THE OPEN UNIVERSITY

Mathematical statistics (M347) Diagnostic Quiz

[Press the ↓ key to begin]

Last Revision Date: March 2015

Version 1.1

1 Introduction

In embarking on M347, basic mathematical competence is as important as basic knowledge of statistics.

You should have a basic knowledge of the ideas and concepts of statistical science at the level of Analysing data (M248). Relevant topics include: normal, Poisson and binomial distributions; the central limit theorem; point estimation; maximum likelihood estimation; confidence intervals; hypothesis testing; simple linear regression; correlation. All these are reviewed in the module.

It would be an advantage if you have also studied Practical modern statistics (M249), especially Book 4 on *Bayesian statistics*. However, such knowledge is not assumed but redeveloped from scratch.

You are also expected to have good university-level mathematical competence. This could be acquired from studying both Essential mathematics 1 (MST124) and Essential mathematics 2 (MST125) or their predecessors MST121 and MS221. The most relevant mathematical techniques are calculus, algebra and matrices. The more at easy you are with basic differentiation and integration the better; there will be quite a lot of algebraic manipulation; matrix properties and manipulations will be kept simple.

This quiz covers a number of key mathematical and statistical areas with which you should be familiar.

Try each question for yourself, using a pencil and paper and your calculator where appropriate. Use the ↑ and ↓ keys to move from Section to Section.

Click on the red arrow (example right) to navigate from the exercise to the solution.

Click on the blue arrow (example right) to navigate from the solution back to the exercise.

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

2 Summation notation

EXERCISE 1

(a) Write out in full the expression $\sum_{i=0}^{3} i^2 a_i x_i$.

(b) Let $x_1 = 1$, $x_2 = 2$ and $x_3 = -2$. Calculate $\sum_{i=1}^{3} x_i^2$.

3 Functions, formulas and equations

EXERCISE 2

Consider the function $f(x) = (1 + x)^2(1 - x)^3$, $-1 < x < 1$. Evaluate $f(0)$ and $f(\frac{1}{2})$.

EXERCISE 3

Consider the following sample data. In a random sample of size $n = 12$, the sample mean was found to be $\bar{x} = 1.521$ and the sample standard deviation $s = 0.614$. 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

$$\bar{x} \pm 1.960 \frac{s}{\sqrt{n}}.$$

Evaluate this confidence interval using the summary statistics given. Give your answer correct to 2 decimal places.

EXERCISE 4

If $\frac{1}{2}y + 4 - 2x = y - 1$, write y as a function of x.

4 Powers

EXERCISE 5

Write each of the following in the form x^n, where n is some number.

(a) $1/x^3$, (b) $x^2 \times x^4$.

EXERCISE 6

Express the fourth root of 0.4 as a power. Use your calculator to evaluate it correct to four decimal places.

5 Logarithms and exponentials

EXERCISE 7

In M347, logarithms are always natural logarithms, that is, logs to base e. This may be ln on your calculator.

(a) To four decimal places, $\log 2 = 0.6931$ and $\log 6 = 1.7918$. Without evaluating any other logarithms, use these values to find $\log 12$ correct to 3 decimal places.

(b) Expand the logarithm of the expression $3\sqrt{x}$ (you need do no arithmetical calculations).

(c) If $\log x = 1.5$, what is x, correct to 3 decimal places?

EXERCISE 8

(a) Suppose that $\exp(x) = 2$, for some number x. Without evaluating x, use this result to find $\exp(-x)$ and $\exp(3x)$.

(b) Simplify $\exp(2\log x)$.

(c) If $\exp(x) = 1.5$, what is x, correct to 3 decimal places?

6 Differentiation

EXERCISE 9

Differentiate each of the following with respect to x.

(a) x^4 and $\dfrac{1}{x^2}$

(b) e^{-3x} and $\log(1 - x^2)$

(c) xe^{2x}

(d) $\dfrac{0.2x}{1 - 0.8x}$

7 Integration

EXERCISE 10

Work out each of th e following integrals.

