Tutorial Letter 202/1/2014

Psychological Research

PYC3704

Semester 1

Department of Psychology

- **The Exam Paper**
- **Feedback to Assignment 02**
- **Additional exercises**

Bar code

Learn without limits.

UNISA university of south africa

Dear Student

This study letter contains answers and explanations of the correct options for each question in Assignment 02. Work your way systematically through the answers we provide, comparing it with your own results, which you will receive separately.

Use the marks that you have received for the assignments as an indication of the standard of your work. If you did not receive pass marks for the assignments, you will have to work much harder if you want to pass the exam.

The questions in the two assignments and those at the end of each topic in the Study Guide for PYC3704 should be used as part of your preparation for the exam. We suggest that you should try to do the questions from the Guide and the assignments without looking up the answers. In this way you will see which questions you can answer, and which ones you cannot answer. This will help you to identify the difficult questions, and force you to rethink the reasons for the answers. Pay special attention to the explanations we provide, since this will give you an indication of the weak spots in your own understanding of the material in the Study Guide. Do not attempt to memorise the questions and their answers. Rather try to *understand* the explanations for the correct option in each question.

1. The Exam Paper

Please read Tutorial Letter 101 again on the format of the exam paper. The exam will consist of a 2 hour paper with 70 multiple choice items, similar to those in the assignments. Each item will have ***four*** possible alternative answers (similar to the assignments), out of which you will have to select the ***best*** option, the one that seems to be ***most*** correct. You can earn a maximum of 70 marks in the exam. Your mark will be converted into a mark out of 80. The other 20% for the exam will be your year mark which you obtained for Assignment 02. If you get 40% or less for the exam paper, the assignment mark will ***not*** be used in calculating your final mark.

The 70 items in the exam will represent the learning material in the PYC3704 Study Guide in *more or less* the following way:

Topic 1: more or less 12 items
Topic 2: more or less 12 items
Topic 3: more or less 20 items
Topic 4: more or less 6 items
Topic 5: more or less 10 items
Topic 6: more or less 10 items

You should know how to calculate the basic descriptive statistics such as a mean, a standard deviation and a variance (Appendix C). You should know how to calculate expected values in a contingency table, a standard error and a z-score as well as how to use the z-tables (Appendix D). You should know about the different measurement levels (Appendix B). You will not be examined on the rules of arithmetic in Appendix E, but you should know how to use them to do calculations. If complex formulas such as those needed for various test statistics are required, we will supply them in an appendix. You should be able to recognise the formulas and know how to use them, but you do not have to memorise them. Because of time limits, we shall not expect of you to do long calculations in the exam (such as – for example – Questions 24 or 28 in Assignment 02, which are discussed below). You will not be expected to do calculations related to the binomial distribution (section 2.2.3 on p. 41 – 46 of the PYC3704 Guide).

A couple of blank pages will be added in the examination booklet for rough work purposes. You may use a pocket calculator in the exams, but it should not be one on which you are able to store text. Make sure you know how your calculator works *before* you enter the exam hall!

If you fail the exam with a mark of between 40% and 49%, you will get the opportunity to write a supplementary exam at the end of the next semester. You have to inform the Examination Department at Unisa if you intend to write the supplementary exam and will have to pay examination fees. *No further* *supplementary exams or special exams will be granted after this.* If you fail again, you will have to reregister for this course.

2. Feedback to Assignment 02 for semester 1

For your convenience, each question is given followed by the appropriate answer and an explanation of the correct response. Work your way systematically through these, comparing it with your own answer. Even if you chose the correct alternative, you may find that the explanations we give are useful. Try to *understand* the explanations. Many of the items measure insight, not just factual knowledge. You will not pass this course if you try and memorise the questions and answers!

Base your answers to Questions 1 to 3 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 1

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 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 a lack of an effect. Option 2 is wrong because the mean IQ in the population used in the comparison is 100, not 0.

Question 2

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. $z_{\bar{x}}$
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 know 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 $\sigma = 15$) the most appropriate test here would be the single sample z-test, which is written $z_{\bar{x}}$. 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 3

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 which implies that the IQs of the gifted children is significantly higher than the population average
2. She cannot reject H_0 which implies that she cannot conclude that the IQs of the gifted children is significantly higher than the population average
3. She should reject H_1 which implies 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

In this question, the words 'which implies that' should be interpreted as referring to the consequence of rejecting the null hypothesis or not rejecting it. To make this more clear, we put an alternative formulation of the question on myUnisa*. Note that this is the same question so the answer is not affected.*

The alternative formulation is as follows:

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 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:

One-tailed p-value = (two-tailed p-value)/2 = 0.082/2 = 0.041 (see p. 81 of the PYC3704 Guide).

This one-tailed p-value is indeed smaller than the chosen level of significance ($p\text{-value} = 0.041 < \alpha = 0.05$), 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 4

When applying a statistical test, the probability of a Type I error is equal to - - - - -.

1. the level of significance
2. the p-value of the test statistic under the null hypothesis
3. the p-value of the test statistic under the alternative hypothesis
4. 0.05 or 0.01

➔ ***Answer:*** Option 2 is correct.

An error of Type I is when a researcher makes an error by rejecting the null hypothesis when it should not be rejected, and the p-value gives the probability that this error will be made based on the specific data being observed (see section 3.3.2 of the PYC3704 Guide). That is to say, the p-value is calculated with the assumption that the null hypothesis is true (that is what 'under the null hypothesis' means) and this possibility is compared with the sample data (in this case, the calculated sample mean). So what is being investigated is the probability that any effect being observed in the data is merely the consequence of chance effects such as measurement error, and this directly reflects the size of the error which will be made if the null hypothesis is rejected (an error of Type I). A researcher would want to keep this value small (smaller than α).

Option 1 is wrong because the level of significance is the *maximum* probability of making an error of type I that we would allow – that is, we want the p-value to fall below this – but it is not a direct estimate of that error for a specific observation of a sample of data. We actually should choose this value *before* making observations by looking at a sample of data. Option 3 is incorrect because it is the probability of the *null hypothesis* being true that is investigated, not the alternative hypothesis. Type I errors are based on calculated p-values for a specific set of data, not predetermined to be 0.05 or 0.01 (unlike the level of significance α), so Option 4 is also wrong.

Question 5

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. accepting the alternative hypothesis

➔ ***Answer:*** Option 3 is correct.

We know that the extent of the type I error that a researcher is willing to make is controlled by the researcher by setting the level of significance (α) in advance. The probability of a type II error (β) is not controlled in advance by the researcher except for the fact that we know that the lower (smaller) the probability of a type I error (α) the greater (larger) the probability of a type II error (β). You could eliminate error of type I (rejecting H_0 when you should not) altogether by never rejecting the null hypothesis, irrespective of how small the p-value becomes, but that would make the probability of not rejecting H_0 when you should reject it (error of type II) an absolute certainty. See page 83 in the Guide for PYC3704.

Question 6

When comparing a sample mean with a specific value like an expected population mean, when should the one sample t-test be used?

1. s is unknown
2. σ is unknown
3. σ is known
4. s is known

➔ ***Answer:*** Option 2 is correct.

Whether a t-test is used or not depends on whether the population standard deviation σ is known (so options 1 and 4 are irrelevant). If the σ is known, the one sample z-test could be used, which rules out option 3. The t-distribution is a refinement of the z-distribution (i.e. the standard normal distribution) and it is to be used when σ is unknown (see Study Unit 4.2 in the Guide for PYC3704). The sample standard deviation (s) can be calculated from sample data and is used to produce an estimate for the value of σ when a t-test is performed.

Question 7

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). 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 one-sample design with two variables being investigated for correlation

➔ ***Answer:*** Option 2 is correct.

The same persons are tested repeatedly, and the first test (before the therapy) is matched with the second test (after the therapy) for each person, which produces matched pairs of measurements. This is typical example of dependent measurements (see pp. 117 – 118 in the Study Guide for PYC3704).

Since the groups are not independent, option 1 is wrong. Note that even though it is one group being tested (a sample of 50 depressed patients), there are *two* samples of measurements being compared, so it is treated as a two-sample design, and options 3 and 4 are therefore wrong. (From a technical point of view, the purpose of the statistical test here is to determine whether two samples of measurements come from the same underlying population of measurements, which is why two samples of measurements can come from repeated measurements on the same group of persons).

Question 8

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

1. The null hypothesis adequately explains the results
2. The alternative hypothesis should be rejected
3. If the null hypothesis were true, the results which were found in the sample data would be unlikely
4. If the alternative hypothesis were true, the results which were found in the sample data would be unlikely.

➔ ***Answer:*** Option 3 is correct.

The null hypothesis states that there is no difference in the means calculated from samples of data from each of of groups A and B. When we calculate the two means from sample data (which we regard as an observation) we may find a difference in the two calculated means, but at least part of this difference could be due to measurement errors. We calculate the p-value (based on a test statistic with a known probability distribution) to find out what the probability is that that these observed differences in the sample data are just a consequence of measurement error if the null hypothesis is assumed to be true. If this probability is low (lower than a pre-determined cut-off level, α), we conclude that the difference in the two means is statistically significant because the probability that the null hypothesis is true is very small. In other words, we conclude that the size of the difference between means found in the sample data would not be likely if the null hypothesis were true.

