

THE OPEN UNIVERSITY

Practical modern statistics (M249) Diagnostic Quiz

[Press ↓ to begin]

1. Introduction

In order to study Practical modern statistics (M249), it is assumed that you have some knowledge of statistics as well as some basic mathematical skills. This quiz covers a number of key mathematical and statistical areas with which you should be familiar.

For M249 you require a good knowledge of statistical ideas and methods at an introductory level. The statistical prerequisites are revised in the Introduction to statistical modelling ([click on link](#)). They include: graphical and numerical data summaries; the basic statistical distributions including the normal, exponential, uniform, binomial and Poisson distributions; confidence intervals and significance tests, correlations and contingency tables. No knowledge of regression is required. If you have taken Analysing data (M248) you should have the required statistical background.

You are also expected to be familiar with mathematical notation, to be able to follow short algebraic arguments, to handle the logarithm and exponential functions, and to use formulas. If you have passed M248, or completed study of mathematics at Stage 1, your mathematical knowledge should be ample. No knowledge of calculus, differentiation or integration is required.

(continued on following page)

Try each question for yourself, using your calculator if you wish, then click on the green section letter (e.g. '(a)') to see the solution. Click on the symbol at the end of the solution to return to the question. Use the \uparrow and \downarrow keys to move from Section to Section.

There is some advice on evaluating your performance at the end of the quiz.

2. Decimals and fractions

EXERCISE 1.

- (a) Evaluate 2.8372×1.8205 to 3 decimal places.
- (b) Evaluate $7.1946 \div 2.011$ to 3 significant figures.

EXERCISE 2.

- (a) Write down $\frac{7}{8} \times \frac{2}{21}$ in its simplest fractional form.
- (b) Write down $\frac{4}{10} \div \frac{16}{15}$ in its simplest fractional form.

3. Formulas and equations

EXERCISE 3.

- (a) Consider the following sample data. In a random sample of size $n = 15$, the sample mean was found to be $\bar{x} = 10.25$ and the sample standard deviation $s = 2.326$. The 95% confidence limits (that is, the end-points of the 95% confidence interval) for the mean μ of the population from which the sample was drawn are defined to be

$$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$$

Evaluate this confidence interval to 2 decimal places, using the summary statistics given.

- (b) If $\frac{5}{2}y + 14 = y + 8$, find y .

4. Powers, logarithms and the exponential function

EXERCISE 4.

- (a) Write $3^3 \times 3^4$ in the form 3^n and hence evaluate it.
- (b) Write $(3^3)^4$ in the form 3^n and hence evaluate it.
- (c) Write $3^4/3^4$ in the form 3^n and hence evaluate it.

EXERCISE 5. This exercise is about logarithms to base e , which are sometimes called ‘natural logarithms’ and may be denoted by ‘ \ln ’, ‘ \log_e ’ or just ‘ \log ’.

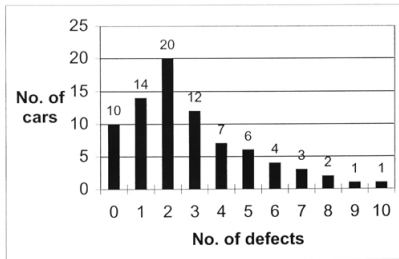
- (a) To 4 decimal places, $\log 3 = 1.0986$ and $\log 4 = 1.3863$. Without evaluating any other logarithms, use these values to find $\log 12$ to 3 decimal places.
- (b) If $\log x = 1.2$, what is x , to 3 decimal places?

EXERCISE 6.

- (a) Suppose that $\exp(x) = 2$, for some number x . Without evaluating x , use this result to find $\exp(-x)$ and $\exp(2x)$.
- (b) If $\exp(x) = 3.2$, what is x , to 3 decimal places?

5. Graphs 1: Bar charts and histograms

EXERCISE 7. The bar chart below displays the numbers of defects found in a sample of 80 cars by the quality control division of a major vehicle manufacturer.



[Press ↓ for questions.]

