7.1 Probability, Significance and Type I and Type II Errors: 30-minute Pack

AQA 7182 · about 30 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions.

1 Distinguish one-tailed and two-tailed tests, name the link to directional/non-directional hypotheses, and apply to a music tempo study.

(Total for Question 1 is 3 marks)

2 Define Type I and Type II errors for an anxiety therapy trial, referring to the null hypothesis and the presence or absence of a real effect.

(Total for Question 2 is 4 marks)

3 Consider a drug trial where $\alpha = 0.01$ is chosen because of side-effect risk. Give two reasons why this is sensible and one disadvantage.

(Total for Question 3 is 3 marks)

© Revision Library | revisionlibrary.co.uk

4 Distinguish statistical significance from practical significance and give an example where significance lacks practical importance.

(Total for Question 4 is 4 marks)

5 Recommend one-tailed or two-tailed test at $\alpha = 0.05$ given prior evidence of increase but serious consequences if effect were in the opposite direction. Justify regarding Type I/II risks.

(Total for Question 5 is 4 marks)

6 Discuss the consequences of choosing significance levels that are too strict or too lenient, evaluating trade-offs between Type I and Type II errors, impact on science and real-world decisions; include applied examples.

(Total for Question 6 is 8 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 7.1 Probability, Significance and Type I and Type II Errors: 30-minute Pack

Question 1

one-tailed test looks for an effect in one specified direction and is used with a directional hypothesis oe **M1**

two-tailed test looks for an effect in either direction and is used with a non-directional hypothesis oe **A1**

applies to example: directional hypothesis 'faster tempo increases speed' would use one-tailed; non-directional 'tempo affects speed' would use two-tailed oe **A1**

Answer: A one-tailed test assesses an effect in a specified direction and matches a directional hypothesis; a two-tailed test assesses effects in either direction and matches a non-directional hypothesis. In the music example, predicting faster tempo increases speed is directional and would justify a one-tailed test; predicting only that tempo affects speed is non-directional and requires a two-tailed test.

Question 2

Type I error: rejecting the null hypothesis when the null hypothesis is actually true, i.e. concluding the therapy works when it does not oe **M1**

labels this a false positive and links to the anxiety therapy context, e.g. offering an ineffective treatment oe **A1**

Type II error: failing to reject the null hypothesis when the null hypothesis is false, i.e. concluding the therapy does not work when it actually does oe **M1**

labels this a false negative and links to the anxiety therapy context, e.g. missing a beneficial treatment oe **A1**

Answer: Type I error: rejecting the null hypothesis when it is true, a false positive, e.g. concluding the new anxiety therapy works when it does not. Type II error: failing to reject the null hypothesis when it is false, a false negative, e.g. concluding the therapy does not work when it actually does.

Question 3

reason 1: reduces the risk of a Type I error, so fewer false positives and lower chance of approving a harmful or ineffective drug oe **B1**

reason 2: protects patients by requiring stronger evidence before concluding the drug is effective, which is important when there are safety concerns oe **B1**

disadvantage: increases risk of a Type II error, so a genuinely effective drug might be missed, delaying a beneficial treatment oe **B1**

Answer: Sensible because it reduces Type I error risk so fewer false positives and reduces chance of approving a harmful or ineffective drug, and it demands stronger evidence before exposing patients to side effects. Disadvantage is higher Type II error risk, so a real beneficial drug might be missed.

Question 4

statistical significance refers to whether a result is unlikely to have occurred by chance according to the chosen α level oe **M1**

practical significance refers to whether the size of the effect is large or important enough to matter in real-world terms oe **M1**

gives an example: a drug that reduces symptom score by 0.5 points on a 100 point scale might be statistically significant in a large sample but clinically trivial oe **A1**

explains why the example shows limited practical importance despite statistical significance oe **A1**

Answer: Statistical significance means the result is unlikely under the null at the chosen α; practical significance asks whether the effect size matters in real life. Example: a treatment that lowers a 100 point symptom scale by 0.5 points may be statistically significant in a large sample but clinically meaningless.

Question 5

recommends a two-tailed test if consequences of a mistaken conclusion in the opposite direction are serious, even if prior evidence suggests an increase oe **M1**

explains that a two-tailed test guards against unexpected effects in either direction, reducing the risk of a Type I error in the unpredicted direction oe **A1**

acknowledges that a one-tailed test would give more power to detect the predicted increase and reduce Type II error risk oe **M1**

justifies the final choice by weighing that avoiding a harmful false positive in the opposite direction is more important than the small gain in power from a one-tailed test oe **A1**

Answer: Recommend a two-tailed test at α = 0.05 because, despite prior evidence for an increase, the serious consequences of being wrong in the opposite direction mean the researcher should guard against effects both ways. A one-tailed test would increase power and reduce Type II risk for the predicted increase but risks missing or misclassifying an effect in the unpredicted direction.

© Revision Library | revisionlibrary.co.uk

Question 6

Level 4 Knowledge and understanding of significance levels and error trade-offs is accurate and detailed. The discussion evaluates consequences for research practice, replication and applied decisions, using well chosen examples. Counterarguments and nuances, such as power and effect size considerations, are used to support evaluation. The answer is well organised and uses appropriate terminology.	**7-8**
Level 3 Knowledge is clear and mostly accurate. The discussion considers consequences for research and real-world decisions and includes some evaluation and examples, although some points may be less developed. Some reference to power or effect size is present.	**5-6**
Level 2 Knowledge is present but basic. The answer lists consequences of strict and lenient alpha levels with limited evaluation or minor inaccuracies. Examples may be general and links to power or replication are superficial.	**3-4**
Level 1 Very limited knowledge. Response is largely descriptive or confused, with little or no evaluation and no useful examples.	**1-2**
Level 0 No creditable content.	**0**

Indicative content:

- Define what is meant by a significance level (alpha) and remind that it sets the probability threshold for rejecting the null hypothesis, linking to Type I error rate.
- Explain the effect of setting alpha very strict (e.g. 0.01): reduces Type I errors but increases Type II errors, lowers statistical power, and may cause real but small effects to be missed. Applied example: in drug safety, strict alpha reduces approving unsafe drugs but may prevent beneficial drugs reaching patients.
- Explain the effect of setting alpha lenient (e.g. 0.10): reduces Type II errors and increases power for detecting effects but raises Type I error risk, increasing false positives. Applied example: in education, lenient alpha might promote many unproven programmes, wasting resources.
- Evaluate impact on scientific progress and replication: strict alpha may slow discovery and discourage publication of marginal but real effects, contributing to file-drawer problems; lenient alpha may flood literature with false positives, reducing replicability and undermining trust.
- Discuss interaction with sample size and effect size: appropriate alpha cannot be chosen in isolation, because larger samples can allow strict alpha while retaining power, and reporting effect sizes and confidence intervals mitigates over-reliance on p-values.
- Consider ethical and practical trade-offs: the context matters, so policy decisions should weigh costs of false positives versus false negatives, and preregistration and replication can reduce the harms of both types of errors.
- Conclude with balanced judgement: neither extreme is universally correct; justify choice by context, use power calculations, report full statistics, and emphasise replication and effect sizes to support robust conclusions.

© Revision Library | revisionlibrary.co.uk