

Binary variable - which take two possible values.

of 1 or ordered categorical

SECTION 2: SHORT QUESTIONS - STATISTICS

Age, educational level - categorical (ordinal)
 Answer either Question 2.1 or Question 2.2

2.1 A randomised controlled trial (RCT) is being conducted to find out whether palliative care improves the mental health of people with advanced HIV. The study also measures age group, gender, time since HIV diagnosis, presence of physical symptoms and pain, and socio-economic status. Mental health is measured from 0 to 100, and decimals are possible (e.g. 55.7). Symptoms and pain are recorded from 0=none to 5=overwhelming, using whole numbers only.

Categorical

ordinal categorical variable

- i. What is the primary outcome of the study? (1 mark)
- ii. What kind of variable is the primary outcome? (1 mark)
- iii. Name:
 - An independent (or exposure) binary variable (1 mark)
 - An independent ordered categorical (or ordinal) variable (1 mark)
 - An independent continuous variable (1 mark)

Categorical.

Gender.

Age group Socio-economic status

iv. What are the secondary outcomes and what kind of variables are they? (3 marks)

Time since HIV diagnosis

Age group, Gender, time since HIV diagnosis, presence of physical symptoms and pain, Socio-economic status

The RCT will include 120 people. After 30 were recruited, some early analysis was carried out. Here are the results using symptom score:

Table 1: symptom score summary data by study arm

	Intervention group	Control group
Minimum	0	0
Maximum	5	4
Median	3	2
Inter-quartile range	2-4	1.5-3

- v. Draw box and whisker plots of the data in Table 1 (3 marks)
- vi. Do you think the symptoms in the two groups are different? Explain your reasoning. (2 marks)
- vii. What statistical test would you use to look for a difference between the groups? (2 marks)
- viii. Why would you use this test, and are there any disadvantages of using it? (3 marks)

You run the test with 5% significance. The test parameter is 11.5 and $p=0.12$.

- ix. What does 5% significance mean? (2 marks)
- x. Explain what $p=0.12$ means in this context as fully as possible. (3 marks)
- xi. Why do you think a study design might include early analysis like this before the full sample has been recruited? (2 marks)

Total of 25 marks

2.2 A telephone survey was conducted in two European countries to find out people's preferences for care at the end of life. One question asked whether participants would choose to die at home or somewhere else.

Table 2: Frequency table of those who wish to die at home by age group

Age group	People sampled	Would want to die at home
16-29	187	128
30-39	259	182
40-49	331	177
50-59	365	182
60-69	364	204

- i. What is the outcome variable in Table 2? What kind of variable is it? (2 marks) *- categorical*
- ii. Calculate the percentage in each age group who would want to die at home. (2 marks) *people's preference - factor*
- iii. Draw a bar chart representing the proportions who wish to die at home by age group. (3 marks)
- iv. Describe the results shown by the graph. (2 marks)

You want to find out whether the relationship between age and wish to die at home is consistent in different countries. The data for two countries, England and Portugal, are presented below. In both countries, age follows an approximately Normal distribution.

Table 3: mean age and preference to die at home by country

Country	n	Mean age	Median	Standard error of mean	95% confidence interval
England	1332	45.3	45	0.54	44.3 to 46.4
Portugal	1223	49.1	51	0.45	48.2 to 50.0

- v. Define mean, median, standard error, and 95% confidence interval (4 marks) *middle score of a set of data that have been averaged in*
- vi. Which is a more useful summary here; the mean or the median? (1 mark) *side of the printed*

mode - most frequent score of our data set

median over even when the data is skewed.

etc. What variable?

vii. Describe and compare the age distributions in the two countries, including the importance of any differences. (3 marks)

viii. What statistical test could you use to compare the age of the two populations? (2 marks)

ix. State the null hypothesis. (2 marks)

You conduct the test, with a 5% significance level. The test parameter is 5.43 and $p < 0.001$.

x. What do the results of this test tell you? (2 marks)

xi. Describe how you think age and country are associated with wish to die at home, giving reasons. (2 marks)

Total of 25 marks

SECTION 2: SHORT QUESTIONS - STATISTICS

Answer either Question 2.1 or Question 2.2

- 2.1 A study was conducted to examine the health related quality of life (QoL) among cancer patients receiving a new palliative care service, and the factors associated with the QoL.

