**An Adventure in Statistics: The Reality Enigma**

This testbank has been adapted from *Discovering Statistics Using IBM SPSS 4e* and has been designed to be used in conjunction with Field, A. P. (2016) *An Adventure in Statistics: The Reality Enigma*. Select questions come with additional feedback on why a particular answer is correct.

**Chapter 1**

1. The purpose of a control condition is to:
2. Show up relationships between predictor variables.
3. Control for participant characteristics.
4. Rule out a *tertium quid*.
5. Allow inferences about cause.

The correct answer is d) Allow inferences about cause.

Feedback: A properly constructed control condition provides you with a reference point to determine what change (if any) occurred when a variable was modified

1. If the scores on a test have a mean of 26 and a standard deviation of 4, what is the *z*-score for a score of 18?
2. -2
3. 2
4. 11
5. -1.41

The correct answer is a) -2.

Feedback: The z scores are calculated by subtracting the mean from the score (18 – 26) = –8 and dividing your answer by the standard deviation (–8 / 4) = –2

1. What is a scientific journal?
2. A piece of scientific research that has not yet been published.
3. A collection of articles written by scientists that have been peer reviewed.
4. A notebook kept by scientists containing important details of all their own experimental research for future reference.
5. A collection of articles written by scientists that have not yet been reviewed by other scientists in the field.

The correct answer is b) A collection of articles written by scientists that have been peer reviewed.

Feedback: Scientific journals contain articles that have been peer reviewed, in an attempt to ensure that articles meet the journal's standards of quality, and scientific validity.

1. In general, as the sample size (*N)* increases:
2. The confidence interval gets wider.
3. The confidence interval gets narrower.
4. The confidence interval is unaffected.
5. The confidence interval becomes less accurate.

The correct answer is b) The confidence interval gets narrower.

1. Confidence intervals:
2. Can be used instead of conventional statistics based on point estimates.
3. Are not frequently used in research articles because they can mislead the reader.
4. Are constructed using subjective evaluations of confidence.
5. None of these options are correct.

The correct answer is a) Confidence intervals can be used instead of conventional statistics based on point estimates.

1. In which of the following situations is the assumption of normality least important?
2. If you have a small sample.
3. If you want to construct confidence intervals around the parameter estimates of your model.
4. If you want only to estimate the parameters of your model.
5. If you want to compute significance tests relating to the parameter estimates of your model.

The correct answer is c) If you want only to estimate the parameters of your model.

1. An experiment was conducted to see how people with eating disorders differ in their need to exert control in different domains. Participants were classified as not having an eating disorder (control), as having anorexia nervosa (anorexic), or as having bulimia nervosa (bulimic). Each participant underwent an experiment that indicated how much they felt the need to exert control in three domains: eating, friendships and the physical world (this final category was a control domain in which the need to have control over things like gravity or the weather was assessed). So all participants gave three responses in the form of a mean reaction time; a low reaction time meant that the person did feel the need to exert control in that domain. The variables have been labelled as **group** (control, anorexic, or bulimic) and **domain** (food, friends, or physical laws). Of the following options, which analysis should be conducted?
2. Two-way repeated-measures ANOVA.
3. Three-way independent ANOVA
4. Two-way mixed ANOVA
5. Analysis of covariance

The correct answer is c) Two-way mixed ANOVA.

Feedback: **Group** is a between-subjects variable and **domain** is a within-subjects variable.

1. An experiment was conducted to see how people with eating disorders differ in their need to exert control in different domains. Participants were classified as not having an eating disorder (control), as having anorexia nervosa (anorexic), or as having bulimia nervosa (bulimic). Each participant underwent an experiment that indicated how much they felt the need to exert control in three domains: eating, friendships and the physical world (this final category was a control domain in which the need to have control over things like gravity or the weather was assessed). So all participants gave three responses in the form of a mean reaction time; a low reaction time meant that the person did feel the need to exert control in that domain. The variables have been labelled as **group** (control, anorexic, or bulimic) and **domain** (food, friends, or physical laws). Looking at the output below, what can we conclude about the main effect of the **domain** variable?



1. There was a significant effect of domain, *F*(2, 54) = 8.02, *p* < .01, on the degree to which people felt the need to exert control.
2. There was not a significant effect of domain, *F*(2, 54) = 8.02, *p* < .05, on the degree to which people felt the need to exert control.
3. People with eating disorders need to exert more control over different domains of their life compared to controls, *F*(1.55, 41.89) = 8.02, *p* < .01.
4. There was a significant effect of domain, *F*(1.55, 41.89) = 8.02, *p* = .001, on the degree to which people felt the need to exert control.

The correct answer is d) There was a significant effect of domain, *F*(1.55, 41.89) = 8.02, *p* = .001, on the degree to which people felt the need to exert control.

Feedback: The significant result for Mauchly’s test indicates that the assumption of sphericity has not been met. However, both the Greenhouse–Geisser and Huynh–Feldt corrections were significant.

**Chapter 2**

1. When the results of an experiment can be applied to real-world conditions, that experiment is said to have:
2. Criterion validity
3. Ecological validity
4. Content validity
5. Factorial validity

The correct answer is b) Ecological validity

Feedback: For a research study to possess ecological validity, the methods, materials and setting of the study must approximate the real-life situation that is under investigation

1. A variable manipulated by a researcher is known as:
2. An independent variable
3. A dependent variable
4. A confounding variable
5. A discrete variable

The correct answer is a) An independent variable

Feedback: An independent variable (or predictor variable) is a variable that is thought to be the cause of some effect. This term is usually used in experimental research to denote a variable that the experimenter has manipulated.

1. A frequency distribution in which low scores are most frequent (i.e. bars on the graph are highest on the **left** hand side) is said to be:
2. Leptokurtic
3. Platykurtic
4. Positively skewed
5. Negatively skewed

The correct answer is c) Positively skewed

Feedback: In a positively skewed distribution the frequent scores are clustered at the lower end and the tail points towards the higher or more positive scores.

1. A frequency distribution in which high scores are most frequent (i.e. bars on the graph are highest on the **right** hand side) is said to be:
2. Negatively skewed
3. Positively skewed
4. Leptokurtic
5. Platykurtic

The correct answer is a) Negatively skewed

Feedback: In a negatively skewed distribution the frequent scores are clustered at the higher end and the tail points towards the lower or more negative scores.

1. Under a null hypothesis, a sample value yields a *p*-value of .015. Which of the following statements is true?
2. This finding is statistically significant at the .01 level of significance.
3. This finding is statistically significant at the .05 level of significance.
4. This finding is statistically significant at the .001 level of significance.
5. This finding is not statistically significant.

The correct answer is b) This finding is statistically significant at the .05 level of significance.

**Chapter 3**

1. Imagine we took a group of smokers, recorded how many cigarettes they smoked each day and then split them randomly into one of two 6-week interventions; ‘hypnosis’ or ‘nicotine patch’. After the 6 weeks, we again recorded how many cigarettes they smoked each day and subtracted this number from the number of cigarettes they each smoked pre-intervention, to produce a intervention success score for each participant. Out of the following options, which would be the best method of displaying the results?
2. A simple bar chart with the variable ‘intervention method’ on the *y*-axis and ‘intervention success’ on the *x*-axis.
3. A simple boxplot with the variable ‘intervention method’ on the y-axis and ‘intervention success’ on the *x*-axis.
4. A clustered boxplot with ‘intervention success’ on the *y*-axis and ‘intervention method’ on the *x*-axis
5. A simple boxplot with the variable ‘intervention method’ on the *x*-axis and ‘intervention success’ on the *y*-axis.

The correct answer is d) A simple boxplot with the variable ‘intervention method’ on the *x*-axis and ‘intervention success’ on the *y*-axis.

Feedback: This option is used when you want to plot a boxplot of a single variable (in this case success), but you want different boxplots produced for different categories in the data (for these success data we could produce separate boxplots for our two intervention groups).

1. Looking at the table below, which of the following statements is the most accurate? (*Hint*: The further the values of skewness and kurtosis are from zero, the more likely it is that the data are not normally distributed.)



1. For the number of hours spent practising, the data are fairly positively skewed.
2. For the level of musical skill, the data are heavily negatively skewed.
3. For the number of hours spent practising, there is an issue with kurtosis.
4. For the number of hours spent practising, there is not an issue with kurtosis.

The correct answer is d) For the number of hours spent practising, there is not an issue with kurtosis.

1. Looking at the table below, which of the following statements is correct?



1. Levene’s test was significant, *F*(1, 118) = 0.93, *p* = .007, indicating that the assumption of homogenity of variance had been met.
2. Levene’s test was non-significant, *F*(1, 118) = 0.01, *p* = .93, indicating that the assumption of homogenity of variance had been met.
3. Levene’s test was non-significant, *F*(1, 118) = 0.01, *p* = .93, indicating that the assumption of homogenity of variance had been violated.
4. Levene’s test was significant, *F*(1, 118) = 0.01, *p* = .93, indicating that the assumption of homogenity of variance had been violated.

The correct answer is b) Levene’s test was non-significant, *F*(1, 118) = 0.01, *p* = .93, indicating that the assumption of homogenity of variance had been met.

1. When it is not necessary to use Levene’s test?
2. When you have equal group sizes.
3. When you have unequal group sizes.
4. When you have a small sample.
5. When you are conducting a two-tailed test.