Option 1 implies the result is not significant, and so does option 2 (if we take ‘rejecting’ the alternative hypothesis as implying accepting the null hypothesis). We test the null hypothesis, not the alternative hypothesis so option 4 is also wrong.

Question 9

Suppose you find that the value of the t-statistic calculated for your research results is zero. Which conclusion is appropriate?

1. The probability of a Type I error is zero
2. The null hypothesis is likely to be true
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 2.

For example, 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 others 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 1 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.

Question 10

The p-value represents the probability of getting the observed results if - - - - -

1. the alternative hypothesis is true
2. the null hypothesis is false
3. the null hypothesis is true
4. the null hypothesis should be rejected

➔ ***Answer:*** Option 3 is correct.

The observed results are the values which you find in your sample(s) of data, for example the sample mean and sample standard deviation, or (if it is relevant), the correlation coefficient which you calculated.

The p-value shows you the probability of seeing some relationship among these variables based on your calculations (such as a difference between means or a high correlation), if in fact this observed relationship is merely the consequence of chance (in other words, *if the null hypothesis was true*). You are in fact comparing the observed relationships in the data with what you would expect if the null hypothesis is true by calculating a relevant test statistic. This test statistic can then be used to find the p-value if we know the probability distribution of the test statistic. If this probability is small, it implies the null hypothesis is probably *not* true. Note that all three other options imply the opposite: i.e., that the p-value refers to the alternative hypothesis being true, which is equivalent to the null hypothesis being false.

Question 11

When using a t-test to compare the means from two samples of data, a researcher finds a significant difference. She decides to also determine the effect size of the result. For what reason would she do this?

1. To double-check the result
2. To determine the power of the statistical test
3. To determine the probability of making an error if she rejects the null hypothesis
4. The see whether the statistical result is also reasonably meaningful in practical terms

➔ ***Answer:*** Option 4 is correct.

A p-value indicates the probability of the t-test statistic obtained under the null hypothesis, but it does not say anything about the practical significance, or size of the effect, of the difference between the two samples. For example, a small difference in the means between the two samples could have produced a statistically significant result and yielded a low p-value if the samples were large (see the PYC3704 Guide, page 86). A low p-value tells you that the result is greater (more significant) than you would expect by chance, but it should not be used as a direct measurement of the size of the effect. To find out about the actual practical significance (the size of the effect) of the difference between the two samples of data, the researcher would have to calculate an effect size. (See the answer to question 26 below).

Question 12

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.

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).

Question 13

A researcher hypothesizes that the drug treatment of hospitalised schizophrenic patients improves their mental alertness. He studies a random sample of 27 patients to see whether there is a relationship between the number of days of drug treatment and patients' scores on the Mental Alertness Test.

Which is an appropriate null hypothesis for this research?

1. H_0: $r \neq 0$
2. H_0: $\mu = 0$
3. H_0: $r = 0$
4. H_0: $\rho = 0$

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

The symbol 'ρ' (the Greek letter 'rho') is used to represent the population parameter being tested when you calculate the Pearson's correlation coefficient 'r.' That is, you calculate r for the sample, then have to decide whether this is likely to represent a significant linear correlation between two variables for the whole population (with this *population* correlation symbolised by ρ), by looking at the p-value associated with this calculated sample statistic r. In a similar way 'μ' represents the population parameter (statistic) for a mean, and 'σ' the population parameter for a standard deviation. (See p. 137 of the PYC3704 Guide).

Question 14

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 pp. 142 to 144 in the PYC3704 Guide).

Question 15

Which of the following does **not** represent a valid value for Pearson's r?

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

➔ ***Answer:*** Option 3 is the correct answer, since a number of smaller than -1 is not a possible value for Pearson's r. This correlation coefficient can range from -1 to 1, so both options 1 and 2 represent possible values. (Note that on a measurement axis, a value of -1.01 will lie to the left of -1, so it is *smaller than* -1)

Question 16

Which of the following alternate hypotheses requires a two-tailed test of significance?

1. The distribution of test marks on a English Comprehension Test of boys differs significantly from the distribution of test marks for girls
2. The correlation coefficient between test marks and examination marks is less than zero
3. The mean verbal ability score for boys is lower than the mean score for girls
4. There is a positive correlation between assignment marks and examination marks for undergraduate students in a psychology course

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

A two-tailed test implies there is some difference, but you have no prediction of the direction in which the difference lies. This is the situation in option 1, where no indication is given of which of the two groups of marks (for boys or girls) is likely to be greater and which is likely to be lesser. Option 3 predicts that the girls will obtain higher marks than the boys and therefore requires a one-tailed test. Options 2 and 4 make predictions about the direction of the correlations between test marks and examination marks. Option 2 states that the correlation will be negative whereas Option 4 states that it will be positive, but note that both options imply a relationship in a specific direction, and therefore both require one-tailed tests of significance.

Question 17

Pearson's correlation coefficient r can vary from - - - - - to - - - - - and indicates a - - - - - relationship.

1. 0; 1; linear
2. -1; 1; positive
3. -1; 1; linear
4. 0; 1; significant

➔ ***Answer:*** Option 3 is correct.

The values of the Pearson's correlation coefficient can range from -1 for a perfect negative correlation to +1 for a perfect positive correlation. The relationship is called linear because Pearson's correlation coefficient measures the extent to which the relationship approximates a straight line. (See page 133 of the PYC 3704 Guide).

Question 18

Which best describes a scatter plot? A graph showing - - - - -.

1. the position of each of a number of sampling units on each of two variables
2. the relative position of a number of sampling units in the sample and population respectively
3. the spread of measurements on two independent variables in a population
4. the way in which measurements on a variable is distributed around a sample or population mean

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

A *scatter plot* is a graph showing the relationship between two numerical variables. In such a graph the data of the one variable are plotted on the horizontal axis (usually referred to as the X axis), and the data of the other variable on the vertical (or Y) axis (see the PYC3704 Guide, p. 132). It is not a comparison of sample and population (option 2), nor has it to do with spread of data or the independence of variables (options 3 or 4).

Question 19

The - - - - - of the correlation coefficient 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 (that is to say, if the X variable gets larger, how does this affect the variable Y?) (See pp. 130 - 133 of the PYC3704 Guide).

Question 20

A contingency table represents - - - - -

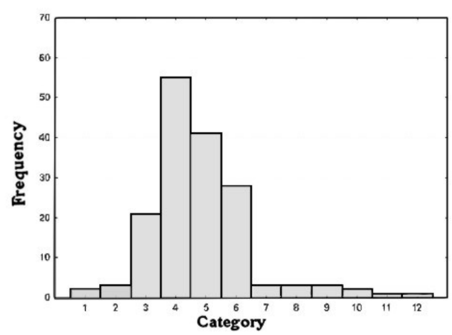
1. the distribution of the data as measured by a variable for a sample
2. the cross classification of frequency counts for two nominal-scale variables
3. the plot of the relationship between two variables
4. the probability distribution of outcomes of a random experiment

➔ ***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 page 142 and table 6.5 on p. 144 of the PYC3704 Guide.

Question 21

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 22

The chi-square (χ^2) test statistic is used to compare - - - - -.

1. the distribution of observed frequency data with the way that the data should be distributed if the null hypothesis is true
2. the variance or spread of observed data with the variance of the data as expected if the null hypothesis is true
3. the way in which each of two variables vary (their variances) with the covariance (how they vary together) of both
4. the extent to which the mean of a variable differs from the mean of another when both are nominal scale measurements

➔ ***Answer:*** Option 1 is correct.

The chi-square test is usually used when you have a cross tabulation of frequency counts of events which are nominal scale measurements. This table is referred to as a contingency table (see the answer to question 20 above). It is used to compare an observed frequency distribution (frequency counts based on a sample of observation) with the frequency distribution which we would expect to find if the null hypothesis of no relationship between two cross-tabulated variables were true. (See p. 140 of the PYC3704 Guide for details).

Base your answers to Questions 23 to 28 on the following scenario:

Alfred is a psychologist who is involved in a series of workshops which provide exercises designed to improve the self-esteem of the participants. Alfred wants to establish the effectiveness of the training programme, so he tests a group of 15 participants on a questionnaire which measures their self-esteem before the training commences and again after the training programme is completed. He also tests them on an anxiety scale before the workshop. The table below shows the results of these measurements.

Key:

Case is a unique number for each participant
Gender is coded with 1 = Male and 2 = Female.
Anxiety was measured on a 5-point scale (ranging from 1='little or no anxiety' to 5='high anxiety').
Self-esteem was measured on a 7-point scale (with 1='very low self-esteem' up to 7='very high self-esteem'). *SelfEst1* indicates the measurement before the training programme, while *SelfEst2* represents the measurement thereafter.

Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Gender	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
Anxiety	1	3	2	4	3	3	5	1	3	4	4	3	3	2	4
SelfEst1	5	7	4	1	3	2	3	5	4	4	4	5	4	5	1
SelfEst2	6	7	6	2	3	3	3	4	5	4	5	5	6	5	3

Question 23

Alfred wants to determine whether there is a gender difference in self-esteem before the workshop. Select the most appropriate way for him to formulate his statistical hypotheses from the options below.