The data collected during the course of the study included age, sex, cancer site, time since cancer diagnosis, *Eastern Cooperative Oncology Group* (ECOG) performance status, and QoL. The ECOG performance status was recorded from 0=fully active to 5=death, using whole numbers only. The QoL was measured by the Visual Analogue Scale (VAS) on a score ranging from 0 to 100 (decimals were possible e.g. 32.5). A higher score means better QoL.

- i. What is the dependent variable in this study? (1 mark)
- ii. What kind of variable is the dependent variable? (1 mark)
- iii. Name:
A continuous variable; (1 mark)
A binary variable; (1 mark)
An ordered categorical (or ordinal) variable (1 mark)

A total of 16 patients were included in the study, the mean QoL score was 36, the standard deviation was 12.5.

- iv. The formula for standard error is $SE = \frac{SD}{\sqrt{N}}$, where SD is the standard deviation, N is the sample size. Calculate the standard error of the mean QoL score, rounded to two decimal places. (2 marks)
- v. Calculate the 95% confidence interval of the mean QoL score (assuming normal distribution, rounded to one decimal place) and explain the meaning of the 95% CI in this context. (3 marks)
- vi. The mean QoL score of patients in another study (A) is 62; do you think it is different from the mean QoL score of the cancer population from which the study sample was drawn? Briefly explain your reasoning. (3 marks)
- vii. What statistical test would you use to examine the difference between the cancer population and the healthy population? (2 marks)
- viii. Why would you use this test, and are there any problems using it? (2 marks)

You run the statistical test and get a p value <0.001.

- ix. Explain what p<0.001 means in this context as fully as possible. (2 marks)
- x. In addition to mean and standard deviation, what other alternative measures can you use to describe the central tendency and the spread of the QoL scores in the study sample? (3 marks)
- xi. Describe when to use which of the above alternative measures of the central tendency and the spread. (3 marks)

2.2 A total of 458 patients with advanced cancer were assessed for symptom prevalence and severity in a cancer clinic. The following table lists the frequency distribution by symptom.

Table 1: Frequency distribution by symptoms

Symptom	Frequency
Pain	380
Lack of energy	279
Anorexia	302
Fatigue	316
Weakness	284
Dry mouth	254

- i. What is the dependent variable in the study? What kind of variable is it? (2 marks)
- ii. Calculate symptom prevalence for the symptoms listed in Table 1, keeping to two decimal places. (3 marks)
- iii. The formula for the standard error is $SE = \frac{P(1-P)}{\sqrt{\text{Sample size}}}$, where P is the prevalence (range: 0-1). Calculate the SE and 95% confidence interval of the pain prevalence. (2 marks)

You want to find out whether the pain prevalence is different between men and women. The frequency distribution is summarised in Table 2.

Table 2: Pain prevalence in advanced cancer patients by sex

	Men	Women	Total
Pain	156	224	380
No pain	34	44	78
Total	190	268	458

iv. What statistical test(s) would you use to compare whether the pain prevalence differed by sex? Briefly explain your reasoning. (2 marks)

v. State your null hypothesis. (2 marks)

You run the test at the significance level of 0.05, and get a p value=0.71.)

vi. Explain what the significance level of 0.05 means. (2 marks)

vii. Explain what $p=0.71$ means in this context. (2 marks)

viii. Do you think the pain prevalence differs by sex? Briefly explain your answer. What other factors might you take into account to reach your conclusion? (4 marks)

The symptom severity was rated on a 5-point Likert scale (0=No and 5=most severe).

ix. What statistical measures would you use to describe the symptom severity score? Briefly explain your answer. (3 marks)

x. What statistical test you would use to compare the differences in symptom severity scores between men and women. Briefly explain your answer. (3 marks)