The correct answer is a) When you have equal group sizes.

Feedback: If you don’t have unequal group sizes, the assumption of homogeneity of variance is pretty much irrelevant and can be ignored.

**Chapter 5**

1. Looking at the graph below, which of the following statements are correct? (*Hint*: Look at the bars – are they in the same direction?)



1. On average, the nicotine intervention was more successful in those who wanted to quit smoking than in those who did not want to quit, whereas the hypnosis intervention was more successful in those who did not want to quit smoking than in those who did
2. On average, for those who wanted to quit smoking, the nicotine patches reduced the number of cigarettes smoked per day, whereas hypnosis actually increased the number of cigarettes smoked per day.
3. Overall, the nicotine intervention was the most successful at reducing the number of cigarettes smoked per day.
4. All of the statements are correct.

The correct answer is d) All of the statements are correct.

1. What is the graph below known as?



1. Stacked bar chart
2. Stacked histogram
3. Frequency polygon
4. Population pyramid

The correct answer is d) Population pyramid.

1. Which of the following statements best describes the graph below?



1. The graph shows that for those who used nicotine patches there is a fairly normal distribution, whereas those who used hypnosis show a skewed distribution, where a very small proportion of people (relative to those using nicotine) smoke more than 2 cigarettes per day.
2. The graph looks fairly symmetrical. This indicates that both groups had a similar spread of scores before the intervention.
3. The graph shows that for those who used hypnosis there is a fairly normal distribution, whereas those who used nicotine patches show a skewed distribution, where a very large proportion of people (relative to those using nicotine) smoke less than 4 cigarettes per day.
4. The graph looks fairly unsymmetrical, indicating that the two groups are from different populations.

The correct answer is b) The graph looks fairly symmetrical. This indicates that both groups had a similar spread of scores before the intervention.

Feedback: A population pyramid can be a very good way to visualize differences in distributions in different groups

1. Approximately what is the mean success score for those who wanted to quit in the hypnosis group?



1. 1.00
2. –1.00
3. 0.00
4. The graph does not display the mean.

The correct answer is d) The graph does not display the mean.

1. What can we say about the graph below?



1. There is a negative relationship between the number of cigarettes smoked per day before the intervention and the number of cigarettes smoked after the intervention
2. The participants who smoked the most cigarettes per day before the intervention, smoked the fewest cigarettes per day after the intervention.
3. There is a positive relationship between the number of cigarettes smoked per day before the intervention and the number of cigarettes smoked after the intervention.
4. There is no relationship between the two variables

The correct answer is c) There is a positive relationship between the number of cigarettes smoked per day before the intervention and the number of cigarettes smoked after the intervention.

Feedback: The scatterplot indicates that the more cigarettes people smoked before the intervention, the more cigarettes they smoked after the intervention, which makes sense.

1. A study was done to investigate the effect of ‘motivation to quit’ on the success rate of a new intervention developed to reduce the number of cigarettes smoked per day in a group of smokers. Looking at the graph below, what can we say about the relationship between motivation to quit and the success rate of the intervention?



1. The medians were the same in people who wanted to quit smoking and those that didn’t.
2. We can’t say anything about the success of the intervention because the graph does not take into account the number of cigarettes smoked per day pre-intervention.
3. Whether a person wanted to quit smoking had no effect on the success of the smoking intervention.
4. There was the same number of people who wanted to quit smoking as who didn’t.

The correct answer is b) We can’t say anything about the success of the intervention because the graph does not take into account the number of cigarettes smoked per day pre-intervention.

1. In IBM SPSS, the following graph is known as a:



1. Summary point plot
2. Simple scatterplot
3. Scatterplot matrix
4. Grouped scatterplot

The correct answer is d) Grouped scatterplot

1. We took a sample of children who had been learning to play a musical instrument for five years. We measured the number of hours they spent practising each week and assessed their musical skill by how many of 8 increasingly difficult exams they had passed. We also asked them whether their parents forced them to practise or not (were their parents pushy?). What does the following graph show?



1. The more time spent practising, the more musically skilled the children were and this relationship was stronger for those who had pushy parents compared to those who did not.
2. Practice causes better exam performance.
3. Children with pushy parents always passed more grade exams than those without.
4. The more time spent practising, the more musically skilled the children were, and this relationship was stronger for children who did not have pushy parents than for those who did.

The correct answer is a) The more time spent practising, the more musically skilled the children were and this relationship was stronger for those who had pushy parents compared to those who did not.

1. The graph below shows the mean success rate of cutting down on smoking (positive score = success) in people who wanted to quit and people who did not want to quit. Which of the following statements is the most true?



1. On average, people who wanted to quit were 25 times more successful than those who did not.
2. On average, success was six times higher in people who wanted to quit than in those who did not.
3. The effect in the population is likely to be the same for those who did and did not want to quit.
4. The average success was significantly higher in people who wanted to quit.

The correct answer is c) The effect in the population is likely to be the same for those who did and did not want to quit.

Feedback: This is correct because the confidence intervals almost entirely overlap.

1. Which of the following statements about Pearson’s correlation coefficient is **not** true?
2. It cannot be used with binary variables (those taking on a value of 0 or 1).
3. It can be used as an effect size measure
4. It varies between –1 and +1.
5. It can be used on ranked data.

The correct answer is a) It cannot be used with binary variables (those taking on a value of 0 or 1).

Feedback: Pearson’s correlation coefficient can be used with binary variables (or categorical variables).

1. A paired-samples *t*-test is used to test for?
2. Differences between means of groups containing different entities when the sampling distribution is normally distributed, and the data have equal variances and are at least interval.
3. Differences between means of groups containing different entities when the sampling distribution is not normally distributed.
4. Differences between means of groups containing the same entities when the sampling distribution is not normally distributed and the data do not have unequal variances.
5. Differences between means of groups containing the same entities when the sampling distribution is normally distributed, and the data have equal variances and are at least interval.

The correct answer is d) Differences between means of groups containing the same entities when the sampling distribution is normally distributed, and the data have equal variances and are at least interval.

**Chapter 8**

1. The 99% confidence interval usually is:
2. Narrower than the 95% confidence interval.
3. Wider than the 95% confidence interval.
4. The same as the 95% confidence interval.
5. A less precise estimate of the effect in the population than the 95% confidence interval.

The correct answer is a) Narrower than the 95% confidence interval.

1. A 95% confidence interval is:
2. The range of values of the statistic that we can be 95% confident contains a significant effect in the population.
3. The range of values of the statistic which we can by 95% certain does not contain the true population effect.
4. The range of values of the statistic which probably contains the true value of the statistic in the population.
5. The range of values of the statistic which we can be 5% confident contains a significant effect in the population.

The correct answer is c) The range of values of the statistic which probably contains the true value of the statistic in the population.

1. Which of the following statements is true?
2. Confidence intervals tell us about the range of possible values of a statistic within the sample.
3. Confidence intervals are known as point estimates.
4. Confidence intervals are not biased by non-normally distributed data.
5. If the confidence interval for the difference between two means does include zero then the difference between the means is statistically significant.

The correct answer is d) If the confidence interval for the difference between two means does include zero then the difference between the means is statistically significant.

1. A 95% confidence interval for the difference between two population means is found to be (−0.08, 0.15). Which of the following statements is true?
2. The two populations cannot have the same means.
3. We can be 95% confident that the true difference between the population means falls between −0.08 and 0.15.
4. The probability is 0.05 that the true difference between the population means is between −0.08 and 0.15
5. The probability is 0.95 that a significant difference between the population means lies between −0.08 and 0.15.

The correct answer is b) We can be 95% confident that the true difference between the population means falls between −0.08 and 0.15.

1. Of what is the standard error a measure?
2. The ‘flatness’ of the distribution of sample scores.
3. The variability in scores in the sample.
4. variability of scores in the population.
5. The
6. The variability of sample estimates of a parameter.

The correct answer is d) The variability of sample estimates of a parameter.

1. Which of the following best describes the relationship between sample size and significance testing? (*Hint*: Remember that test statistics are basically a signal-to-noise ratio, so given that large samples have less ‘noise’ they make it easier to find the ‘signal’.)
2. Large effects tend to be significant only in large samples.
3. Large effects tend to be significant only in small samples.
4. In small samples only small effects will be deemed ‘significant’.
5. In large samples even small effects can be deemed ‘significant’.

The correct answer is d) In large samples even small effects can be deemed ‘significant’.

1. Which of the following is not an assumption of the general linear model?
2. Normally distributed residuals
3. Linearity
4. Additivity
5. Dependence

The correct answer is d) Dependence

Feedback: Independence is an assumption of parametric tests, not dependence.

1. Participants rated their mood score out of 20 before and after listening to *Reign in Blood* by the thrash metal band Slayer. What is the null hypothesis of this study?

|  |  |
| --- | --- |
| Before Listening to Slayer | After Listening to Slayer |
| 5 | 14 |
| 8 | 5 |
| 9 | 17 |
| 4 | 18 |
| 3 | 8 |
| 15 | 19 |
| 12 | 14 |
| 6 | 16 |

1. Listening to Slayer is no better than listening to no music at improving mood score.
2. Listening to Slayer decreases mood score.
3. Listening to Slayer does not affect mood.
4. Listening to Slayer increases mood score.