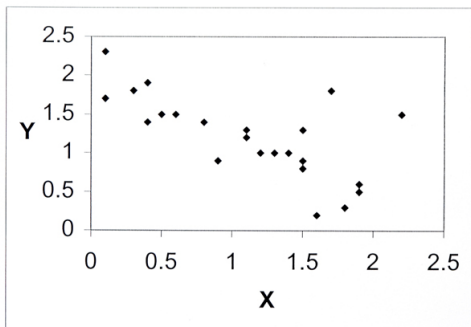
- (a) What proportion of cars in the sample have at most 4 faults?
- (b) What proportion of cars in the sample have between 4 and 6 faults inclusive?
- (c) Without performing any calculations, would you expect the sample mean number of faults to be greater than the sample median? Give a reason for your answer.

[Press ↑ to return to graph.]

[Press ↓ for next exercise.]

6. Graphs 2: Scatterplots

EXERCISE 8. Consider the scatterplot below:



[Press ↓ for questions.]

- (a) How would you characterise the trend in the data displayed by this graph? (For example, would you say that it is linear or non-linear, increasing, decreasing or constant?)
- (b) How would you describe the degree of scatter? (For example, would you say that it is increasing, decreasing or constant as x increases?)

[Press \uparrow to return to graph.]

[Press \downarrow for next exercise.]

7. Measures of location and dispersion

EXERCISE 9. Numbers of defective components in a random sample of 4 production batches were found to be 8, 2, 2 and 20.

- (a) Find the mean, median and mode of this sample.
- (b) Calculate the sample standard deviation, to 3 decimal places.

8. Probabilities from tables

EXERCISE 10. Consider the data tabulated below, which summarises the beer preferences of a sample of 150 beer drinkers, categorised by gender:

	Bitter	Lager	Stout
Male	40	20	20
Female	30	35	5

- (a) What percentage of the sample drink lager?
- (b) What percentage of the sample comprises female lager drinkers?
- (c) What percentage of females in the sample drink lager?
- (d) What percentage of lager drinkers in the sample are female?

9. Probability distributions 1: Discrete distributions

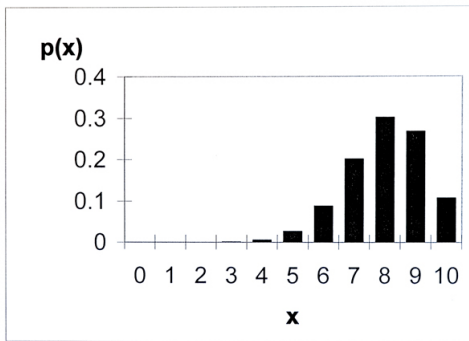
EXERCISE 11. Consider the probability distribution of the discrete random variable x tabulated below.

x	-1	0	1	2
$p(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{16}$	$\frac{1}{16}$

- (a) Find the mean of the distribution.
- (b) Find $p(x \geq 1)$.

10. Probability distributions 2: Identify the distribution

EXERCISE 12. Consider the probability distribution represented by the following probability mass function.



[Press ↓ for questions.]

Do you think that this probability mass function (p.m.f.) represents:

- (a) a normal distribution?
- (b) a binomial distribution?
- (c) a Poisson distribution?

[Press \uparrow to return to graph.]

[Press \downarrow for next exercise.]

11. Hypothesis testing

EXERCISE 13. A hypothesis test has been undertaken to test the null hypothesis that there is no difference between the proportions of girls and boys who get a grade 'A' in GCSE Mathematics. The significance probability (or p value) was found to be 0.0034.

(a) Which of the following two statements is true?

(A) 'There is strong evidence that the underlying proportions are the same.'

OR

(B) 'There is strong evidence that the underlying proportions are *not* the same.'

12. Confidence intervals

EXERCISE 14. Suppose that $(-2.6, 3.4)$ is a 95% confidence interval for some parameter θ .