1. $H_0: \mu_M = \mu_F$ and $H_1: \mu_M \neq \mu_F$
2. $H_0: \mu_M = \mu_F$ and $H_1: \mu_M > \mu_F$
3. $H_0: \overline{x}_M = \overline{x}_F$ and $H_1: \overline{x}_M \neq \overline{x}_F$
4. $H_0: \overline{x}_M = \overline{x}_F$ and $H_1: \overline{x}_M < \overline{x}_F$

➔ ***Answer:*** Option 1 is correct.

The question states that 'Alfred wants to determine whether there is a gender difference in self-esteem before the workshop' which implies taking the self-esteem measurements *before* the workshop (*SelfEst1*) of males (Gender = 1) and females (Gender = 2) *separately*, and comparing their means. The null hypothesis would be that no such difference exists, i.e. $\mu_M = \mu_F$. No indication is given of whether males are expected to have either a higher or lower self-esteem, which implies a *non-directional test* with an alternative hypothesis of $\mu_M \neq \mu_F$. Note that in formal hypothesis testing we refer to the population parameters, which in the case of comparing means would make use of the symbol μ for population mean, not $\overline{x}$, which would refer to a sample mean. You calculate sample statistics to make conclusions about population parameters (see section 3.2.1 starting on p. 76 in the PYC3704 Guide).

Question 24

Which is the most appropriate test statistic for determining whether a difference exists *before* the workshop between *males* and *females* (in the '*gender'* variable)?

1. The t_d test for differences between dependent samples
2. The one sample t-test ($t_{\overline{x}}$)
3. The t_c test for differences between independent samples
4. The chi-square (χ^2) test for two variables.

➔ ***Answer:*** Option 3 indicates the most appropriate test statistic to use.

It should be kept in mind that this question refers to the scenario above, and forms a continuation of the previous question (we clarified this in an announcement on *myUnisa*). So what is required is a test for comparing difference in the mean scores of self-esteem before the workshop for the two gender groups (males and females); in other words to test the hypothesis suggested in question 23. These are two independent groups, so a t-test for comparing the means of two independent groups would be effective – which is to say, the t_c-test.

These are not dependent groups forming matched pairs of data, so option 1 is wrong (also note that the sample sizes for the gender groups differ, so it is not really possible to identify matched pairs of measurements). There are two groups being compared, which should be thought of as testing whether two samples come from the same population (which would imply the variable Gender is irrelevant), so a one-sample t-test as suggested in option 2 is not appropriate. Self esteem is a measurement on a 7-point scale which is a stronger level of measurement than merely a nominal level, so the chi-square test (option 4) is not the most appropriate test to use here (although using it would be technically possible).

Question 25

Alfred calculates the value of the appropriate test statistic (as selected in Question 25 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 1 is correct.

It will be necessary to calculate the t_c test statistic to answer this question.

First, calculate the mean and standard deviation for each of the two groups from the sample data in the scenario. For males, the mean would be the mean of all the values of *SelfEst1* with *Gender* indicated as 1:

$\bar{x}_m = \frac{1}{n_m}\sum x_i$ = (5 + 7 + 4 + 1 + 3 + 2 +3)/7 = 25/7 = 3.57 (rounded off)

The standard deviation for the seven males can be calculated with the formula for the sample standard deviation (given in Appendix C of the PYC3704 Study Guide):

$$s_m = \sqrt{\frac{\Sigma(x-\bar{x}_m)^2}{n_m-1}} = \sqrt{\frac{(5-3.57)^2+(7-3.57)^2+\ (4-3.57)^2+(1-3.57)^2+(3-3.57)^2+(2-3.57)^2+(3-3.57)^2}{7-1}}$$

$$= \sqrt{\frac{1.43^2+3.43^2+0.43^2(-2.57)^2+(-0.57)^2+(-1.57)^2+(-0.57)^2}{6}}$$

$$= \sqrt{\frac{2.04+11.76+0.18+6.60+0.32+2.46+0.32}{6}} = \sqrt{\frac{23.68}{6}}$$

$$= \sqrt{3.94} = 1.987$$

The mean and standard deviation for the 8 females (with Gender indicated by 2 in the scenario) can be calculated in the same way, and should come to:

$\bar{x}_f = 4.00$ and $s_f = 1.31$

These values should now be substituted in the formula for the independent-sample t-test (given on p. 114 of the PYC3704 Study Guide):

$$t_c = \frac{(\bar{x}_m - \bar{x}_f)}{\sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}}} = \frac{(3.57 - 4.00)}{\sqrt{\frac{1.98^2}{7} + \frac{1.31^2}{8}}} = \frac{-0.43}{\sqrt{\frac{3.92}{7} + \frac{1.72}{8}}} = \frac{-0.43}{\sqrt{0.56 + 0.22}}$$

$$= \frac{-0.43}{\sqrt{0.78}} = \frac{-0.43}{0.88} = -0.4996$$

So t_c = -0.4996 and the absolute value of t_c is $|t_c|$ = |-.4996| = 0.4996, which can be rounded off to 0.50. This falls in the interval below 1, which is why option 1 is correct.

Question 26

Alfred is also interested to know whether the difference which was observed between the genders on the *self-esteem* scores before the workshop is of practical significance. In other words, he wants to establish whether the effect can be regarded as big or small, irrespective of its statistical significance. After calculating the relevant measurement of the size of the effect, Alfred finds that the effect is - - - - -.

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

➔ ***Answer:*** Option 1 gives the best description of the effect size.

In this case, the issue is not the *statistical significance* of the difference between the means that is the issue (whether it is greater than may be expected by chance for samples of this size) but the *practical significance* (what importance can be given to the difference if a difference is found). This is determined by calculating the *effect size* of the difference between means for the sample data (see section 3.3.3 starting on p. 86 of the Guide for PYC3704).

The reason why one would calculate an effect size is that if we find statistical significance, it just tells us that the effect (the relationship between variables) is greater than can be expected by chance. The size of the p-value (statistical significance) should not be used as an indication of the overall (practical) importance of the effect, because the p-value is very sensitive to sample size. A small statistical effect is visible in a big sample but for a small sample the effect has to be quite large before statistical (above chance) significance will be admitted. (See the discussion in section 6.1.4.1 on p. 139 in the PYC3704 Guide, which uses the test for correlation as an example to explain this phenomenon).

First, we will require the value of the pooled standard deviation; that is to say, the overall standard deviation of all 15 measurements of the variable *SelfEst1*, ignoring group membership (Gender). Using the sample standard deviation formula (in question 25 above), we calculated this to be s_p = 1.61. A quick alternative estimate which can be used if the two sizes of the two groups are nearly equal, is the mean of the two group standard deviations s_p = (1.988+1.309)/2 = 1.648. (You can use this shortcut when the standard deviations of the two groups do not differ too greatly).

Now the effect size for two independent means which are being compared will have to be computed. The formula for this in the case of two independent means is given on p. 116 of the Guide for PYC3704.

$$Cohen's\ d = \frac{\bar{x}_m - \bar{x}_f}{s_p} = \frac{3.57 - 4.00}{1.65} = \frac{-0.43}{1.61} = -0.27$$

Note that for an effect size, we do not care what the sign is since (in this particular case) the decision of which mean to subtract from which (male from female or the other way round) is arbitrary. In other words, we use the *absolute value* of Cohen's d, written as |d| = 0.27 (see Appendix E).

From this we can conclude that the effect size is *small*, using the guideline given on p. 88 of the PYC3704 Guide.

Question 27

Alfred is also interested to determine whether a relationship exists between *anxiety* and *self-esteem* 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 t_c test statistic for independent variables
3. The t_d test statistic for dependent samples
4. The Pearson's correlation (r) test statistic

➔ ***Answer:*** Option 4 is correct.

We want to know the relationship between two measurements of the same group of research participants, so a statistic based on the Pearson correlation coefficient is appropriate (see Topic 6). We are not comparing two samples or two group means, so neither option 1 or option 2 is appropriate. Both of the measurement levels are stronger than nominal scale measurements (see Appendix B), so one is not likely to use the chi-square test, referred to in option 1.

Question 28

Which of the following is the result if this test statistic (in Question 27) 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.

In the answer to question 27 we indicated that the test for Pearson's correlation coefficient (r) will have to be used. The best way to do this would be to first create a table such as the following (similar to the table on p. 135 of the PYC3704 Guide):

Case	**Anxiety**	**SelfEst1**	**Anxiety^2**	**SelfEst1^2**	**Anxiety X SelfEst1**
1	1	5	1	25	5
2	3	7	9	49	21
3	2	4	4	16	8
4	4	1	16	1	4
5	3	3	9	9	9
6	3	2	9	4	6
7	5	3	25	9	15
8	1	5	1	25	5
9	3	4	9	16	12
10	4	4	16	16	16
11	4	4	16	16	16
12	3	5	9	25	15
13	3	4	9	16	12
14	2	5	4	25	10
15	4	1	16	1	4
Sum	45	57	153	253	158

Note that in the fourth column **Anxiety²** refers to Anxiety squared (multiplied with itself) and in the fifth column **SelfEst1²** refers to the square of Self Esteem before the Workshop. The last column is ***Anxiety*** multiplied with ***SelfEst1***.