The correct answer is c) Listening to Slayer does not affect mood.

1. A small standard error of differences tells us what? (*Hint*: The standard error of differences is a measure of the unsystematic variation within the data.)
2. That the differences between scores are normally distributed.
3. That most pairs of samples from a population will have very similar means.
4. That sample means can deviate quite a lot from the population mean and so differences between pairs of samples can be quite large by chance alone.
5. That the differences between scores are not normally distributed.

The correct answer is b) That most pairs of samples from a population will have very similar means.

Feedback: The difference between sample means should normally be very small.

1. A researcher wanted to see the effects of different learning strategies. A control group simply read the book *Discovering Statistics* (Book), a second group read the book and completed the ‘end of chapter exercises’ (Book & Exercises), and a third group read the book, did the end of chapter exercises and also completed the web materials (All Activities). The researcher predicted that the ‘all activities’ and ‘book and exercises’ groups would perform better than the book group on a subsequent test, but that the ‘book and exercises’ group would perform worse than the ‘all activities’ group. Which coding scheme would test these hypotheses in a set of planned comparisons? (*Hint*: The sum of weights for a comparison should be zero. If you add up the weights for a given contrast the result should be zero.)
2. 
3. 
4. 



The correct answer is a)

 

**Chapter 10**

1. A Type II error occurs when:

(*Hint*: This would occur when we obtain a small test statistic (perhaps because there is a lot of natural variation between our samples.)

1. The data we have typed into SPSS is different from the data collected.
2. We conclude that there is not an effect in the population when in fact there is.
3. We conclude that the test statistic is significant when in fact it is not.
4. We conclude that there is an effect in the population when in fact there is not.

The correct answer is b) We conclude that there is not an effect in the population when in fact there is.

Feedback: A Type II error would occur when we obtain a small test statistic (perhaps because there is a lot of natural variation between our samples).

1. If we calculated an effect size and found it was *r* = .21 which expression would best describe the size of effect? (*Hint*: The value of *r* can lie between 0 (no effect) and 1 (a perfect effect).
2. medium to large
3. large
4. small
5. small to medium

The correct answer is d) Small to medium.

Feedback: While this is correct, it is worth remembering that these ‘canned’ effect sizes are no substitute for evaluating an effect size within the context of the research domain in which it is being used.

1. What is the null hypothesis for the following question: Is there a relationship between heart rate and the number of cups of coffee drunk within the last 4 hours?
2. There will be a significant relationship between the number of cups of coffee drunk within the last 4 hours and heart rate.
3. There will be no relationship between heart rate and the number of cups of coffee drunk within the last 4 hours.
4. People who drink more cups of coffee will have significantly lower heart rates.
5. People who drink more coffee will have significantly higher heart rates.

The correct answer is b) There will be no relationship between heart rate and the number of cups of coffee drunk within the last 4 hours.

Feedback: The null hypothesis is the opposite of the alternative hypothesis and so usually states that an effect is absent.

1. What is the alternative hypothesis for the following question: Does eating salmon make your skin glow?
2. People who eat salmon will have a more glowing complexion compared to those who don’t.
3. People who eat salmon will have a similar complexion to those who do not.
4. Eating salmon does not predict the glow of skin.
5. There will be no difference in the appearance of the skin of people who eat salmon compared to those who don’t.

The correct answer is a) People who eat salmon will have a more glowing complexion compared to those who don’t.

Feedback: The hypothesis or prediction from your theory would normally be that an effect will be present. This is known as the alternative hypothesis.

1. What are variables?
2. Variables estimate the centre of the distribution.
3. Variables are estimated from the data and are (usually) constants believed to represent some fundamental truth about the relations in the model.
4. Variables are measured constructs that vary across entities in the sample.
5. Variables estimate the relationship between two parameters.

The correct answer is c) Variables are measured constructs that vary across entities in the sample.

1. What are parameters?
2. Parameters are measured constructs that vary across entities in the sample.
3. A parameter tells us about how well the mean represents the sample data.
4. Parameters are estimated from the data and are (usually) constructs believed to represent some fundamental truth about the relations between variables in the model.
5. All of the options describe parameters.

The correct answer is c) Parameters are estimated from the data and are (usually) constructs believed to represent some fundamental truth about the relations between variables in the model.

1. Assume a researcher found that the correlation between a test she had developed and exam performance was .5 in a study of 25 students. She had previously been informed that correlations under .30 are considered unacceptable. The 95% confidence interval was [0.131, 0.747]. Can you be confident that the true correlation is at least 0.30?
2. No you cannot, because the lower boundary of the confidence interval is .131, which is less than .30, and so the true correlation could be less than .30.
3. Yes you can, because the correlation coefficient is .5 (which is above .30) and falls within the boundaries of the confidence interval.
4. No you cannot, because the sample size was too small.
5. Yes you can, because the upper boundary of the confidence interval is above .30 we can be 95% confident that the true correlation will be above .30

The correct answer is a) No you cannot, because the lower boundary of the confidence interval is .131, which is less than .30, and so the true correlation could be less than .30.

1. Why is the standard error important?
2. It is unaffected by outliers.
3. It gives you a measure of how well your sample parameter represents the population value.
4. It tells us the precise value of the variance within the population.
5. It is unaffected by the distribution of scores.

The correct answer is b) It gives you a measure of how well your sample parameter represents the population value.

1. Other things being equal, compared to the paired-samples (or dependent) *t*-test, the independent *t*-test:
2. Has less power to find an effect.
3. Has more power to find an effect.
4. Has the same amount of power, the data are just collected differently.
5. Is less robust.

The correct answer is a) Has less power to find an affect.

Feedback: When the same participants are used across conditions the unsystematic variance (often called the error variance) is reduced dramatically, making it easier to detect any systematic variance.

**Chapter 11**

[no questions available]

**Chapter 12**

1. A \_\_\_\_\_ is a numerical characteristic of a sample and a \_\_\_\_\_ is a numerical characteristic of a population.
2. distribution, variable
3. variable, distribution
4. parameter, statistic
5. statistic, parameter

The correct answer is d) statistic, parameter

1. A kurtosis value of –2.89 indicates:

*(Hint*: Positive values of kurtosis indicate too many scores in the tails of the distribution and that the ditribution is too peaked, whereas negative values indicate too few scores in the tails and that the distribution is quite flat).

1. There is a mistake in your calculation.
2. A pointy and heavy-tailed distribution
3. A flat and heavy-tailed distribution
4. A flat and light-tailed distribution

The correct answer is d) A flat and light-tailed distribution

Feedback: The further the value is from zero, the more likely it is that the data are not normally distributed.

1. What does the graph below indicate about the normality of our data?



1. The histogram reveals that the data deviate substantially from normal.
2. The histogram reveals that the data are more or less normal.
3. The histogram reveals that the data have multivariate normality.
4. We cannot infer anything about the normality of our data from this graph.

The correct answer is b) The histogram reveals that the data are more or less normal.

Feedback: The graph is fairly symmetrical, with the characteristic bell shaped curve.

1. Is it possible to calculate the skewness of a set of numerical scores?
2. Yes.
3. Only if you have used an independent-measures design.
4. No.
5. Only if you have a large sample size.

The correct answer is a) yes.

1. A researcher investigating ‘Pygmalion in the classroom’ measured teachers’ perceptions of male and female students’ mathematical abilities. She collected teacher ratings of 97 male and female students. Based on the output, what can you say about the data?



1. Mean perceptions of male and female students were significantly different.
2. Mean and median ratings were similar for males and females
3. The variances of ratings were significantly different for males and females.
4. Homogeneity of variance can be assumed.

The correct answer is c) The variances of ratings were significantly different for males and females.

Feedback: The significant result (*p* < .05) means that homogeneity of variance cannot be assumed.

1. Levene’s test can be used to measure:
2. Whether scores are normally distributed.
3. Whether scores are independent.
4. Whether group means are equal.
5. Whether group variances differ.

The correct answer is d) Whether group variances differ.

1. To get a sample of a certain size, scores are taken one-by-one from the observed data and each time replaced. The parameter of interest (e.g., the mean or *b* in regression) is computed within the sample. This process is repeated numerous times. The resulting parameter estimates are used to compute a confidence interval. The process I am describing is:
2. Bootstrapping
3. Significance testing
4. Sampling
5. The standard error

The correct answer is a) Bootstrapping.

1. A psychologist was interested in predicting how depressed people are from the amount of news they watch. Based on the output, do you think the psychologist will end up with a model that can be generalized beyond the sample?



1. No, because the errors lack linearity.
2. No, because the errors show heteroscedasticity.
3. Yes, because errors are normally distributed.
4. Yes, because errors are independent.

The correct answer is b) No, because the errors show heteroscedasticity.

Feedback: The funnel shape of the residuals indicates heteroscedasticity.

1. The following graph shows:



1. Heteroscedasticity
2. Heteroscedasticity and non-linearity
3. Non-linearity
4. Regression assumptions that have been met

The correct answer is d) Regression assumptions that have been met.

Feedback: This is correct because the residuals are spread out evenly and do not curve or funnel out.