(a) $\int x\,dx$ and $\int_2^{\infty} \frac{1}{x^3}\,dx$

(b) $\int_1^3 \left(1 + \frac{3}{x^2}\right) dx$

(c) Using integration by parts,

$\int_0^{\infty} xe^{-x}\,dx$

(d) Using the substitution $u = x^2 + 4$,

$\int_0^2 \frac{2x}{x^2 + 4}\,dx.$

8 Matrices

EXERCISE 11

A and B are 2×2 matrices given by

$$A = \begin{pmatrix} 0.5 & 0.1 \\ 0.6 & 0.0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}.$$

Work out each of the following.

(a) $A + B$

(b) AB.

9 Measures of location and dispersion

EXERCISE 12

Numbers of blue tits observed in a random sample of 5 gardens were found to be 8, 2, 0, 5 and 1.

(a) Find the mean and median of this sample.

(b) Calculate the sample standard deviation.

10 Probability distributions

EXERCISE 13

In M347, the notation $N(\mu, \sigma^2)$ is used for the normal distribution with mean μ and variance σ^2. Similarly, $B(n, p)$ is used for the binomial distribution with parameters n (a positive integer representing the number of independent Bernoulli trials) and $p \in (0, 1)$ representing the probability of success in a single trial. Poisson(μ) refers to the Poisson distribution with mean and variance μ.

Classify the likely distributions applicable in the scenarios that follow as binomial, Poisson or normal.

(a) Hourly counts of vehicles passing a census point on a quiet country road.

(b) The number of working days each week in the winter that Chris forgets to take his spare sweater with him.

(c) Heights of children in Leeds on entering secondary school.

EXERCISE 14

Suppose that the distribution of X is $N(1, 4)$.

(a) What type of distribution does $Y = 1 - 2X$ have?

(b) Calculate the mean of the distribution of Y.

(c) Calculate the variance of the distribution of Y.

11 Confidence intervals

EXERCISE 15

A 95% confidence interval for the mean of a population is $(-1.25, 2.46)$. For each of the following statements, decide whether it is true or false.

(a) The probability that the population mean is between -1.25 and 2.46 is 0.95.

(b) In repeated sampling, the interval calculated in the same way will include the population mean nineteen times out of twenty.

(c) In repeated sampling, the interval calculated in the same way will include zero nineteen times out of twenty.

(d) On the basis of these data, it is very implausible that the population mean is zero.

12 Hypothesis tests

EXERCISE 16

A hypothesis test was undertaken at the 5% level to test the null hypothesis that there is no difference between the average sizes of tumours in cancer patients after treatment with standard and experimental drugs. The value of the test statistic was 1.52 and the rejection region of the test was values of the test statistic greater than 3.84. Which of the following two statements is true?

(A) The null hypothesis is not rejected, so there is little or no evidence that the average tumour sizes are not the same.

(B) The null hypothesis is rejected, so there is some evidence that the average tumour sizes are not the same.

EXERCISE 17

Another hypothesis test was undertaken to test the null hypothesis that there is no difference between the proportions of girls and boys who get a grade 'A' in GSCE Mathematics. The significance probability (or p value) was found to be 0.0029. Which of the following two statements is true?

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

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

13 Linear regression

EXERCISE 18

A regression line has been fitted to data on how the time taken by males to complete a marathon (Y, in minutes) depends on the distance run in weekly training (x, in miles). The fitted line has the equation

$$Y = 303 - 1.59x.$$

Interpret what the fitted line is telling us.

14 Are you ready for M347?

You should be familiar with the techniques covered by Exercises 1 to 18 before embarking on M347. Exercises 1 to 11 covered the mathematical competence you will need before starting M347, Exercises 12 to 18 the statistical.

