row total
Management	*4*	*2*	*6*
Clerical	*7*	*14*	*21*
Technical	*6*	*2*	*8*
Human resources	*3*	*7*	*10*
Column Total	*20*	*25*	*45*

To get the expected frequency (the way data would be distributed purely by chance) you need to multiply the total number of people in *human resources* (total of row 4) with the total number of *males* (total of column 1), and divide this by the overall total, using the formula:

$$e_{ij} = (t_{row} \times t_{col}) / t_{total}$$

Therefore: e_{41} = (10 x 20)/45 = 200/45 = 4.4444 (which can be rounded off to 4.44)

Note that Option 4 is incorrect because it gives the *observed* frequency for tall followers, while the question asks for the *expected* frequency if the frequencies were distributed equally with no interaction between gender and job category. Option 1 is the total divided by the number of possible categories (i.e. 45/8) but this does not take account of the fact that the expected distribution of the classifications should be *proportional* to the overall distribution of the data. Option 2 is the number of males. (See pp. 143-5 of the PYC3704 Guide for details on calculating expected frequencies from a contingency table).