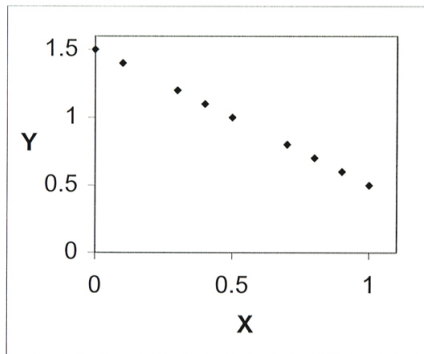
Classify the following statements as true or false:

- (a) 3.3 lies outside the corresponding 99% confidence interval for θ .
- (b) On the basis of these data, it is very implausible that θ should equal zero.

13. Correlation

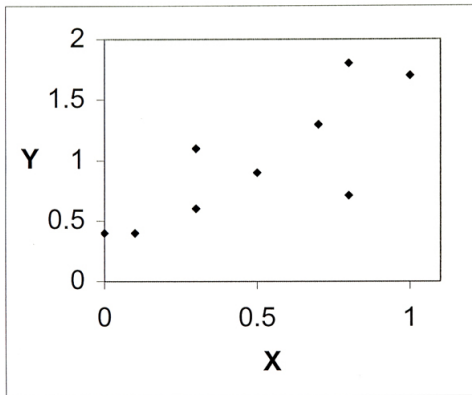
EXERCISE 15.

- (a) For the data in the scatterplot below, do you think the correlation coefficient is likely to be -1 , 0 or 0.8 ?



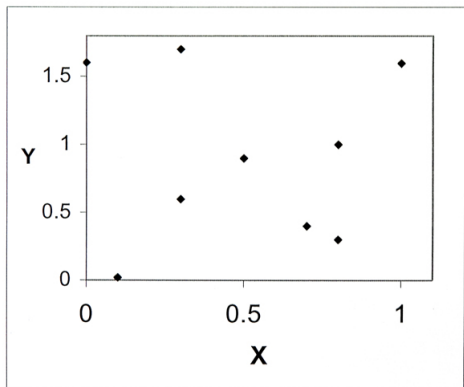
EXERCISE 16.

- (a) For the data in the scatterplot below, do you think the correlation coefficient is likely to be -1 , 0 or 0.8 ?



EXERCISE 17.

- (a) For the data in the scatterplot below, do you think the correlation coefficient is likely to be -1, 0 or 0.8?



14. Post-mortem

If you had difficulty in answering some of these questions, you might find it useful to look at the pre-registration version of the Introduction to statistical modelling, (click on link) which comprises a revision of the pre-requisites for M249. You should only attempt to read Sections 1, 3, 4 and 5; Sections 2 and 6 introduce the statistical package SPSS which you will not receive until you have registered for the course. (If you decide to study M249, you will need to work through this unit again, including the computing sections.)

If you found the mathematics exercises 1 - 6 difficult, you might consider studying one of the mathematics entry level modules, MU123 or MST124, before embarking on M249. On the other hand, if you found the statistics exercises 7 - 17 difficult, you would be advised to study Analysing data (M248) before M249.

If you have any queries about your suitability for the course, you should contact your Student Support Team via StudentHome.

Solutions to Exercises

Exercise 1(a) $2.8372 \times 1.8205 = 5.1651226$, which is 5.165 to 3 decimal places.

Remember that, when rounding to 3 decimal places, if the digit in the 4th decimal place is in the range $0 \cdots 4$, the number is rounded down, that is, the digit in the 3rd decimal place is left unchanged. If it is in the range $5 \cdots 9$, the number is rounded up, that is, the digit in the 3rd decimal place is increased by 1.

This rule can be applied however many decimal places you are asked to round to.



Exercise 1(b) $7.1946 \div 2.011 = 3.577623073$ to the limits of calculator accuracy, which is 3.58 to 3 significant figures.



Exercise 2(a) Applying the usual cancellation rules, we have

$$\begin{aligned}\frac{7}{8} \times \frac{2}{21} &= \frac{1}{8} \times \frac{2}{3} \\ &= \frac{1}{4} \times \frac{1}{3} \\ &= \frac{1}{12}\end{aligned}$$

(after dividing above and below by 7, then by 2).