If you found Exercises 1 to 11 particularly difficult, and have not studied MST124 Essential Mathematics 1 and MST125 Essential Mathematics 2 (or their predecessors, MST121 Using Mathematics and MS221 Exploring Mathematics), you would be well advised to consider taking MST124 and MST125 before starting M347. If you had difficulty only with the later parts of Exercises 9 and 10, then this should not delay your registering for M347. However, you would be well advised to reinforce your understanding with the help of the relevant units in MST124 and MST125 (or MST121 and MS221), or of any basic textbook on calculus.

You should also be familiar with the statistical techniques covered by Exercises 12 to 18 before embarking on M347. In particular, you should have a sound grasp of the concepts behind confidence intervals and hypothesis testing. These and the other statistical questions are covered in the Level 2 statistics course M248 *Analysing Data*. In any case, you would be well advised to study M248 before M347, or to ensure that you have covered material of this sort by some other means.

Do contact your Student Support Team via StudentHome if you have any queries about your suitability for M347.

Solutions to Exercises

Exercise 1(a)

$$\sum_{i=0}^{3} i^2 a_i x_i = 0^2 \times a_0x_0 + 1^2 \times a_1x_1 + 2^2 \times a_2x_2 + 3^2 \times a_3x_3$$
$$= a_1x_1 + 4a_2x_2 + 9a_3x_3.$$

Exercise 1(b)

$$\sum_{i=1}^{3} x_i^2 = 1^2 + 2^2 + (-2)^2$$
$$= 1 + 4 + 4$$
$$= 9.$$

Exercise 2

Putting $x = 0$ yields

$$\begin{aligned} f(0) &= (1+0)^2 \times (1-0)^3 \\ &= 1^2 \times 1^3 \\ &= 1 \times 1 \\ &= 1. \end{aligned}$$

Putting $x = \frac{1}{2}$ yields

$$\begin{aligned} f\left(\frac{1}{2}\right) &= \left(1+\frac{1}{2}\right)^2 \times \left(1-\frac{1}{2}\right)^3 \\ &= \left(\frac{3}{2}\right)^2 \times \left(\frac{1}{2}\right)^3 \\ &= \frac{9}{4} \times \frac{1}{8} \\ &= \frac{9}{32}. \end{aligned}$$

A fractional answer like this is fine, while in decimals the answer is 0.28125.

Exercise 3

Substituting in the given formula, we get

$$\bar{x} \pm 1.960 \frac{s}{\sqrt{n}} = 1.521 \pm 1.960 \times \frac{0.614}{\sqrt{12}}$$

$$= 1.521 \pm (1.960 \times 0.177246532)$$

$$= 1.521 \pm 0.347403204.$$

To 2 decimal places, the required interval is therefore (1.17, 1.87).

Exercise 4

Applying the algebraic manipulation rules to

$$\tfrac{1}{2}y + 4 - 2x = y - 1$$

yields

$$4 - 2x = \tfrac{1}{2}y - 1$$

(by subtracting $\frac{1}{2}y$ from both sides) and then

$$5 - 2x = \tfrac{1}{2}y$$

(by adding 1 to both sides),

$$2(5 - 2x) = y$$

(by multiplying both sides by 2) and finally

$$y = 2(5 - 2x)$$

(by swapping sides), or equivalently $y = 10 - 4x$.

Exercise 5(a)

Using the rule $1/x^a = x^{-a}$ gives

$$\frac{1}{x^3} = x^{-3}.$$

Exercise 5(b)

Using the rule $(x^a)(x^b) = x^{a+b}$ gives

$$x^2 \times x^4 = x^6.$$

Exercise 6

The fourth root of 0.4 is $0.4^{1/4}$.

From your calculator, this takes the value 0.7953, correct to 4 decimal places.