The formula on p. 134 of the PYC3704 Guide is then used with values from this table substituted to calculate r, as follows:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = \frac{15(158) - (45)(57)}{\sqrt{[15(153) - (45)^2][15(253) - (57)^2]}}$$

$$= \frac{2370 - 2565}{\sqrt{(2295 - 2025)(3795 - 3249)}} = \frac{-195}{\sqrt{(270)(546)}} = \frac{-195}{\sqrt{147420}}$$

$$= \frac{-195}{383.9531} = -0.5079$$

This is less than – 0.4 (notice the negative sign which implies a negative correlation, and – 0.5 is smaller than – 0.4), so the Pearson correlation coefficient r falls in the interval indicated in option 1.

Use the following scenario to answer Questions 29 and 30:

Lebo wonders whether a relationship 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 table below.

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

Question 29

Which test should Lebo use to determine whether the relationship exists?

1. A t-test for independent samples to compare tall and short persons
2. A test based on the correlation coefficient of the relationship between 'length' and 'leadership ability'
3. A chi-square (χ^2) test
4. A t-test for dependent samples because the data all comes from a single sample.

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

The two variables (Person's Length and Leadership Ability) can both be regarded as categorical or nominal level measurements (see Appendix B in the PYC3704 Study Guide). Tall vs. Short is actually an ordinal variable (you can order it in a list) but since it is a dichotomy (only two categories), we cannot really use it as a measurement. When two nominal measurements are cross-classified like this, a chi-square test is the appropriate test to determine whether a significant relationship exists between them (see Study Unit 6.2 in the Guide as well as the feedback to question 22 above). None of the other tests which are suggested in the other options are appropriate for cross-classified nominal scale data.

Question 30

What would you expect the number of people who can be classified as both 'Tall' and as 'Followers' to be under the null hypothesis?

1. Close to 43
2. Close to 20
3. Close to 16
4. Close to 10

➔ ***Answer:*** Option 3 is correct.

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*

Note that the expected frequency for Tall Followers is the frequency value which one would predict if the null hypothesis were true (the *expected* frequency), while the number 22 we see in the table is the *observed* value; i.e. the count of Tall Followers which we observed in an actual sample of data. To get this expected frequency (the way data would be distributed purely by chance) you need to multiple the total for the first column of frequencies (tall people) with the total of the second row (Followers) and divide by the overall total, as follows:

Overall total = 95
Total number of Tall persons = 43
Total number of Followers = 36

So, the expected number of Tall Followers would be as follows:

e_{12} = (43x36)/95 = 1548/95 = 16.2947 (See p. 144 in the PYC3704 Guide).

This is close to 16, which is indicated in option 3.

3. Additional Exercises

Some additional explanations and exercises are given below. We suggest that you use the exercises as a mock exam, to help you prepare for the final exam. In other words, after studying the relevant material for each topic, try doing these exercises without first looking up the answers. You can then mark your own work using the supplied answers, to see how well you understand the material in the PYC3704 Guide. Note that the subjects we choose to discuss here are just some things that you may find useful. It does ***not*** imply that these issues are more important than the rest of the material in the Guide.

Exercises concerning the notion of probability

The excercises that follow are related to the concept of **probability**. Probability is the *liklihood* that something is the case or will happen. A definition of "the probability of an event" is that it is the number of ***favourable*** outcomes divided by the total number of ***possible*** outcomes. An example is how often the number 2 is likely to occur if you throw a six-sided die.

(Note that the discussion below repeats the information in the Study Guide for PYC3704 given in Study Unit 2.1, but it is a simplified description that should be helpful to those of you who have still have difficulty in understanding how probabilities are determined).

> *The probability of a particular event **E** is the number of occurrences of that particular event relative to the total number of events which can occur.*

This concept is used in statistical decision making to draw conclusions about the likelihood that some event will occur, out of a range of possible events.

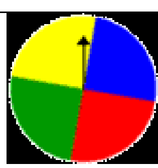
For example, if you think you have a 90% chance of passing the exam, this can be expressed as follows:

Probability of passing = p(pass) = 90/100 = 9/10 = 0.9

Remember that ‘percent’ means ‘out of 100’, so 90% is the same as 0.9. The symbol ***p*** is conventionally used to indicate the probability of an event, and it is usual to express a probability ***p*** as a number between **0** and **1**, where *0* indicates that the probability is zero (an event is impossible), and *1* means absolute certainty. A probability is therefore always expressed as a ratio (decimal proportion) and will, by definition, always have a value of between 0 and 1 (inclusive).

In mathematical form, this can be expressed as $0 \leq p(\text{event } E) \leq 1$. The words or symbols in brackets provide an indication of the specific event in question.

Here is a simple example of a statistical experiment to illustrate the calculation of probabilities.

 **Problems**	A spinning wheel has 4 equal sectors coloured **yellow**, **blue**, **green** and **red**. • What are the chances of landing on **blue** after spinning the spinner? • What are the chances of landing on **red**? • What are the chances of landing on **green OR yellow?**
Answers	• The chances of landing on **blue** are 1 in 4, or one fourth. • The chances of landing on **red** are 1 in 4, or one fourth. • The chances of landing on **green OR yellow** are 2/4, or ½ (=50%)

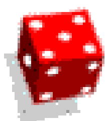
In order to measure probabilities, mathematicians have devised the following formula for finding the probability of an event E:

$$p(E) = \frac{Number\ of\ ways\ in\ which event\ E\ can\ occur}{Total\ number\ of\ possible\ events}$$

Or, if we want to estimate it from data:

$$p(E) = \frac{Number\ of\ observations\ of\ specific\ event\ E}{Number\ of\ times\ the\ experiment\ was\ performed}$$

Let us take a look at an application of this idea.



Question:

A single 6-sided die is rolled. What is the probability of each outcome? What is the probability of rolling an even number? What is the probability of rolling an odd number?

Answer:

The possible *outcomes* of this experiment are **1, 2, 3, 4, 5** and **6.** The *probabilities* of each possible outcome are given below:

$$p(1) = \frac{\text{number of ways to roll a 1}}{\text{total number of sides}} = \frac{1}{6}$$

$$p(2) = \frac{\text{number of ways to roll a 2}}{\text{total number of sides}} = \frac{1}{6}$$

$$p(3) = \frac{\text{number of ways to roll a 3}}{\text{total number of sides}} = \frac{1}{6}$$

$$p(4) = \frac{\text{number of ways to roll a 4}}{\text{total number of sides}} = \frac{1}{6}$$

$$p(5) = \frac{\text{number of ways to roll a 5}}{\text{total number of sides}} = \frac{1}{6}$$

$$p(6) = \frac{\text{number of ways to roll a 6}}{\text{total number of sides}} = \frac{1}{6}$$

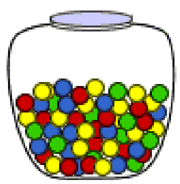
$$p(\text{even number}) = \frac{\text{number of ways to roll an even number}}{\text{total number of sides}} = \frac{3}{6} = \frac{1}{2}$$

$$p(\text{odd number}) = \frac{\text{number of ways to roll an odd number}}{\text{total number of sides}} = \frac{3}{6} = \frac{1}{2}$$

In the case of a six-sided die the outcomes are **discrete**: any one outcome can only take on the specific values 1, 2, 3, 4, 5 or 6. An outcome of, for example, 3.78 (lying between 3 and 4) is not really possible. It is also possible to get events with **continuous** outcomes, where the outcome can be any of all possible numbers in a range, including all possible fractions (for example the z-distribution, which is discussed in the next section, below).

The range of probabilities of all possible outcomes is the *probability distribution*. A die gives an example of an ***even distribution of probabilities***, because the probability of each possible outcome is the same at 1/6 (= 0.166…). All possible outcomes are *equally likely* to occur because the probability of rolling any particular number on the die is always 1/6 (unlike, for example, the normal distribution, where the probability of an outcome close to the mean is higher than the probabilities of outcomes further away from it).

Let us look at an experiment in which the outcomes are *not* equally likely.



A glass jar 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 of choosing a red marble? A green marble? A blue marble? A yellow marble?

The outcomes in this experiment are not equally likely to occur. You are more likely to choose a blue marble than any other colour. You are least likely to choose a yellow marble.

The probabilities for each colour can be calculated as follows:

p(red)	=	$\frac{\text{Number of red marbles}}{\text{Total number of marbles}}$	=	$\frac{6}{22}$
p(green)	=	$\frac{\text{Number of green marbles}}{\text{Total number of marbles}}$	=	$\frac{5}{22}$
p(blue)	=	$\frac{\text{Number of blue marbles}}{\text{Total number of marbles}}$	=	$\frac{8}{22}$
p(yellow)	=	$\frac{\text{Number of yellow marbles}}{\text{Total number of marbles}}$	=	$\frac{3}{22}$

What is the probability that a marble would be blue OR red? Since these are mutually exclusive possibilities, we can apply the additive rule directly (see section 2.1.4.1 in the PYC3704 Guide).

$$p(blue\ OR\ red) = p(blue) + p(red) = \frac{\text{Number of } blue \text{ marbles}}{\text{Total number of marbles}} + \frac{\text{Number of } red \text{ marbles}}{\text{Total number of marbles}}$$

$$= \frac{8}{22} + \frac{6}{22} = \frac{8+6}{22} = \frac{14}{22} = 0.636$$

Summary:

The probability of an event is the measure of the chance that the event will occur. The probability of an event A can be quantified as the *number of ways event A can occur* divided *by the total number of possible outcomes* and this can be estimated in an experiment by *the number of observations of event A* divided by *all the observations*.