Exercise 2(b) Remembering that $\frac{a}{b} \div \frac{c}{d} = \frac{a}{b} \times \frac{d}{c}$, where a, b, c, d represent any non-zero real numbers, and applying the cancellation rules, we get

$$\begin{aligned}\frac{4}{10} \div \frac{16}{15} &= \frac{4}{10} \times \frac{15}{16} \\ &= \frac{4}{2} \times \frac{3}{16} \\ &= \frac{1}{2} \times \frac{3}{4} \\ &= \frac{3}{8}\end{aligned}$$



Exercise 3(a) Substituting in the given formula, we get

$$\begin{aligned}\bar{x} \pm 1.96 \frac{s}{\sqrt{n}} &= 10.25 \pm 1.96 \left(\frac{2.326}{\sqrt{15}} \right) \\ &= 10.25 \pm (1.96 \times 0.600570617) \\ &= 10.25 \pm 1.17711841\end{aligned}$$

To 2 decimal places, the required interval is therefore (9.07, 11.43).



Exercise 3(b) Applying the manipulation rules yields

$$\begin{aligned} & \frac{5}{2}y + 14 = y + 8 \\ \iff & \frac{5}{2}y - y = 8 - 14 \\ \iff & \frac{3}{2}y = -6 \\ \iff & y = -4 \end{aligned}$$

(The logical symbol \iff means ‘if and only if’ or ‘is equivalent to’.)



Exercise 4(a) Using the rule $x^a \times x^b = x^{a+b}$ gives

$$\begin{aligned}3^3 \times 3^4 &= 3^{3+4} \\ &= 3^7 \\ &= 2187\end{aligned}$$



Exercise 4(b) Using the rule $(x^a)^b = x^{ab}$ gives

$$\begin{aligned} (3^3)^4 &= 3^{3 \times 4} \\ &= 3^{12} \\ &= 531441 \end{aligned}$$



Exercise 4(c) Using the rule $x^a/x^b = x^{a-b}$ gives

$$\begin{aligned}3^4/3^4 &= 3^{4-4} \\ &= 3^0 \\ &= 1\end{aligned}$$

remembering that, for any number x , $x^0 = 1$.



Exercise 5(a) Remembering that, for any positive real numbers x and y , $\log(xy) = \log x + \log y$, we get

$$\begin{aligned}\log 12 &= \log(3 \times 4) \\ &= \log 3 + \log 4 \\ &= 1.0986 + 1.3863 \\ &= 2.4849 \\ &= 2.485\end{aligned}$$

to 3 decimal places.



Exercise 5(b) Remembering that \log is the inverse function of \exp , the exponential function, we have

$$\begin{aligned}x &= \exp(1.2) \\ &= e^{1.2} \\ &= 3.320116923 \\ &= 3.320\end{aligned}$$

to 3 decimal places.



Exercise 6(a) Since $\exp(-x) = e^{-x} = \frac{1}{e^x}$ for all x , we have

$$\begin{aligned}\exp(-x) &= \frac{1}{\exp(x)} \\ &= \frac{1}{2}\end{aligned}$$

Since $\exp(2x) = e^{2x} = (e^x)^2$ for all x , we have

$$\begin{aligned}\exp(2x) &= (\exp(x))^2 \\ &= 2^2 \\ &= 4\end{aligned}$$



Exercise 6(b) Again remembering that \log is the inverse function of \exp , we have

$$\begin{aligned}x &= \log(3.2) \\ &= 1.16315081 \\ &= 1.163\end{aligned}$$

to 3 decimal places.



Exercise 7(a) The number of cars with at most 4 faults is

$$10 + 14 + 20 + 12 + 7 = 63.$$

The proportion is therefore $\frac{63}{80} \cong 0.79$.



Exercise 7(b) The number of cars with between 4 and 6 faults inclusive is $7 + 6 + 4 = 17$. The proportion is therefore $\frac{17}{80} \cong 0.21$.



Exercise 7(c) Yes. For right-skew data, the sample mean is greater than the sample median.



Exercise 8(a) A visual inspection suggests that x and y could well be linearly related. y appears to be decreasing as x increases.



Exercise 8(b) The degree of scatter appears to be increasing as x increases.



Exercise 9(a) The mean is $\frac{8+2+2+20}{4} = 8$.