Exercise 7(a)

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

$$\begin{aligned}\log 12 &= \log(2 \times 6)\\ &= \log 2 + \log 6\\ &\simeq 0.6931 + 1.7918\\ &= 2.4849\\ &\simeq 2.485, \qquad \text{correct to 3 decimal places.}\end{aligned}$$

Exercise 7(b)

Here, we also need the rule that, for any real number a and any $b > 0$, $\log(b^a) = a \log b$. This yields

$$\log(3\sqrt{x}) = \log 3 + \log(\sqrt{x})$$
$$= \log 3 + \tfrac{1}{2} \log x.$$

Exercise 7(c)

Remembering that the exponential function, exp, is the inverse function of log, we have

$$\begin{aligned} x &= \exp(1.5) \\ &= e^{1.5} \\ &\simeq 4.482, \qquad \text{correct to three decimal places.} \end{aligned}$$

Exercise 8(a)

Since $\exp(-x) = e^{-x} = 1/e^{x}$ for all x, we have

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

Since $\exp(3x) = e^{3x} = (e^{x})^{3}$ for all x, we have

$$\begin{aligned}\exp(3x) &= \{\exp(x)\}^{3}\\ &= 2^{3}\\ &= 8.\end{aligned}$$

Exercise 8(b)

This again uses the rules used in Exercises 7(b) and 7(c):

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

Exercise 8(c)

Again using the inverse relationship between exp and log,

$$\begin{aligned} x &= \log(1.5) \\ &\simeq 0.405, \qquad \text{correct to three decimal places.} \end{aligned}$$

Exercise 9(a)

The derivative of x^n is nx^{n-1}.

So,

$$\frac{d}{dx}(x^4) = 4x^3.$$

Since $\frac{1}{x^2} = x^{-2}$, its derivative is

$$\frac{d}{dx}\left(\frac{1}{x^2}\right) = -2x^{-3} \quad \text{or} \quad -\frac{2}{x^3}.$$

Exercise 9(b)

The *chain rule*, for differentiating the composition of one function with a second one, is needed to differentiate these.

First, $\exp(-3x) = \exp(y)$, where $y = -3x$. Therefore, its derivative is

$$-3e^{-3x}$$

because $d\exp(y)/dy = \exp(y)$ and $d(-3x)/dx = -3$.

Second, $\log(1 - x^2) = \log y$, where $y = 1 - x^2$. Therefore, its derivative is

$$-\frac{2x}{1 - x^2}$$

because $d\log(y)/dy = 1/y$ and $d(1 - x^2)/dx = -2x$.

Exercise 9(c)

This is the *product* of two functions.

The derivative is

$$\frac{d}{dx}(x)\, e^{2x} + x\frac{d}{dx}(e^{2x}) = 1\ e^{2x} + x\, 2e^{2x} = (1 + 2x)e^{2x}.$$

Exercise 9(d)

This is the *quotient* of two functions. The derivative is

$$\frac{\frac{d}{dx}(0.2x)\times(1-0.8x)-0.2x\frac{d}{dx}(1-0.8x)}{(1-0.8x)^2} = \frac{0.2(1-0.8x)-0.2x(-0.8)}{(1-0.8x)^2}$$

$$= \frac{0.2-0.16x+0.16x}{(1-0.8x)^2}$$

$$= \frac{0.2}{(1-0.8x)^2}.$$

Exercise 10(a)

For $n \neq -1$, $\int x^n \, dx = \dfrac{x^{n+1}}{n+1} + c.$

So,

$$\int x^2 \, dx = \frac{x^3}{3} + c.$$

Since $\dfrac{1}{x^3} = x^{-3}$,

$$\int \frac{1}{x^3} \, dx = \int x^{-3} \, dx$$

$$= \frac{x^{-2}}{-2} + c$$

$$= -\frac{1}{2x^2} + c.$$

Exercise 10(b)

$$\int_1^3 \left(1 + \frac{3}{x^2}\right) dx = \left[x + 3\frac{x^{-1}}{-1}\right]_1^3$$

$$= \left[x - \frac{3}{x}\right]_1^3$$

$$= (3 - 1) - (1 - 3)$$

$$= 4.$$

Exercise 10(c)