1. Recent research has shown that lecturers are among the most stressed workers. A researcher wanted to know exactly what it was about being a lecturer that created this stress and subsequent burnout. She recruited 75 lecturers and administered several questionnaires that measured: **Burnout** (high score = burnt out), **Perceived Control** (high score = low perceived control), **Coping Ability** (high score = low ability to cope with stress), **Stress from Teaching** (high score = teaching creates a lot of stress for the person), **Stress from Research** (high score = research creates a lot of stress for the person), and **Stress from Providing Pastoral Care** (high score = providing pastoral care creates a lot of stress for the person). The outcome of interest was burnout, and Cooper’s (1988) model of stress indicates that perceived control and coping style are important predictors of this variable. The remaining predictors were measured to see the unique contribution of different aspects of a lecturer’s work to their burnout. What analysis has been carried out?



1. Hierarchical multiple regression
2. Multilevel model
3. Reliability analysis
4. Factor analysis

The correct answer is a) Hierarchical multiple regression

1. Recent research has shown that lecturers are among the most stressed workers. A researcher wanted to know exactly what it was about being a lecturer that created this stress and subsequent burnout. She recruited 75 lecturers and administered several questionnaires that measured: **Burnout** (high score = burnt out), **Perceived Control** (high score = low perceived control), **Coping Ability** (high score = low ability to cope with stress), **Stress from Teaching** (high score = teaching creates a lot of stress for the person), **Stress from Research** (high score = research creates a lot of stress for the person), and **Stress from Providing Pastoral Care** (high score = providing pastoral care creates a lot of stress for the person). The outcome of interest was burnout, and Cooper’s (1988) model of stress indicates that perceived control and coping style are important predictors of this variable. The remaining predictors were measured to see the unique contribution of different aspects of a lecturer’s work to their burnout. Which of the predictor variables does **not** predict burnout?



1. Stress from providing pastoral care
2. Stress from teaching
3. Stress from research
4. Perceived control

The correct answer is c) Stress from research.

1. Recent research has shown that lecturers are among the most stressed workers. A researcher wanted to know exactly what it was about being a lecturer that created this stress and subsequent burnout. She recruited 75 lecturers and administered several questionnaires that measured: **Burnout** (high score = burnt out), **Perceived Control** (high score = low perceived control), **Coping Ability** (high score = low ability to cope with stress), **Stress from Teaching** (high score = teaching creates a lot of stress for the person), **Stress from Research** (high score = research creates a lot of stress for the person), and **Stress from Providing Pastoral Care** (high score = providing pastoral care creates a lot of stress for the person). The outcome of interest was burnout, and Cooper’s (1988) model of stress indicates that perceived control and coping style are important predictors of this variable. The remaining predictors were measured to see the unique contribution of different aspects of a lecturer’s work to their burnout. How much variance in burnout does the final model explain for the sample?



1. 89.6%
2. 8.3%
3. 79.2%
4. 80.3%

The correct answer is d) 80.3%

1. The student welfare office was interested in trying to enhance students’ exam performance by investigating the effects of various interventions. They took five groups of students before their statistics exams and gave them one of five interventions: (1) a control group just sat in a room contemplating the task ahead; (2) the second group had a yoga class to relax them; (3) the third group were told they would get monetary rewards contingent upon the grade they received in the exam; (4) the fourth group were given beta-blockers to calm their nerves; and (5) the fifth group were encouraged to sit around winding each other up about how much revision they had/hadn’t done (a bit like what usually happens). The final percentage obtained in the exam was the dependent variable. Using the critical values for *F*, how would you report the result in the table below?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *SS* | *Df* | *MS* | *F* |
| Model | 1213.6 | 4 | 303.4 | 12.43\*\* |
| Residual | 707.9 | 29 | 21.4 |  |
| Total | 1921.5 | 33 |  |  |

1. Type of intervention had a significant effect on levels of exam performance, *F*(4, 33) = 12.43, *p* < .01.
2. Type of intervention did not have a significant effect on levels of exam performance, *F*(4, 33) = 12.43, *p* > .01.
3. Type of intervention had a significant effect on levels of exam performance, *F*(4, 29) = 12.43, *p* < .01.
4. Type of intervention did not have a significant effect on levels of exam performance, *F*(4, 29) = 12.43, *p* > .05.

The correct answer is c) Type of intervention had a significant effect on levels of exam performance, *F*(4, 29) = 12.43, *p* < .01.

**Chapter 13**

1. Imagine a researcher wanted to investigate whether there was a significant correlation between IQ and annual income, but she had reason to believe that work ethic would influence both of these variables. What should she do?
2. Conduct a partial correlation to look at the relationship between IQ and annual income while partialling out the effect of work ethic.
3. Conduct a partial correlation to look at the relationship between work ethic and annual income partialling out the effect of IQ.
4. Conduct a semi-partial correlation to look at the relationship between IQ and annual income while partialling out the effect of work ethic.
5. Conduct a semi-partial correlation to look at the relationship between IQ and work ethic while partialling out the effect of annual income.

The correct answer is a) Conduct a partial correlation to look at the relationship between IQ and annual income while partialling out the effect of work ethic.

Feedback: Partial correlation partials out the effect that the third variable has on *both* variables in the correlation.

1. The table below contains scores from six people on two different scales that measure attitudes towards reality TV shows.

|  |  |
| --- | --- |
| **Attitudes towards Watching Reality TV Scale** | **The General Reality TV Scale** |
| 3 | 2 |
| 7 | 5 |
| 4 | 3 |
| 1 | 1 |
| 8 | 7 |
| 6 | 7 |

Using the scores above, the two scales are likely to:

(*Hint*: If two variables are related, then changes in one variable should be met with similar changes in the other variable.)

1. Have identical means.
2. Correlate positively.
3. Be uncorrelated.
4. Correlate negatively.

The correct answer is b) Correlate positively

Feedback: High scores on one scale tend to produce high scores on the other, and low scores on one also correspond with low scores on the other

1. The table below contains scores from six people on two different scales that measure attitudes towards reality TV shows.

|  |  |
| --- | --- |
| **Attitudes towards Watching Reality TV Scale** | **General Attitudes towards Reality TV Scale**  |
| 3 | 2 |
| 7 | 5 |
| 4 | 3 |
| 1 | 1 |
| 8 | 7 |
| 6 | 7 |

Based on intuition rather than computation, which of the following is the value of the coefficient of determination between the two scales?

1. .85
2. –.85
3. 85
4. .085

The correct answer is a) .85

Feedback: The coefficient of determination is *r* squared, and there is a very strong positive relationship between the scales.

1. How much greater is the shared variance between two variables if the Pearson correlation coefficient between them is –.4 than if it is .2?
2. Four times as great
3. Two times as great
4. Half as much
5. A quarter as much

The correct answer is a) Four times as great

1. Which correlation coefficient would you use to look at the correlation between gender and time spent on the phone talking to your mother?
2. The biserial correlation coefficient, *r*b
3. Kendall’s correlation coefficient, *τ*
4. The point-biserial correlation coefficient, *r*pb
5. Pearson’s correlation coefficient, *r*

The correct answer is c) The point-biserial correlation coefficient, *r*pb

Feedback: The point-biserial correlation coefficient, *r*pb, is used when one variable is a discrete dichotomy (e.g., gender)

1. A Pearson’s correlation coefficient of –.5 would be represented by a scatterplot in which:
2. The data cloud looks like a circle and the regression line is flat.
3. The regression line slopes upwards.
4. Half of the data points sit perfectly on the line.
5. There is a moderately good fit between the regression line and the individual data points on the scatterplot.

The correct answer is d) There is a moderately good fit between the regression line and the individual data points on the scatterplot.

1. If you have a curvilinear relationship, then:

(*Hint*: The two most important sources of bias in this context are probably linearity and normality.)

1. Transforming the data won’t help.
2. You can use Pearson’s correlation; you just need to remember that a curve indicates that the variables are not linearly related.
3. It is not appropriate to use Pearson’s correlation because it assumes a linear relationship between variables.
4. Pearson’s correlation can be used in the same way as it is for linear relationships.

The correct answer is c) It is not appropriate to use Pearson’s correlation because it assumes a linear relationship between variables.

Feedback: Remember that we’re fitting a linear model to the data, so if the relationship between variables is not linear then this model is invalid, and you might need to transform the data.

1. If a correlation coefficient has an associated probability value of .02 then:
2. There is only a 2% chance that we would get a correlation coefficient this big (or bigger) if the null hypothesis were true.
3. The results are important.
4. We should accept the null hypothesis.
5. The hypothesis has been proven.

The correct answer is a) There is only a 2% chance that we would get a correlation coefficient this big (or bigger) if the null hypothesis were true.

Feedback: A probability value of .02 indicates that the correlation coefficient is significant.

1. Which of the following could not be a correlation coefficient:
2. 0
3. .27
4. –.27
5. 2.7

The correct answer is d) 2.7

Feedback: A correlation coefficient can only range from –1 to +1.

1. Looking at the table below, which variables were the most strongly correlated?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Work ethic | Annual income | IQ |
| Work ethic | Pearson’s correlation | 1.000 | .72 | .66 |
|  | Sig. (2-tail) | . | .001 | .000 |
|  | *N* | 550 | 550 | 550 |
| Annual income | Pearson’s correlation | .72 | 1.000 | .47 |
|  | Sig. (2-tail) | .000 | . | .03 |
|  | *N* | 550 | 550 | 550 |
| IQ | Pearson’s correlation | .66 | .47 | 1.000 |
|  | Sig. (2-tail) | .000 | .03 | . |
|  | *N* | 550 | 550 | 550 |

(*Hint*: Larger values of Pearson’s correlation (ignoring the sign) indicate stronger correlations between variables.)