The ordered sample is 2, 2, 8, 20, so the median is $\frac{2+8}{2} = 5$.

The mode is the most frequently-occurring value, in this case 2.



Exercise 9(b)

$$\begin{aligned}s &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \sqrt{\frac{1}{3} \left\{ (8-8)^2 + (2-8)^2 + (2-8)^2 + (20-8)^2 \right\}} \\ &= \sqrt{\frac{216}{3}} \\ &= \sqrt{72} \\ &= 8.485281374 \\ &\cong 8.485,\end{aligned}$$

to 3 decimal places.



Exercise 10(a) The expanded table is:

	Bitter	Lager	Stout	Totals
Male	40	20	20	80
Female	30	35	5	70
Totals	70	55	25	150

The percentage of the sample who drink lager is therefore

$$\frac{55}{150} \cong 36.67\%,$$

to 2 decimal places.



Exercise 10(b) The expanded table is:

	Bitter	Lager	Stout	Totals
Male	40	20	20	80
Female	30	35	5	70
Totals	70	55	25	150

The percentage of the sample who are female lager drinkers is therefore

$$\frac{35}{150} \cong 23.33\%,$$

to 2 decimal places.



Exercise 10(c) The expanded table is:

	Bitter	Lager	Stout	Totals
Male	40	20	20	80
Female	30	35	5	70
Totals	70	55	25	150

The percentage of females in the sample who drink lager is therefore

$$\frac{35}{70} = 50\%.$$



Exercise 10(d) The expanded table is:

	Bitter	Lager	Stout	Totals
Male	40	20	20	80
Female	30	35	5	70
Totals	70	55	25	150

The percentage of lager drinkers in the sample who are female is therefore

$$\frac{35}{55} \cong 63.64\%,$$

to 2 decimal places.



Exercise 11(a) The mean is

$$\begin{aligned} E(x) &= \left((-1) \times \frac{1}{4} \right) + \left(0 \times \frac{1}{2} \right) + \left(1 \times \frac{3}{16} \right) + \left(2 \times \frac{1}{16} \right) \\ &= -\frac{1}{4} + 0 + \frac{3}{16} + \frac{1}{8} \\ &= \frac{1}{16} \end{aligned}$$



Exercise 11(b)

$$\begin{aligned} p(x \geq 1) &= p(x = 1 \text{ or } x = 2) \\ &= p(x = 1) + p(x = 2) \\ &= \frac{3}{16} + \frac{1}{16} \\ &= \frac{1}{4} \end{aligned}$$



Exercise 12(a) The p.m.f. does not represent a normal distribution since it is discrete, whereas the normal distribution is continuous. Also, the p.m.f. is not symmetrical, whereas the normal probability density function is symmetrical about the mean.



Exercise 12(b) The p.m.f. has the characteristics associated with a binomial distribution, in that it is discrete and unimodal and has a finite range. In fact, it is the distribution $B(10, 0.8)$.



Exercise 12(c) The p.m.f. does not represent a Poisson distribution since the Poisson distribution is right-skew.



Exercise 13(a) The significance probability is the probability of obtaining data which are at least as extreme as those observed if the null hypothesis were true. This probability is very small and so we have strong evidence against the null hypothesis. It is highly unlikely that the underlying proportions are the same. Thus statement (A) is incorrect and statement (B) is correct.



Exercise 14(a) The statement is false. The 99% confidence interval for θ is centred on the same value (in this case 0.4) as the 95% confidence interval and is wider. So, since 3.3 lies inside the 95% confidence interval, it must also lie inside the 99% confidence interval.



Exercise 14(b) The statement is false. The 95% confidence interval gives a range of values of θ which are plausible at the 95% confidence level. Since this interval contains zero, it is plausible that $\theta = 0$.



Exercise 15(a) The data points lie on a straight line with negative gradient. The correlation coefficient is -1 .



Exercise 16(a) The data points appear to lie fairly close to a straight line with positive gradient. A correlation coefficient of 0.8 would be appropriate here.



Exercise 17(a) There appears to be no evidence of any linear trend in these data. A correlation coefficient of 0 would be appropriate here.