The probability of an event A, symbolized by p(A), is always a number that measures the likelihood of an event in the following way:

- $0 \leq p(A) \leq 1$ implies that the probability of event A is a number between 0 and 1, inclusive;
- If $p(A) > p(B)$ then event A is more likely to occur than event B;
- If $p(A) = p(B)$ then events A and B are equally likely to occur.

➔ *Some questions to practice on*

Read each question below. Select your answer from the options given. Try this yourself before looking at the answers given at the end of this study letter.

Question 1.1

Which of the following experiments does **NOT** have equally likely outcomes?

1. Choose a number at random from 1 to 7
2. Toss a coin
3. Choose a letter at random from the word SCHOOL

Question 1.2

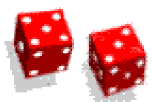
What is the probability of choosing a vowel from the alphabet?

1. 21/26
2. 5/26
3. 1/21

Question 1.3

A number from 1 to 11 is chosen at random. What is the probability of choosing an odd number?
1. 1/11
2. 5/11
3. 6/11

Question 1.4



An experimenter casts two dice, and adds the resulting outcomes for each die together after each throw. Calculate the probability that the sum of the outcomes will add up to 6.

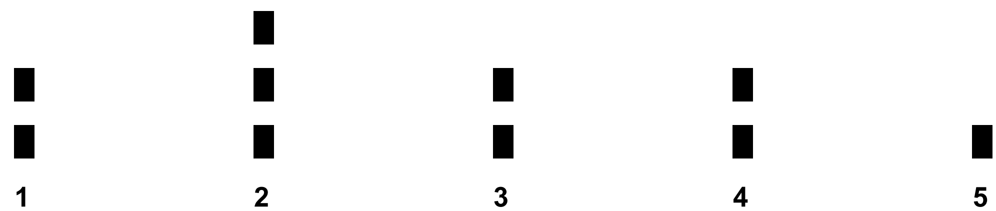
1. 0.139
2. 0.167
3. 0.500

➔ ***Note:*** *answers to these questions will be given at the end of this Study Letter.*

Probability distributions

As described above, probabilities are based on the frequencies with which certain outcomes occur relative to all events that could occur, when specific events are studied in a carefully controlled environment. When all possible probabilities for a type of event are ordered in a systematic way, we refer to it as a *probability distribution.* The probabilities in such a distribution can be estimated from data that were collected in an appropriate experiment. You would typically determine the frequencies of each type of outcome in a frequency diagram which gives you an estimate of the probability of each outcome relative to the rest.

In the drawing below, an experiment was repeated ten times. There are five possible outcomes to this experiment, numbered 1 – 5. Each time an event out of the five possibilities occurs, a block is drawn to indicate that the particular event occurred:



- ***Question:*** What is $p(x=3)$? ***Answer:*** 2/10 (event ‘3’ occurred twice, so the probability is estimated as 2 out of the 10 repetitions of the experiment)

- ***Question:*** What is $p(x \geq 2)$? ***Answer***: 8/10 (event ‘2’ plus all events greater than ‘2’ adds up to eight out of the 10 repetitions of the experiment)

The way in which probabilities was determined above is purely empirical (and is described in the Guide for PYC3704 as the relative frequency approach; see section 2.1.1.2). We use the experiment to ‘simulate’ the typical distribution of outcomes and use that as a basis for an estimate of the probability. We do not know if some kind of formal procedure or formula exists by which the precise distribution of probabilities can found.

There are however cases where the probability of a particular outcome or range of outcomes can be calculated more precisely, where a formula (or equation) was established (by statisticians), which gives a universal description of the probability distribution relative to all possible outcomes. Examples of this are the binomial distribution, used when outcomes will be one of two discrete events (section 2.2.3 in the PYC3704 Guide) and the normal distribution, where probabilities for a continuous range of possible outcomes are spread symmetrically around a mean (section 2.3.2 in the PYC3704 Guide). There are other distributions for which formulae or equations are known, and which are implied in the Guide but not explained in any depth, such as the t-distribution (an adjusted z-distribution) and the chi-square distribution.

In such cases, the probability for a specific outcome (or range of outcomes) can be directly calculated (this usually requires a computer) or read from a table which was developed by the use of a computer. It is these probabilities that are referred to as p-values and, as in the case of the distribution of frequencies of occurrences in the graph above, these probabilities are directly related to the area under a curve (in this case, a mathematical curve based on the equation that describes the probabilities of all possible outcomes). One such curve is the well-known bell-shaped normal distribution. When this distribution is transformed in such a way that its mean is always 0 and its standard deviation is always 1, it is usually referred to as the *standard normal distribution* or the *z-distribution* (see section 2.3.3 in the PYC3704 Guide).

[***Note:*** *You will not be asked questions regarding the binomial distribution in the exam. It is included in the PYC3704 Guide to demonstrate the way in which a probability distribution can be determined for discrete events*]

Exercises concerning z-scores the normal distribution

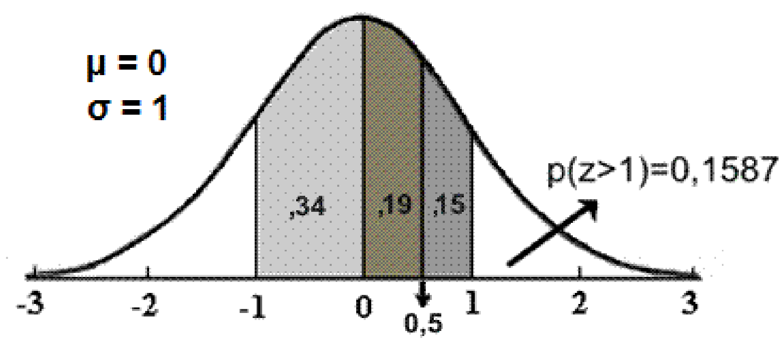
Exercise 1

Consider the standard normal distribution pictured below. Note that the horizontal axis consists of z- values (z-scores) ranging from about -3 to about +3 (Actually z-values range from minus infinity to plus infinity, but less than a quarter of a percent of the area falls outside the range of -3 to +3). For the standard normal distribution, the mean is always 0 and the standard deviation (which is not shown in the diagram) is always 1.

Note that the area between -1 and 0 is 0.34. This area represents the probability of a z-value between -1 and 0. In other words, the probability of someone obtaining a z-score between -1 and 0 is 0.34.
Since the distribution is perfectly symmetrical around the mean of 0, the probability of obtaining a z-score between 0 and +1 is also 0.34 (= 0.19 + 0.15 in the graph below). You can read this value from the z-table in your PYC3704 Study Guide (Appendix D). Find the value 1 under the column "z", and then read the value in the column "mean to z" which is 0.3413 or 0.34 when rounded off.

Question: What is the probability that someone would score between 0 and 0.5?

Solution: The probability is 0.19. See if you can find this answer from the z-table. We say: The probability that someone randomly chosen from the standard normal distribution will have a z-score of between 0 and 0.5 is 0.19. The formal symbolic way of expressing this is $p(0 \leq z \leq 0.5) = 0.19$.



Suppose the question was rather: What is the probability that someone randomly selected from this normal distributed population would score 1 or more than 1? We now have to find the area under the curve to the right of the value 1 on the horizontal axis. Think of it this way: The area under the curve to the left of 1 is larger than the area under the curve to the right of 1. The area to the left is therefore called the *larger area* and the area to the right the *smaller area*. If you go to the z-table and find the row where the z-value is equal to 1, you will find the value 0.1587 under the column marked "Smaller Portion". This gives you the probability that someone from this population would obtain a z-score of z=1 or more; that is to say, $p(z \geq 1) = 0.1587$.

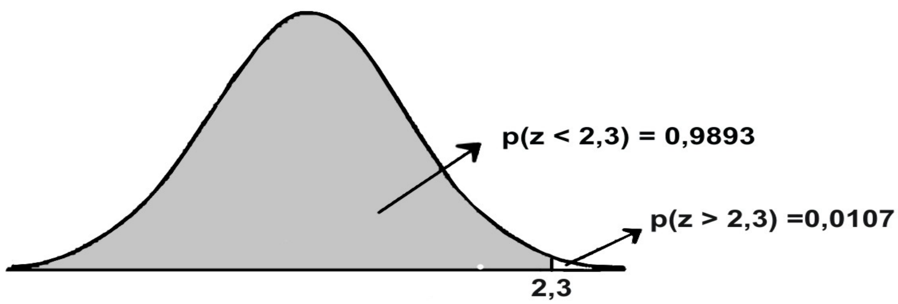
On the other hand, the probability that someone would obtain a z-score of 1 *or less* is 0.8413 (look under "Larger Portion"), so $p(z \leq 1) = 0.8413$. Note what happens if you add the two values: 0.8413 + 0.1587 = 1.000. The area under the curve to the left of the point where z = 1 plus the area to the right of this point include the total area under the bell curve, and the total area under the curve is always equal to 1. This would be true for any given value of z.