1. Work ethic and annual income
2. Work ethic and IQ
3. Annual income and IQ
4. None of the correlations are significant

The correct answer is a) Work ethic and annual income.

Feedback: The value of Pearson’s *r* was .72, which was the largest in the table.

1. The coefficient of determination:
2. Is the square root of the correlation coefficient.
3. Is the square root of the variance.
4. Is a measure of the amount of variability in one variable that is shared by the other.
5. Indicates whether the correlation coefficient is significant.

The correct answer is c) Is a measure of the amount of variability in one variable that is shared by the other.

1. A Pearson’s correlation of –.71 was found between number of hours spent at work and energy levels in a sample of 300 participants. Which of the following conclusions can be drawn from this finding?
2. The estimate of the correlation will be imprecise.
3. Spending more time at work caused participants to have less energy.
4. Amount of time spent at work accounted for 71% of the variance in energy levels.
5. There was a strong negative relationship between the number of hours spent at work and energy levels.

The correct answer is d) There was a strong negative relationship between the number of hours spent at work and energy levels.

1. Which of the following statements about the *F*-ratio is true?
2. The *F*-ratio is the ratio of variance explained by the model to the error in the model.
3. The *F*-ratio is the ratio of variance explained by the model to the total variance in the outcome variable.
4. The *F*-ratio is the ratio of error variance to the total variance.
5. The *F*-ratio is the proportion of variance explained by the regression model.

The correct answer is a) The *F*-ratio is the ratio of variance explained by the model to the error in the model.

1. A researcher asked 933 people which type of programme they prefer to watch on television. Results are below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | News | Documentaries | Soaps | Sport | Total |
| Women | 108 | 123 | 187 |  62 | 480 |
| Men | 130 | 123 |  68 | 132 | 453 |
| Total | 238 | 246 | 255 | 194 | 933 |

A chi-square test produced the SPSS output below. What can we conclude from this output?



1. Men and women watch similar types of programmes.
2. The profile of programmes watched was significantly different between men and women.
3. Significantly more soap operas were watched.
4. Women watched significantly more programmes than men.

The correct answer is b) The profile of programmes watched was significantly different between men and women.

Feedback: The *p*-value is less than .05 for the row labelled *Pearson Chi-Square* and underneath the table SPSS tells us that no expected frequencies are less than 5.

1. Are directional hypotheses possible with chi-square?
2. Yes, but only when your sample is greater than 200.
3. Yes, but only when you have a 2 × 2 design.
4. Yes, but only when there are 12 or more degrees of freedom.
5. Directional hypotheses are never possible with the chi-squared test.

The correct answer is b) Yes, but only when you have a 2 × 2 design.

Feedback: Directional alternative hypotheses using the chi-square test are only valid for 2 × 2 contingency tables – any larger and the chi-square test is testing compound hypotheses.

1. Men and women were asked which type of animal they thought made the best pets. Data are in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Reptiles | Mammals | Birds |
| Men | 24 | 35 | 20 |
| Women | 15 | 47 | 12 |

If the expected frequencies rule for chi-square had been violated by the data, which categories could be combined together in a meaningful way to increase the expected frequencies?

1. It depends on the research hypothesis.
2. Reptiles, mammals and neither
3. Men and women
4. Reptiles and mammals

The correct answer is a) It depends on the research hypothesis.

Feedback: All of the answers are correct in a sense, but it depends on the question that you want to answer. If you want to know whether men and women like different types of pets it doesn’t make sense to combine males and females, but it might be acceptable to combine reptiles and birds so you could see gender effects on mammals vs. other animals

1. With 2 × 2 contingency tables (i.e., two categorical variables both with two categories) no expected values should be below \_\_\_\_.
2. 10
3. 0.8
4. 1
5. 5

The correct answer is d) 5.

1. What does Fisher’s exact probability show? (*Hint*: Fisher came up with a method for computing the exact probability of the chi-square statistic that is accurate when sample sizes are small.)
2. It is the amount of variation that one variable can explain in the other variable.
3. It is the probability of obtaining a chi-square value at least as big as the one observed if the null hypothesis were true.
4. It tests whether the assumption of independence has been met.
5. It tests whether the assumption of expected frequencies has been met.

The correct answer is b) It is the probability of obtaining a chi-square value at least as big as the one observed if the null hypothesis were true.

1. What are phi and Cramér’s *V* used for?
2. They are measures of the strength of association between two categorical variables.
3. They can be used as remedies when the assumption of expected frequencies has been violated.
4. They are used as a correction when the assumption of independence has been violated.
5. They are measures of the strength of association between two continuous variables.

The correct answer is a) They are measures of the strength of association between two categorical variables.

Feedback: They are measures of the size of the effect.

1. Which of the hypotheses below would be suited for testing with a one-variable chi-square test?
2. People who choose the number 7 as their ‘lucky’ number are significantly more superstitious than people who choose the number 13 as their ‘lucky’ number.
3. Choice of ‘lucky’ number is directly related to measures superstition.
4. It was hypothesized that more people would choose the number 7 as their ‘lucky’ number than any other number.
5. All of these.

The correct answer is c) It was hypothesized that more people would choose the number 7 as their ‘lucky’ number than any other number.

1. How are the degrees of freedom calculated for a chi-square test? (*Hint*: *r* is the number of rows and *c* is the number of columns.)
2. (*r* − 1) / (*c* − 1)
3. (*r* + 1) − (*c* + 1)
4. (*r* − 1)(*c* −1)
5. (*r* × 2) − (*c* × 2)

The correct answer is c) (*r* − 1)(*c* −1)

1. Which of the following statements about loglinear analysis is false?
2. Loglinear analysis relies on all of the same assumptions of a linear model.
3. Loglinear analysis is hierarchical: the initial model contains all main effects and interactions.
4. The likelihood ratio statistic is used to assess each model in loglinear analysis.
5. Loglinear analysis can be used when you want to test the relationship between more than two categorical variables.

The correct answer is a) Loglinear analysis relies on all of the same assumptions of a linear model.

Feedback: The above statement is false because categorical data cannot have a normal sampling distribution because they aren’t continuous.

1. A recent story in the media has claimed that women who eat breakfast every day are more likely to have boy babies than girl babies. Imagine you conducted a study to investigate this in women from two different age groups (18–30 and 31–43 years). Looking at the output tables below, which of the following sentences best describes the results?







1. The model is a poor fit of the data.
2. There was a significant two-way interaction between eating breakfast and age group of the mother.
3. Women who ate breakfast were significantly more likely to give birth to baby boys than girls.
4. Whether or not a woman eats breakfast significantly affects the gender of her baby at any age.

The correct answer is d) Whether or not a woman eats breakfast significantly affects the gender of her baby at any age.

Feedback: To interpret this interaction, we could perform a chi-square test on breakfast and gender.

1. The interpretation of the odds ratio, Exp(*B*), can be generalized to the population if:
2. The confidence interval of Exp(*B*) does cross 0.
3. The confidence interval of Exp(*B*) does cross 1.
4. The confidence interval of Exp(*B*) does not cross 0.
5. The confidence interval of Exp(*B*) does not cross 1.

The correct answer is d) The confidence interval of Exp(*B*) does not cross 1.

Feedback: If our confidence interval does not cross zero then we can be confident that the direction of the relationship we have observed is true in the population.

**Chapter 14**

1. Which of the following statements about the *t*-statistic in regression is **not** true?
2. The *t*-statistic is equal to the regression coefficient divided by its standard deviation.
3. The *t*-statistic can be used to see whether a predictor variables makes a statistically significant contribution to the regression model.
4. The *t*-statistic provides some idea of how well a predictor predicts the outcome variable
5. The *t*-statistic tests whether the regression coefficient, *b*, is equal to 0.

The correct answer is a) The *t*-statistic is equal to the regression coefficient divided by its standard deviation.

Feedback: This is correct as the statement is not true.

1. What is multicollinearity? (*Hint*: It is an assumption of the linear model. For this assumption to be met, we want there to be no multicollinearity in our data set.)
2. When predictor variables are independent
3. When predictor variables are correlated with variables not in the regression model
4. When predictor variables correlate very highly with each other
5. When predictor variables have a linear relationship with the outcome variable

The correct answer is c) When predictor variables correlate very highly with each other

1. Which of the following is **not** a reason why multicollinearity a problem in regression?
2. It leads to unstable regression coefficients.
3. It creates heteroscedasticity in the data.
4. It limits the size of *R*.
5. It makes it difficult to assess the importance of individual predictors.

The correct answer is b) It creates heteroscedasticity in the data.

Feedback: This is correct, because it is not true. When the variances are very unequal there is said to be heteroscedasticity. Violating this assumption invalidates our confidence intervals and significance tests. Multicollinearity, however, is when predictor variables correlate too highly.

1. Which of these statements is **not** true?
2. Tolerance values above 0.2 may indicate multicollinearity in the data.
3. The tolerance is 1 divided by the VIF (variance inflation factor).
4. Multicollinearity in the data is shown by a VIF (variance inflation factor) greater than 10.
5. If the average variance inflation factor is greater than 1 then the regression model might be biased.

The correct answer is a) Tolerance values above 0.2 may indicate multicollinearity in the data.

Feedback: Yes, this is correct, because it is not true. Tolerance below 0.1 indicates a serious problem and tolerance below 0.2 indicates a potential problem.

**Chapter 15**

1. What does the independent *t*-test test assume?
2. The sampling distribution is normally distributed.
3. The data are normally distributed
4. There are no differences between the mean scores.
5. The means of two sets of scores are correlated.

The correct answer is a) The sampling distribution is normally distributed.

1. When conducting an independent *t*-test, what is the dependent variable?
2. The experimental conditions
3. The scores
4. One of the independent variables becomes the dependent variable in the analysis.
5. The term ‘dependent variable’ does not apply to the *t*-test.

The correct answer is b) The scores

1. Which of the following sentences is an accurate description of the standard error?
2. The standard deviation squared.
3. It is the same as the standard deviation.
4. It is the standard deviation of the sampling distribution of a statistic.
5. It is the observed difference between sample means minus the expected difference between population means (if the null hypothesis is true).

The correct answer is c) It is the standard deviation of the sampling distribution of a statistic.

1. If you use a paired samples *t*-test:
2. There ought to be less unsystematic variance compared to the independent *t*-test.
3. The same participants take part in both experimental conditions.
4. Other things being equal, you do not need as many participants as you would for an independent samples design.
5. All of these are correct.

The correct answer is d) All of these are correct

Feedback: When the same participants are used across conditions the unsystematic variance (often called the error variance) is reduced dramatically, making it easier to detect any systematic variance.

1. Which of the following statements about the *t* distribution is correct?
2. It is skewed.
3. In small samples it is narrower than the normal distribution
4. It follows an exponential curve.
5. As the degrees of freedom increase, the distribution becomes closer to normal.

The correct answer is d) As the degrees of freedom increase, the distribution becomes closer to normal.

1. The degrees of freedom for the paired samples *t*-test are:
2. *N* − 1
3. √*N* − 1
4. *N*
5. *N* − 2

The correct answer is a) *N* − 1

Feedback: When the same participants have been used, the degrees of freedom are the sample size minus 1.

1. Participants rated their mood score out of 20 before and after listening to *Reign in Blood* by the thrash metal band Slayer.

|  |  |
| --- | --- |
| Before Listening to Slayer | After Listening to Slayer |
| 5 | 14 |
| 8 | 5 |
| 9 | 17 |
| 4 | 18 |
| 3 | 8 |
| 15 | 19 |
| 12 | 14 |
| 6 | 16 |

What are the degrees of freedom for this study?

1. 8
2. 7
3. 16
4. 15

The correct answer is b) 7

Feedback: When the same participants have been used, the degrees of freedom are the sample size minus 1.

1. A recent study investigated whether vodka is less likely to give you a hangover than wine. Twenty participants on a night out were asked to drink only vodka for the whole evening then rate how they felt the next day out of 10 (0 = I feel fantastic, 10 = I can't move my head in case it explodes). The following month, they were asked to do the same again, only this time they were asked to drink only white wine. The *t*-score was 2.56. Which of the sentences below is correct? (*Hint*: You will need to look up the *p*-value in the table of critical values of the *t-*distribution in the Appendix.)
2. It is not significant with a two-tailed test.
3. It is not significant with a one-tailed test.
4. It is significant at the 1% level with a two-tailed test.
5. It is significant at the 5% level but not the 1% level with a two-tailed test.

The correct answer is d) It is significant at the 5% level but not the 1% level with a two-tailed test.

Feedback: Our value of *t* was *t*(19) = 2.56, which is larger than the .05 critical value of 2.09 and smaller than the .01 critical value of 2.86.

1. The *t*-statistic:
2. When significant, indicates an important finding.
3. Is accurate only when using large samples.
4. Is the ratio of the systematic variation to the unsystematic variation.
5. Is the standard deviation of the sampling distribution of a statistic.

The correct answer is c) Is the ratio of the systematic variation to the unsystematic variation.

1. Two samples of data are collected and the sample means calculated. If the samples come from the same population, then:
2. Their means should be roughly equal.
3. The difference between the samples we have collected is likely to be larger than we would expect based on the standard error.
4. Their means should differ significantly.
5. The experiment is unreliable.

The correct answer is a) Their means should be roughly equal.

Feedback: Although it is possible for their means to differ by chance alone, we would expect large differences between sample means to occur very infrequently.

1. What does the variance sum law state?
2. That the variance of a difference between two independent variables is smaller than the sum of their variances.
3. That the variance of a difference between two independent variables is equal to the sum of their variances.
4. That the variance of a difference between two independent variables is larger than the sum of their variances.
5. That the sum of the variances of two independent variables is larger than the sum of their individual variances.

The correct answer is b) That the variance of a difference between two independent variables is equal to the sum of their variances.

Feedback: This statement means that the variance of the sampling distribution is equal to the sum of the variances of the two populations from which the samples were taken.

**Chapter 16**

1. The table below contains the length of time (minutes) for which different groups of students were able to stay awake to revise statistics after consuming 500 ml of one of three different types of stimulants. What is the variation in scores from groups A to B to C known as?

|  |  |  |
| --- | --- | --- |
| **A** | **B** | **C** |
| 20 | 15 | 40 |
| 15 | 12 | 33 |
| 120 | 7 | 50 |
| 57 | 18 | 135 |

1. The between-groups variance
2. The within-groups variance
3. The grand variance
4. Homogeneity of variance

The correct answer is a) The between-groups variance

1. When the between-groups variance is a lot larger than the within-groups variance, the *F*-value is \_\_\_\_ and the likelihood of such a result occurring because of sampling error is \_\_\_\_\_.
2. small; low
3. large; high
4. large; low
5. small; high

The correct answer is c) large; low

Feedback: Yes, this is correct. If the differences between group means are large enough, then the resulting model will be a better fit of the data than the grand mean.

1. Subsequent to obtaining a significant result from an exploratory one-way independent ANOVA, a researcher decided to conduct three *t*-tests to investigate where the differences between groups lie. Which of the following statements is correct?
2. The researcher should accept as statistically significant tests with a probability value of less than 0.016 to avoid making a Type I error.
3. This is the correct method to use. The researcher did not make any predictions about which groups will differ before running the experiment, therefore contrasts and *post hoc* tests cannot be used.
4. The researcher should have conducted orthogonal contrasts instead of *t*-tests to avoid making a Type I error.
5. None of these options are correct.

The correct answer is a) The researcher should accept as statistically significant tests with a probability value of less than 0.016 to avoid making a Type I error.

Feedback: Yes, this is correct. Conducting multiple *t*-tests increases the familywise error rate, so if you are going to do this, it is important to divide the accepted probability level (.05) by the number of *t*-tests you conduct (in this case 3). Alternatively, you could run *post hoc* tests, which control the familywise error rate.

1. When conducting a one-way independent ANOVA with three levels on the independent variable, an *F*-ratio that is large enough to be statistically significant tells us:
2. That all of the differences between means are statistically significant.
3. That there is a significant three-way interaction.
4. That the model fitted to the data accounts for less variation than extraneous factors, but it doesn’t tell us where the differences between groups lie.
5. That one or more of the differences between means is statistically significant but not where the differences between groups lie.

The correct answer is d) That one or more of the differences between means is statistically significant but not where the differences between groups lie.

Feedback: It is necessary after conducting an ANOVA to carry out further analysis to find out which groups differ.

1. What is the overall effect of an independent variable on a dependent variable known as?
2. The indirect effect
3. The main effect
4. The direct effect
5. The interaction effect

The correct answer is b) The main effect

1. After an ANOVA you need more analysis to find out which groups differ. When you did not generate specific hypotheses before the experiment use:

(*Hint*: We need a way to contrast the different groups without inflating the Type I error rate.)

1. *Post hoc* tests
2. Bootstrapping
3. Planned contrasts
4. *t*-tests

The correct answer is a) *Post hoc* tests

Feedback: *Post hoc* tests compare every group (as if conducting several *t*-tests) but use a stricter acceptance criterion such that the familywise error rate does not rise above .05

1. Imagine we conduct a one-way independent ANOVA with four levels on our independent variable and obtain a significant result. Given that we had equal sample sizes, we did not make any predictions about which groups would differ before the experiment and we want guaranteed control over the Type I error rate, which would be the best test to investigate which groups differ?

(*Hint*: *Post hoc* tests do not require specific *a priori* predictions about which groups will differ.)

1. Helmert
2. Hochberg’s GT2
3. Bonferroni
4. Orthogonal contrasts

The correct answer is d) Orthogonal contrasts

Feedback: If you want guaranteed control over the Type I error rate then use Bonferroni.

1. A music teacher had noticed that some students went to pieces during exams. He wanted to test whether this performance anxiety was different for people playing different instruments. He took groups of guitarists, drummers and pianists (variable = ‘Instru’) and measured their anxiety (variable = ‘Anxiety’) during the exam. He also noted the type of exam they were performing (in the UK, musical instrument exams are known as ‘grades’ and range from 1 to 8). He wanted to see whether the type of instrument played affected performance anxiety when accounting for the grade of the exam. Which of the following statements best reflects what the effect of GRADE in the table tells us?