Note also that because the z-distribution is continuous, there are an infinite number of fractions between any two numbers. For this reason we do not distinguish between $z < 1$ and $z \leq 1$. The area on the left of 1 is the same irrespective of whether the number 1 itself is included. The difference (between 1 and 0.99999999999 ...) is infinitely small. It follows that $p(z \leq 1) = p(z < 1) = 0.8413$.

Exercise 2

Question: Find the probability that z is smaller than 2.30; in other words, find $p(z < 2.30)$.

Solution: The value 2.30 on the horizontal axis cuts the area under the normal distribution in a larger area to the left of 2.30 and a smaller area to the right of 2.30 (see the drawing below).



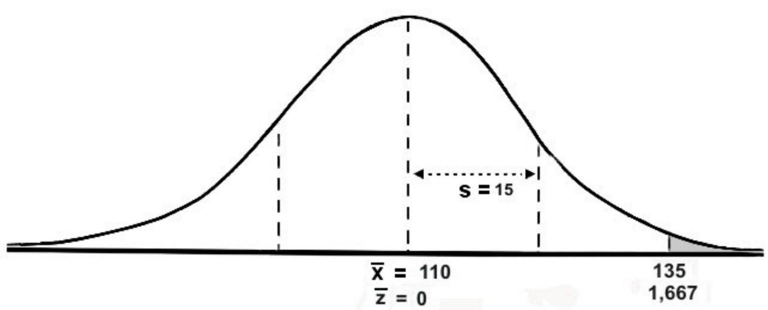
Since we are interested in the area where z is smaller than 2.3, it is the whole area to the left of that point which need to be considered (coloured grey in the figure). Therefore, we should first find the z-value of 2.30 in the z-table (in Appendix D in the PYC3704 Guide) and then read off the probability value for the "larger portion," which is 0.9893.

Note that the "smaller portion" is 0.0107 and that 0.9893 + 0.0107 = 1.000. The whole area under the normal curve always adds up to 1.

Exercise 3

Question: Suppose IQ scores of university students are distributed normally with a mean of 110 and a standard deviation of 15. What is the probability that a randomly selected student would have an IQ of greater than 135?

Solution: In cases like this, it is useful to first make a drawing of the problem, as follows:



Since we are interested in what the probability of an x-value of *above* 135 would be, we need to consider the area under the curve to the *right* of 135 (coloured grey in the figure).

When a problem is formulated in terms of a variable which is normally distributed with a known mean and standard deviation, the first step is to convert it into a z-score, so that we can use the z-tables. So before we can look up this probability, we need to convert the given x-value values of 135 to a z-score:

$$z_{135} = \frac{x - \bar{x}}{s} = \frac{135 - 110}{15} = \frac{25}{15} = 1{,}667$$

This tells us that an x-score of 135 is equivalent to a z-score of 1.667. Since we are only interested in the small area to the right of the mark where z = 1.667, we look up the probability for the smaller area in the table of z-values, to find that p(x>135) = p(z>1.667) = 0.048. So we can conclude that the probability that a randomly selected student would have an IQ of greater than 135 is 0.048 or 4.8%.

➔ *A scenario with some practice exercises*

An instructor in the Underwater Demolition Institute (UDI) believes that he has developed a new technique for staying under water longer. The head of the institute gives the instructor permission to try his new technique with a student who has been randomly selected from the current team. As part of their qualifying exam, all students are tested to see how long they can stay under water without an air tank. Past records show that the scores are normally distributed with a mean of 130 seconds and a standard deviation of 14. Calculate what the probabilities in the two questions below would be, if the new technique has no additional effect.

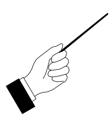
Question 2.1

Calculate the probability that the randomly selected students will stay under water for over 150 seconds.

Question 2.2

Calculate the probability that the randomly selected students will stay under water for between 115 and 135 seconds.

➔ ***Note:*** *answers to these questions will be given at the end of this Study Letter.*



Exercises related to the process of testing hypotheses using inferential statistics

First, let us consider a summary of the process that is involved in testing statistical hypotheses (as set out in detail in the Study Guide for PYC3704). In quantitative research we would normally begin with a research question, conceived as a possible relationship among variables (where variables are measurements of so-called 'constructs'; see Topic 1 in the Guide).

The general steps to follow are:

- *Specify the specific statistical hypotheses that need to be tested;*
- *Select an appropriate test statistic to test the hypotheses and calculate it;*
- *Determine the p-value and use this to evaluate the hypotheses.*

In setting up the formal hypotheses, the possibility that no relationship exists (a null hypothesis) is contrasted with the possibility that some suspected relationship does in fact exist. The actual form of the hypothesis depends on issues such as the nature of the research design and the measurements that could be made.

Some possibilities that were covered in this course include (and there are of course many other possibilities):

- *A measurement parameter (such as a mean) compared with a constant value (see Topic 4 in the V)*

- *A parameter compared between two samples, for example two means. The samples can be independent or dependent. In the latter case you will find that the measurements have to represent matched pairs of data. (See Topic 5 in the Study Guide for PYC3704 on the comparison of samples))*

- *Two measurements consisting of two different variables from a single sample which are compared (see the discussion of relationships between pairs of variables in Topic 6).*

A further consideration in setting up hypotheses is the form of the postulated relationship, which will influence whether the null hypothesis is contrasted with a non-directional hypothesis (a relationship exists but it is unclear in which direction) or a directional hypothesis (a relationship in a specific direction is predicted).

The testing procedure consists of making an observation of the relationship that exists in data from a sample of measurements. You then calculate a test statistic which is a comparison of the situation which you have observed in the sample (the relationship among variables or the pattern you are looking for), and the situation as predicted in the null hypothesis (what you should find if there is in fact no relationship among the variables).

The choice of test statistic will depend on the *kind of hypotheses* you are testing and how much information you have about issues such as the *distributions* of your variables (which will affect the distribution of the test statistic), the *parametric qualities* of the variables, (for example, whether the population mean and/or population standard deviation are known) and the *sample size*.

Following this, you have to determine how probable this value of the test statistic would be if the relationship which you observe among the variables in your sample of data is merely a chance effect. This probability is the p-value, which in effect tells you the probability of finding this observed relationship between variables in your sample, if in fact no relationship exists (i.e. if H_0 is true).

You have to set a cut-off limit to indicate the lowest highest probability of the null hypothesis being true which you are willing to accept (called the level of significance and symbolized by α). If the calculated p-value is lower than this, it is accepted that the null hypothesis is probably false. So significance level represents the maximum risk which you are willing to take that the effect (some kind of relationship among variables) which you found in the sample data is due purely to chance (i.e. the *maximum* probability of making an error of Type I if you reject the null hypothesis that you would be willing to take; see sections 3.3.1 and 3.3.2 in the PYC3704 Guide).

The example of a research scenario that is presented below is followed by a number of questions. Along with similar problems in the PYC3704 Guide and Assignments, this will show you the type of problems you have to be able to solve.

➔ *A scenario with some questions to practice on*

A researcher hypothesizes that people in the acting profession should do better on a test for memory than people in general, because their profession requires that they memorize a lot of text. She draws a random sample of 30 actors from various dramatic societies and actors' training schools and measures their memories by means of a standardized memory test developed. The scores of the general population, as based on a large sample during test construction, were found to be normally distributed with a mean of 52. Suppose the researcher finds that the mean score of his sample of actors is 59.5, and the standard deviation of the scores is 12.2.

Question 3.1

Which research design did the researcher use?

1. Single-sample groups design
2. Two-groups design
3. Two-groups design with a known population mean

Question 3.2

Which is the appropriate alternative hypothesis for testing the research hypothesis?

1. H_1: $\overline{x} > 52$
2. H_1: $\mu > 52$
3. H_1: $\mu \neq 52$

Question 3.3

Which is the appropriate test statistic to calculate?

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

Question 3.4

What value of the standard deviation of the sampling distribution of the mean of 30 memory scores does the researcher use when calculating the test statistic?

1. 0.41
2. 2.23
3. 12.2

Question 3.5

Calculate the appropriate test statistic.

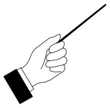
1. 8.45
2. 3.37
3. 0.62

Question 3.6

Suppose the above value of the test statistic is found to lead to a two-tailed p-value of 0.002. What would you conclude if you decide to test to the 1% level of significance?

1. Reject the null hypothesis and conclude that actors have better memories than the rest of the population
2. Reject the alternative hypothesis, so no conclusion can be made about whether actors have better memories than the rest of the population
3. Reject the null hypothesis and conclude that actors do not have better memories than the rest of the population
4.