(*Hint*: ANCOVA looks at the relationship between an independent and dependent variable, taking into account the effect of a covariate.)

1. The grade of exam had a significant relationship with the level of anxiety experienced.
2. The grade of exam did not have a significant relationship with the level of anxiety experienced.
3. The type of instrument played did not have a significant relationship with the grade of exam that was being taken.
4. The type of instrument played had a significant relationship with the grade of exam that was being taken.

The correct answer is a) The grade of exam had a significant relationship with the level of anxiety experienced.

1. An experiment was done to look at whether there is an effect of the number of hours spent practising a musical instrument and gender on the level of musical ability. A sample of 30 (15 men and 15 women) participants who had never learnt to play a musical instrument before were recruited. Participants were randomly allocated to one of three groups that varied in the number of hours they would spend practising every day for 1 year (0 hours, 1 hours, 2 hours). Men and women were divided equally across groups. All participants had a one-hour lesson each week over the course of the year, after which their level of musical skill was measured on a 10-point scale ranging from 0 (you can’t play for toffee) to 10 (‘Are you Mozart reincarnated?’). An ANOVA was conducted on the data from the experiment. Looking at the output below, which of the following sentences is correct?





1. There was a significant main effect of gender. Looking at the group means, we can see that women achieved a significantly higher level of musical skill (*M* = 5.53) than males (*M* = 4.87).
2. There was a significant main effect of practice. This means that men and women significantly differed in the amount of practice that they did.
3. There was no significant main effect of gender. This means that overall, when we ignore the number of hours spent practising, the gender of the participant did not have a significant effect on the level of musical skill.
4. There was no significant practice × gender interaction effect, indicating that if we ignore the gender of the participant, the number of hours spent practising did not significantly affect the level of musical skill.

The correct answer is c) There was no significant main effect of gender. This means that overall, when we ignore the number of hours spent practising, the gender of the participant did not have a significant effect on the level of musical skill.

Feedback: If you look at the means in the table labelled *Gender* you can see that although women scored slightly higher (*M* = 5.53) than men (*M* = 4.87), the means were very similar.

1. The table below shows hypothetical data from an experiment with three conditions.

|  |  |  |
| --- | --- | --- |
| Condition A | Condition B | Condition C |
| 12 | 10 | 9 |
| 15 | 15 | 13 |
| 20 | 25 | 31 |
| 32 | 30 | 27 |
| 54 | 46 | 50 |

For these data, sphericity will hold when:

(*Hint*: Sphericity refers to the equality of variances of the *differences* between treatment levels.)

1. $variance\_{A-B}\ne variance\_{B-C}\ne variance\_{A-C}$
2. $variance\_{A+B}≈variance\_{B+C}≈variance\_{A+C}$
3. $variance\_{A}≈variance\_{B}≈variance\_{C}$
4. $variance\_{A-B}≈variance\_{B-C}≈variance\_{A-C}$

The correct answer is d) $variance\_{A-B}≈variance\_{B-C}≈variance\_{A-C}$

Feedback: Sphericity is met when the variances of the differences between treatment levels are roughly equal.

1. Which of the following regarding the assumption of sphericity is **false**?

(*Hint*: Even very small departures from sphericity produce large biases in the *F*-test.)

1. Sphericity does not affect *post hoc* tests.
2. When sphericity is violated, the Bonferroni *post hoc* test can be used.
3. The effect of violating sphericity is a loss of power (i.e., an increased probability of a Type II error).
4. When sphericity is violated, we need to adjust the degrees of freedom for any *F*-ratios affected by the violation.

The correct answer is a) Sphericity does not affect *post hoc* tests.

Feedback: This statement is false, therefore it is the correct answer. Sphericity does cause some complications for *post hoc* tests.

1. When entering data for a repeated-measures design in SPSS:
2. Each row of the data editor should represent a level of a variable, while each column represents data from one entity.
3. Each row of the data editor should represent data from one entity, while each column represents a coding variable.
4. Each row of the data editor should represent data from one entity, while each column represents a level of a variable.
5. It doesn’t matter how you enter repeated-measures data into SPSS.

The correct answer is c) Each row of the data editor should represent data from one entity, while each column represents a level of a variable.

Feedback: Separate columns should represent each level of a repeated-measures variable.

1. When an experimental manipulation is carried out on the same entities, the within-participant variance will be made up of:
2. The effect of the manipulation and individual differences in performance.
3. Unsystematic variance only.
4. Individual differences in performance only.
5. The effect of the manipulation only.

The correct answer is a) The effect of the manipulation and individual differences in performance.

1. The results of a one-way repeated-measures ANOVA with four levels on the independent variable revealed a significance value for Mauchly’s test of *p* = 0.048. What does this mean?
2. The assumption of sphericity has been met.
3. The assumption of sphericity has been violated.
4. That Tukey’s test should be used.
5. This value can be ignored because sphericity is not an issue in a one-way repeated-measures ANOVA design.

The correct answer is b) The assumption of sphericity has been violated.

Feedback: The significance value (.048) is less than the critical value of .05, which means that the assumption of sphericity has been violated.

1. Imagine I ran a one-way repeated-measures ANOVA with five levels on the independent variable. The results revealed a Greenhouse–Geisser estimate, $\hat{ε} $= .977, and the Huynh–Feldt estimate, $\tilde{ε}$ = .999. What do these values tells us about the assumption of sphericity? (*Hint*: When the data are perfectly spherical, these estimates will be 1.)
2. We do not need to check the assumption of sphericity for a one-way repeated-measures ANOVA.
3. We cannot tell anything about the assumption of sphericity from these values.
4. The assumption of sphericity is likely to have been violated.
5. The assumption of sphericity is likely to have been met.

The correct answer is d) The assumption of sphericity is likely to have been met.

Feedback: For the Greenhouse–Geisser estimate, the lowest possible value is ${1}/{\left(κ-1\right)},$ which with five conditions will be ${1}/{\left(5-1\right)}=.25 $. The Greenhouse–Geisser estimate is closer to the upper limit of 1 than to the lower limit of .25, so we do not have much deviation from sphericity at all.

1. Imagine we were interested in the effect of supporters singing on the number of goals scored by soccer teams. We took 10 groups of supporters of 10 different soccer teams and asked them to attend three home games, one at which they were instructed to sing in support of their team (e.g., ‘Come on, you Reds!’), one at which they were instructed to sing negative songs towards the opposition (e.g., ‘You’re getting sacked in the morning!’) and one at which they were instructed to sit quietly. The order of chanting was counterbalanced across groups. Which of the following could be used to analyse these data?

(*Hint*: All participants took part in all experimental conditions.)

1. Loglinear analysis
2. One-way repeated-measures ANOVA
3. Three-way repeated-measures ANOVA
4. One-way between-groups ANOVA

The correct answer is b) One-way repeated-measures ANOVA

Feedback: There was one independent variable (singing) with three levels (positive, negative, and none), and all participants took part in all three conditions.

1. Imagine we were interested in the effect of supporters singing on the number of goals scored by soccer teams. We took 10 groups of supporters of 10 different soccer teams and asked them to attend three home games, one at which they were instructed to sing in support of their team (e.g., ‘Come on, you Reds!’), one at which they were instructed to sing negative songs towards the opposition (e.g., ‘You’re getting sacked in the morning!’) and one at which they were instructed to sit quietly. The order of chanting was counterbalanced across groups. Which of the following sentences regarding the output from Mauchly’s test of sphericity below is correct?



1. Mauchly’s test indicated that the assumption of sphericity had been met, *χ*2 (2) = 2.50, *p* = .29.
2. Mauchly’s test indicated that the assumption of sphericity had been met, *χ*2 (2) = 2.50, *p* < .05.
3. Mauchly’s test indicated that the assumption of sphericity had been violated, *χ*2 (2) = .73, *p* = .29.
4. Mauchly’s test indicated that the assumption of sphericity had not been met, *χ*2 (2) = 2.50, *p* > .05.

The correct answer is a) Mauchly’s test indicated that the assumption of sphericity had been met, *χ*2 (2) = 2.50, *p* = .29.

Feedback: The significance value of Mauchly’s *W* is .287, which is larger than .05, indicating that the assumption of sphericity had been met.

1. Which of the following statements is false? When you have data that violate the assumption of sphericity:
2. The Greenhouse–Geisser or Huynh–Feldt correction should be applied.
3. The degrees of freedom are adjusted for any *F*-ratios affected by the violation using estimates of sphericity.
4. The means are adjusted for any groups that are affected by the violation using estimates of sphericity.
5. You can use multivariate test statistics (MANOVA) instead.

The correct answer is c) The means are adjusted for any groups that are affected by the violation using estimates of sphericity.

Feedback: This statement is false, therefore it is correct. When sphericity is violated, the means stay the same, it is the degrees of freedom and the resulting *F*-ratio that change.

1. Imagine we were interested in the effect of supporters singing on the number of goals scored by soccer teams. We took 10 groups of supporters of 10 different soccer teams and asked them to attend three home games, one at which they were instructed to sing in support of their team (e.g., ‘Come on, you Reds!’), one at which they were instructed to sing negative songs towards the opposition (e.g., ‘You’re getting sacked in the morning!’) and one at which they were instructed to sit quietly. The order of chanting was counterbalanced across groups. An ANOVA with a simple contrasts using the last category as a reference was conducted. Looking at the output tables below, what does the first contrast (Level 1 vs. Level 3) compare?