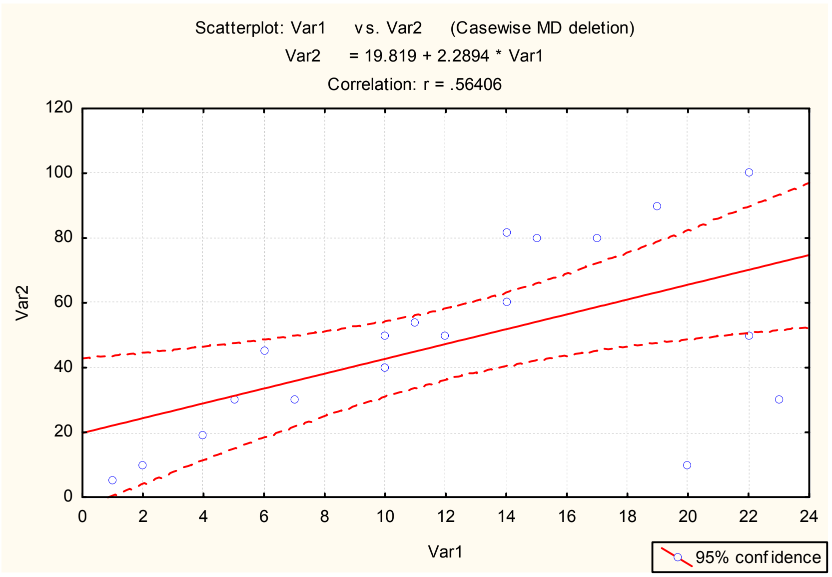
➔ ***Note:*** *answers to these questions will be given at the end of this Study Letter.*



Exercises concerning correlation

The term ***correlation*** refers to the strength of the relationship between (two or more) variables and a ***correlation coefficient*** is a number that indicates the strength of the relationship between variables. The most commonly used correlation coefficient is the ***Pearson product-moment correlation coefficient,*** which is discussed in Topic 6 of the PYC3704 Guide.

This correlation coefficient is indicated by a small letter 'r', and varies from -1 (a negative correlation) through 0 (no correlation at all) up to 1 (a perfect positive correlation). A strong correlation implies that as the one variable changes, the other one changes also, to a similar degree. A positive correlation means that as one gets larger, so does the other, while a negative correlation means that as one gets larger, the other gets smaller.



Look at the graph, which was created by a computer program to describe the relationship between two variables (called *Var1* and *Var2*). The correlation coefficient in this case was calculated to be r = 0.56.

The straight line describes the *linear relationship* between the two variables. It is called the *regression line* and it implies that if you have the value for *Var1*, you could calculate what *Var2* would be. (The two dotted curved lines indicate that 95% of the data falls between them; but this is not very important at the moment).

This is a *positive* relationship. You can see this by the sign of the r-value (r=+0.56); but, more importantly, by the fact that the line moves from lower left to upper right in the graph. A negative relationship would run from upper left to lower right (see the example below, as well as p. 133 of the PYC3704 Guide).

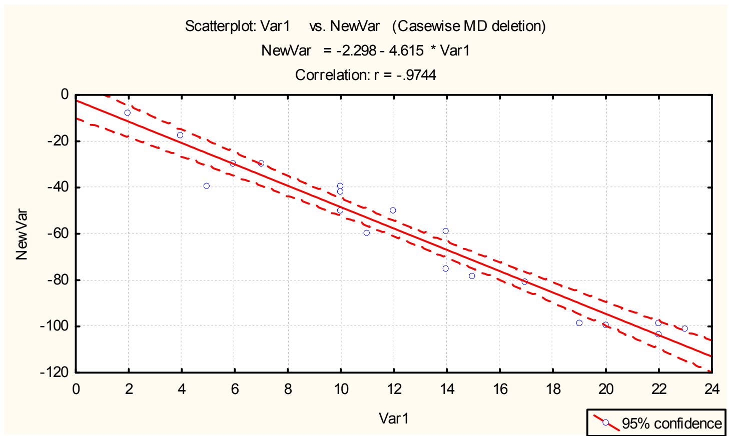
Now consider how the data is plotted (see pp. 130-1 of the Guide).

Take a number like *Var1*=2. If you drew a vertical like upwards from *Var1*=2, the line would cut the graph where *Var2*=27 (more or less). Now imagine that the Var1 value increases, for example to *Var1*=6. *What happens to* Var2? You should find that the place where a vertical line cuts the graph is at *Var2*=34 (more or less). Which is why we say, as one variable gets bigger, so does the other, if the correlation is positive.

What would happen if *Var1* got smaller? Take *Var1*=10, then *Var2*=43 (more or less). Now take a smaller value of *Var1*; anything smaller than 10 would do. *What happens to* Var2? At what point on the y-axis (*Var2*) would a vertical line from this smaller *Var1*-value cut the regression line?

This procedure also works if you exchange the two axes. The x-axis is the independent value and the y-axis is the dependent value, but this is just a convention. If you put *Var1* on the x-axis and *Var2* on the y axis the result (and conclusion) would be the same.

Below is an example where the correlation is negative (i.e., r < 0):



See what happens to *NewVar* (on the vertical axis) if you increase the x-axis value (*Var1*), and if you decrease the *Var1* value? (Keep in mind that, for example, -120 is *smaller than* -100).

➔ ***Some questions to practice on relating to correlation***

Question 4.1

Which of the possible results given below can you expect if you calculate Pearson's r between the variables X and Y in the table?

X	2	5	7	2	9	12	3	5	10	14
Y	6	16	20	5	25	39	10	16	30	42

1. r < 0
2. r > 0
3. r ≈ 0 *

[* *Note:* The symbol '≈' is interpreted as 'approximately' (see Appendix E in the PYC3704 Guide)]

Question 4.2

Which of the possibilities given below are closest to the expected Pearson's r between the variables X and Y in the table?

X	2	2	2	2	2	2	2	2	2	2
Y	15	16	17	18	19	20	21	22	23	24

1. 1.0
2. -1.0
3. 0.0

ANSWERS TO THE EXCERCISE QUESTIONS

Answer to Question 1.1

Option 3 is the correct answer. All possible choices are not equally probable when you choose a letter at random out of the sequence of letters in 'SCHOOL,' since the letter 'O' appears twice.

Answer to Question 1.2

Option 2 is correct. The vowels are A, E, I, O, U, so there are 5 in total in the alphabet of 26 characters – so the answer is 5/26.

Answer to Question 1.3

The answer is alternative 3, that is, 6/11. There are 6 odd numbers in the sequence of whole numbers (integers) from 1 to 11, as follows:

1 Odd	2 Even	3 Odd	4 Even	5 Odd	6 Even	7 Odd	8 Even	9 Odd	10 Even	11 Odd

Answer to Question 1.4

The correct answer is option 1. There are 36 possible outcomes when a pair of dice is thrown, as shown below (in the first two columns):

Die 1	Die 2	Sum
1	1	2
1	2	3
1	3	4
1	4	5
1	**5**	**6**
1	6	7
2	1	3
2	2	4
2	3	5
2	**4**	**6**
2	5	7
2	6	8

Die 1	Die 2	Sum
3	1	4
3	2	5
3	**3**	**6**
3	4	7
3	5	8
3	6	9
4	1	5
4	**2**	**6**
4	3	7
4	4	8
4	5	9
4	6	10

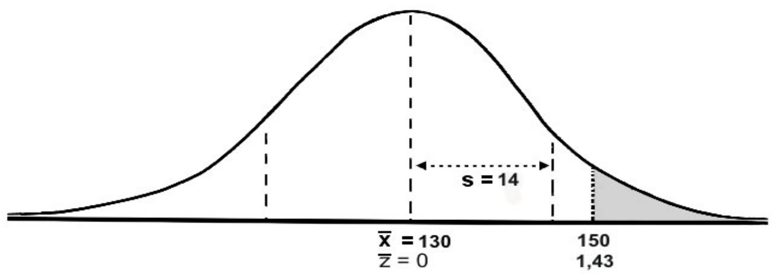
Die 1	Die 2	Sum
5	**1**	**6**
5	2	7
5	3	8
5	4	9
5	5	10
5	6	11
6	1	7
6	2	8
6	3	9
6	4	10
6	5	11
6	6	12

To calculate the probability that the sum of the two dice will equal 6, calculate the number of outcomes that sum to 6 (marked in **bold** in the table) and divide by the total number of outcomes (36).

Since *five* of the outcomes have a total of 6 (1 + 5; 2 + 4; 3 + 3; 4 + 2; 5 + 1), the probability of the two dice adding up to 6 is 5/36 = 0.139.

Answer to Question 2.1

It is useful to start problems like this by drawing a graph of the normal distribution, like the one below.