1. We cannot tell which groups the levels represent from this output.
2. Negative singing vs. positive singing
3. No singing vs. negative singing
4. Positive singing vs. no singing

The correct answer is d) Positive singing vs. no singing

Feedback: If you look at the *Within-Subjects Factors* box, you can see that level 1 = positive singing, level 2 = negative singing and level 3 = no singing.

1. Imagine we were interested in the effect of supporters singing on the number of goals scored by soccer teams. We took 10 groups of supporters of 10 different soccer teams and asked them to attend three home games, one at which they were instructed to sing in support of their team (e.g., ‘Come on, you reds!’), one at which they were instructed to sing negative songs towards the opposition (e.g., ‘You’re getting sacked in the morning!’) and one at which they were instructed to sit quietly. The order of chanting was counterbalanced across groups. A one-way repeated-measures ANOVA was conducted and *post hoc* tests were selected. Looking at the output below, which of the following sentences regarding the pairwise comparisons is correct?









(*Hint*: Look in the box labelled *Within-Subjects Factors* to see which variable corresponded with which level.)

1. Soccer teams scored significantly more goals when their supporters sang positive songs towards them compared to when they sang negative songs towards the opposition, but not compared to when they did not sing at all. Soccer teams scored significantly fewer goals when their supporters sang negative songs compared to when they did not sing at all.
2. Soccer teams scored significantly more goals when their supporters sang positive songs towards them compared to when they sang negative songs towards the opposition, but not compared to when they did not sing at all. Soccer teams did not significantly differ in the number of goals scored when their supporters sang negative songs compared to when they did not sing at all.
3. Soccer teams scored significantly more goals when their supporters sang positive songs towards them compared to when they sang negative songs towards the opposition and compared to when they did not sing at all. Soccer teams did not significantly differ in the number of goals scored when their supporters sang negative songs compared to when they did not sing at all.
4. Soccer teams scored significantly more goals when their supporters sang positive songs towards them compared to when they sang negative songs towards the opposition and when they did not sing at all. Soccer teams scored significantly more goals when their supporters sang negative songs compared to when they did not sing at all.

The correct answer is b) Soccer teams scored significantly more goals when their supporters sang positive songs towards them compared to when they sang negative songs towards the opposition, but not compared to when they did not sing at all. Soccer teams did not significantly differ in the number of goals scored when their supporters sang negative songs compared to when they did not sing at all.

1. Two-way repeated-measures ANOVA compares:
2. Several means when there are two independent variables, and the same entities have been used in some of the conditions.
3. Two means when there are more than two independent variables, and the same entities have been used in all conditions.
4. Several means when there are two independent variables, and the same entities have been used in all conditions.
5. Several means when there are more than two independent variables, and some have been manipulated using the same entities and others have used different entities.

The correct answer is c) Several means when there are two independent variables, and the same entities have been used in all conditions.

1. How many effects will there be from a two-way repeated-measures ANOVA?
2. 4
3. 2
4. 3
5. It will vary depending on how many levels of your independent variable there are.

The correct answer is c) 3.

Feedback: There will be a main effect of each variable and the interaction between the two.

1. When conducting a repeated-measures ANOVA which of the following assumptions is not relevant? (*Hint*: Repeated-measures ANOVA is an extension of the linear model and so all of the sources of potential bias (and counteractive measures) apply.)
2. Homogeneity of variance
3. They are all relevant
4. Sphericity
5. Independent residuals

The correct answer is a) Homogeneity of variance

Feedback: The assumption of homogeneity of variance (the assumption that the variances between groups are roughly equal) is relevant for between-groups ANOVA and not for repeated-measures ANOVA.

**Chapter 17**

1. Which of the following is calculated or presented in the same way for both repeated-measures and between-subjects designs?
2. The simple effects
3. The degrees of freedom
4. The model error
5. Partial eta squared

The correct answer is d) Partial eta squared.

1. A study was conducted to look at whether caffeine improves productivity at work in different conditions. There were two independent variables. The first independent variable was **email**, which had two levels: ‘email access’ and ‘no email access’. The second independent variable was **caffeine**, which also had two levels: ‘caffeinated drink’ and ‘decaffeinated drink’. Different participants took part in each condition. Productivity was recorded at the end of the day on a scale of 0 (I may as well have stayed in bed) to 20 (wow! I got enough work done today to last all year). Looking at the group means in the table below, which of the following statements best describes the data?

|  |  |  |
| --- | --- | --- |
|  | Decaffeinated Drink | Caffeinated Drink |
| No Email | 12.08 | 19.83 |
| Email | 11.98 | 5.49 |

1. A significant interaction effect is likely to be present between caffeine consumption and email access.
2. The effect of caffeine is about the same regardless of whether the person had email access.
3. The effect of email is relatively unaffected by whether the drink was caffeinated.
4. There is likely to be a significant main effect of caffeine.

The correct answer is a) A significant interaction effect is likely to be present between caffeine consumption and email access.

Feedback: Yes, this is correct: for decaffeinated drinks there is little difference between email and no email, but for caffeinated drinks there is.

1. A study was conducted to look at whether caffeine improves productivity at work in different conditions. There were two independent variables. The first independent variable was **email**, which had two levels: ‘email access’ and ‘no email access’. The second independent variable was **caffeine**, which also had two levels: ‘caffeinated drink’ and ‘decaffeinated drink’. Different participants took part in each condition. Productivity was recorded at the end of the day on a scale of 0 (I may as well have stayed in bed) to 20 (wow! I got enough work done today to last all year). Looking at the group means in the table below, which of the interpretations below is correct?

|  |  |  |
| --- | --- | --- |
|  | Decaffeinated Drink | Caffeinated Drink |
| No Email | 12.08 | 19.83 |
| Email | 11.98 | 5.49 |

1. A simple effects analysis is likely to show an effect of email on productivity at both levels of caffeine.
2. A simple effects analysis is likely to show an effect of caffeine on productivity at both levels of email.
3. A simple effects analysis is likely to show an effect of email on productivity for decaffeinated drinks but not caffeinated ones.
4. A simple effects analysis is likely to show an effect of caffeine on productivity only for ‘no email’.

The correct answer is b) A simple effects analysis is likely to show an effect of caffeine on productivity at both levels of email.

Feedback: The effect of caffeine for no email is 19.83 – 12.08 = 7.75; and for email 5.49 – 11.98 = –6.49, both of which are not 0.

1. What type of ANOVA is used when there are two independent variables each with more than two levels, and with different participants taking part in each condition?
2. Factorial
3. One-way independent
4. Mixed
5. One-way between subjects

The correct answer is a) Factorial.

1. Two-way ANOVA is basically the same as one-way ANOVA, except that:
2. We calculate the model sum of squares by looking at the difference between each group mean and the overall mean.
3. The residual sum of squares represents individual differences in performance.
4. The model sum of squares is partitioned into two parts.
5. The model sum of squares is partitioned into three parts.

The correct answer is d) The model sum of squares is partitioned into three parts.

Feedback: The model sum of squares is partitioned into the effect of each of the independent variables and the effect of how these variables interact.

1. An experiment was done to look at whether there is an effect of the number of hours spent practising a musical instrument and gender on the level of musical ability. A sample of 30 (15 men and 15 women) participants who had never learnt to play a musical instrument before were recruited. Participants were randomly allocated to one of three groups that varied in the number of hours they would spend practising every day for 1 year (0 hours, 1 hours, 2 hours). Men and women were divided equally across groups. All participants had a one-hour lesson each week over the course of the year, after which their level of musical skill was measured on a 10-point scale ranging from 0 (you can’t play for toffee) to 10 (‘Are you Mozart reincarnated?’). Which of the following tests could we use to analyse these data?
2. Two-way independent ANOVA
3. Two-way repeated-measures ANOVA
4. Three-way ANOVA
5. *t*-test

The correct answer is a) Two-way independent ANOVA

Feedback: We have two independent variables, ‘gender’ and ‘number of hours spent practising’, and different participants took part in each condition.

1. Field and Lawson (2003) reported the effects of giving children aged 7–9 years positive, negative or no information about novel animals (Australian marsupials). This variable was called ‘Infotype’. The gender of the child was also examined. The outcome was the time taken for the children to put their hand in a box in which they believed either the positive, negative, or no information animal was housed (positive values = longer than average approach times, negative values = shorter than average approach times). Some simple contrasts were performed on the data. Based on the SPSS output given, which of the following statements is true? (Levels of Infotype were entered in the following order: negative information, positive information, no information.) (*Hint*: The order that the levels of the independent variable are entered corresponds to the level of the contrast in the output: negative information = level 1, positive information = level 2, no information = level 3.)







1. Approach times for the box containing the positive animal were significantly shorter to the box containing the control (no information) animal.
2. Approach times for the box containing the negative animal were significantly longer than for the box containing the control (no information) animal.
3. Approach times for the box containing the negative animal were not significantly different from those for the box containing the positive information animal.
4. The profile of results were different for boys and girls.

The correct answer is b) Approach times for the box containing the negative animal were significantly longer than for the box containing the control (no information) animal.