By putting the mean value of x-as $\bar{x}$=130 and the value of x = 150 in the appropriate positions on the x-axis, you can see that the area under the curve that needs to be determined is at the far right of the figure (the grey area which indicates where x = 150 or larger). To find the probability associated with this area, we first have to convert the raw x-value to a z-value, like this:

$$z = \frac{x - \bar{x}}{s} = \frac{150 - 130}{14} = \frac{20}{14} = 1.43$$

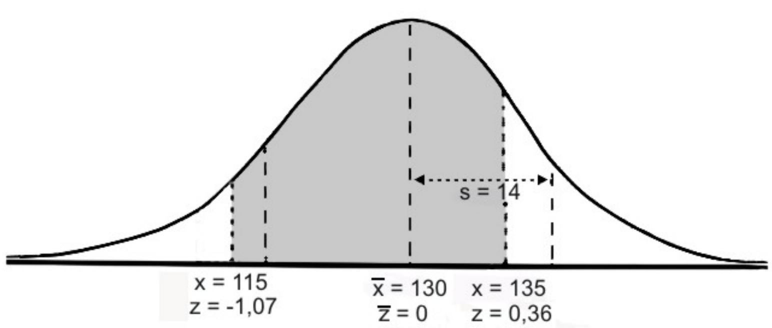
Now look up the probability associated with z = 1.43 in your z-table in the PYC3704 Guide (from the graph you will see that the grey area is the *smaller* portion; this is why drawing a picture is useful).

You should find that the answer is 0.0764. In other words, the probability of selecting a student that can stay under water purely at random is only 7.64% (0.0764 X 100). In formal terms, we would write this as $p(x > 150) = 0.0764$.

Note: in the exam, a question like this one will of course be given in the form of a multiple choice question. There will be four answers presented to you, of which you have to choose the one closest to the correct one (keep in mind when working with calculations that slight differences are always possible due to rounding-off errors).

Answer to Question 2.2

Once again, it would be a good idea to begin by drawing a picture of the normal distribution, like the one below.



Given that the mean is 130 ($\bar{x}$ = 130) and the standard deviation is s = 14, you can now estimate roughly where the values x = 115 and x = 135 would lie on the graph. The probability that a value for a particular value x lies between the values 115 and 135 (which can be written as $p(115 < x < 135)$, that is, x is larger than 115 but smaller than 135) will be equivalent to the grey area in the graph above.
To find out what this probability is from the tables, you first have to transform each value of x into its corresponding z-value.

The two calculations are as follows:

$$z_1 = \frac{x_1 - \bar{x}}{s} = \frac{115 - 130}{14} = \frac{-15}{14} = -1.07$$

$$z_2 = \frac{x_2 - \bar{x}}{s} = \frac{135 - 130}{14} = \frac{5}{14} = 0.36$$

It follows that p(115 < x < 135) = p(-1.07 < z < 0.36), that is to say, the probability that x lies between 115 and 135 is the same as the probability that z lies between -1.07 and 0.36. To find out what this is, first look up the value of p(z < 0.36), which is 0.6406. This gives you the area from minus infinity (the far left of the graph) up to z = 0.36 (you can see from the graph that you need to look at the z-values for the **large** area under the curve).

Now you look up the area from minus infinity up to z = - 1.07 (the leftmost 'triangle' under the curve) which gives you p(z<-1.07) = 0.1423 (can you see that we are interested in the **smaller** area in the z tables this time? Also, remember that you can look up the value for the smaller section at z = +1.07 in the tables, because the probability value for positive and negative values of z will be the same). Since we are NOT interested in this area (it is not part of the grey area), we have to subtract this from the previous result. This will give us the desired probability value.

So we get

p(115 < x < 135) = p(-1.07 < z < 0.36)
= p(z < 0.36) – p(z < -1.07) = 0.6406 – 0.1423
= 0.4983.

It follows that the probability that a value for x lies between the values 115 and 135 is p = 0.498; or a probability of 49.8%.

Answer to Question 3.1

The correct option is 1. There is only one group, the actors, being compared with the population mean. Determining 'the scores of the general population' does not form part of this particular research project, but refers to data that is already available. Here the goal is to compare a sample of data (obtained from the actors) with an existing value, the population mean, which can be estimated to be 52, based on previous research. The population standard deviation is not given. So this is a group mean compared to a constant value of 52, and not two groups compared with each other. Therefore it is a single-samples group design (see section 4.1.2 in the PYC3704 Guide. The expression '*single-sample groups design*' is explained on p. 101).

Answer to Question 3.2

Option 2 is the correct answer. The research question is that people in the acting profession should do better on a test for memory than people in general. We know that people in general have a score of 52, so we want to know whether the score of the sample of actors is significantly greater than 52, which implies a directional or one-tailed test. This eliminates option 3, which states that actors' scores would differ significantly, but gives no information on the direction of the difference (lesser or greater). Option 1 can be eliminated because, although we calculate the mean from the sample, we are testing a hypothesis in relation to the population (the sample data acts as representation of the population), so we should use the symbol μ for the population mean.

Answer to Question 3.3

The correct answer is option 3. From the answer to Question 3.1 above, we know we are testing for a single-samples group design, which eliminates option 2. The z-statistic for the mean of a single sample can only be used when the population standard deviation σ is known (see section 4.1.2 in the PYC3704 Study Guide), but it was not provided. Therefore option 1 cannot be the correct answer.

Answer to Question 3.4

The correct answer is option 2. To estimate the standard deviation of the sampling distribution of the mean of the 30 actors' memory scores, we use the standard error s (see p. 61 and p. 105 in the PYC3704 Study Guide). Here, ***s*** is the sample standard deviation, given above as s = 12.2, and ***n*** is the sample size, given as 30.

The calculation is as follows: $s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{12.2}{\sqrt{30}} = \frac{12.2}{5.48} = 2.226$

This makes option 2 the best answer (the slight difference is due to rounding off).

You will get option 1 if you calculate 12.2/30 (forgetting to take the square root), and option 3 is the sample standard deviation s.

Answer to Question 3.5

Option 2 is the correct answer. We know which t-test to calculate from question 3.3. The formula for the t-statistic for the mean of a single sample is given in the Study Guide for PYC3704 on p. 104:

$$t_{\bar{x}} = \frac{(\bar{x} - \mu_{\bar{x}})}{s_{\bar{x}}}$$

We know what the value of $s_{\bar{x}}$ is from the answer to Question 3.4, and the other values are provided in the scenario, so this t-value can be worked out quite easily as follows:

$$t_{\bar{x}} = \frac{(59.5 - 52)}{2.226} = \frac{7.5}{2.226} = 3.369$$

[*Note that we will provide these formulas for the exam, but we might not tell you which is which*]

Answer to Question 3.6

The correct alternative is option 1. Testing to the 1% level of significance implies that the cut-off point for the p-value is set to a value of α=0.01. Note that we require a one-tailed p-value for this test (the hypothesis in Question 3.2 is a directional hypothesis), so to get the one-tailed p-value we first have to divide the two-tailed p-value by 2: p-value = 0.002/2 = 0.001.

Since this p-value is less than the specified level of significance (i.e., $p < \alpha$), the null hypothesis can be rejected in favour of the alternative hypothesis, and we can conclude that people in the acting profession do better than the general population on the memory test. Note that in option 3 the decision to reject the null hypothesis is correct but the conclusion is wrong.

Answer to Question 4.1

Option 2 is correct. It is not really necessary to calculate r to see this. If you take the values of Y and arrange them in a rising order (along with matching values of X) you get the following:

Y	2	2	3	5	5	7	9	10	12	14
X	5	6	10	16	16	20	25	30	39	42

It is clear from this that as X gets larger, Y also rises (or *vice versa*). This implies that a correlation exists, and that it is positive (which can also be written as $r > 0$, which means r greater than 1, or a positive value for r). If you do calculate the Pearson correlation coefficient you should find that it is very close to a perfect positive correlation, with $r = 0.994$.

Answer to Question 4.2

From the data in the table it is clear that variable Y does not change. This implies that changes in the size of X has no effect on the size of Y (as X changes, Y remains the same). Correlation measures the extent to which one variable affects the other, and here it is clear that variable X does not affect variable Y at all.

Another way of coming to the same conclusion is by using the fact that variable Y has a variance of zero (and therefore a standard deviation of zero). So even though one of the variables (variable X) has a variance, we can never find a correlation between two variables when one of them has no variance. If you try to calculate this value of Pearson's r with the formula on p. 134 in the PYC3704 Guide, you will end up trying to divide by zero, which is impossible. The answer is therefore Option 3, namely $r = 0.0$. Note that if variables X and Y were interchanged, the implication would be the same.

4. Some of the symbols used in statistics

This table may help you with the interpretation and pronunciation of some symbols used in the module

Symbol	Pronunciation and meaning
μ	muu (population mean)
$\bar{x}$	x-bar (sample mean)
σ	sigma (population standard deviation)
χ	chi (χ^2 is used to refer to the chi-square test)
α	alpha (level of significance)
ρ	rho (population correlation coefficient)
$\sum$	capital letter sigma (used to indicate a sum or total of a number of values)
$\neq$	'not equal to'
$<$	'smaller than'
$>$	'greater than'
$\leq$	'smaller than or equal to'
$\geq$	'greater than or equal to'
X	'multiplied by'. But if x and y are two variables, xy implies 'x multiplied with y'
/	'divided by'
$\sqrt{n}$	square root of n
H_0	Null hypothesis
H_1	Alternative hypothesis
$\|$	'given that'; the horizontal bar is used to indicate a conditional statement – so 'A|B' means 'A on condition of B')

The PYC3704 team wishes you all the best for the exams!