Here, integration by parts is required. First integrate e^{-x} to get $-e^{-x}$, then

$$\int_0^\infty xe^{-x}\,dx = x(-e^{-x})\Big|_0^\infty - \int_0^\infty \frac{d}{dx}(x)\,(-e^{-x})\,dx$$

$$= 0 - 0 + \int_0^\infty e^{-x}\,dx$$

$$= -e^{-x}\Big|_0^\infty$$

$$= -(0-1)$$

$$= 1.$$

Exercise 10(d)

Using the substitution $u = x^2 + 4$, $du = 2x\,dx$, so that when $x = 0$, $u = 0^2 + 4 = 4$ and when $x = 2$, $u = 2^2 + 4 = 4 + 4 = 8$,

$$\int_0^2 \frac{2x}{x^2+4}\,dx = \int_4^8 \frac{1}{u}\,du$$

$$= [\log u]_4^8$$

$$= \log 8 - \log 4$$

$$= \log(8/4) = \log 2.$$

Exercise 11(a)

$$A + B = \begin{pmatrix} 0.5 & 0.1 \\ 0.6 & 0.0 \end{pmatrix} + \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$$

$$= \begin{pmatrix} 0.5+2 & 0.1+1 \\ 0.6+1 & 0.0+5 \end{pmatrix}$$

$$= \begin{pmatrix} 2.5 & 1.1 \\ 1.6 & 5.0 \end{pmatrix}.$$

Exercise 11(b)

$$AB = \begin{pmatrix} 0.5 & 0.1 \\ 0.6 & 0.0 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$$

$$= \begin{pmatrix} 0.5\times 2 + 0.1\times 1 & 0.5\times 1 + 0.1\times 5 \\ 0.6\times 2 + 0.0\times 1 & 0.6\times 1 + 0.0\times 5 \end{pmatrix}$$

$$= \begin{pmatrix} 1.1 & 1.0 \\ 1.2 & 0.6 \end{pmatrix}.$$

Exercise 12(a)

The mean is

$$\bar{x} = \frac{8+2+0+5+1}{5} = \frac{16}{5} = 3.2.$$

The ordered sample is 0, 1, 2, 5, 8, so the median is the middle value, 2.

Exercise 12(b)

The sample standard deviation is

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{4}\{(8-3.2)^2 + (2-3.2)^2 + (0-3.2)^2 + (5-3.2)^2 + (1-3.2)^2\}}$$

$$= \sqrt{\frac{1}{4}\{4.8^2 + (-1.2)^2 + (-3.2)^2 + 1.8^2 + (-2.2)^2\}}$$

$$= \sqrt{\frac{42.8}{4}}$$

$$= \sqrt{10.7}$$

$$= 3.271085447$$

$$\simeq 3.3.$$

Alternatively,

$$s = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)}$$

$$= \sqrt{\frac{1}{4}\{8^2 + 2^2 + 0^2 + 5^2 + 1^2 - (5 \times 3.2^2)\}}$$

$$= \sqrt{\frac{42.8}{4}}$$

$$= \sqrt{10.7}$$

$$= 3.271085447$$

$$\simeq 3.3.$$

Exercise 13(a)

This could be Poisson. The data are discrete and non-negative (they are counts) and there is no fixed upper limit.

Exercise 13(b)

This could be binomial. We would have $n = 5$ (or perhaps some other fixed number) and p, the probability of Chris's forgetting to take his spare sweater, would be (hopefully) quite small.

Exercise 13(c)

This could be normal. The data are real and continuous, with no fixed limits on their range (although there is a range within which one would sensibly expect heights to lie).

Exercise 14(a)

The distribution of $Y = 1 - 2X$ is normal.

Exercise 14(b)

The mean of the distribution of Y is

$$1 - 2 \times 1 = 1 - 2 = -1.$$

Exercise 14(c)

The variance of the distribution of Y is

$$2^2 \times 4 = 4 \times 4 = 16.$$

Exercise 15(a)

This statement is false. The population mean is a fixed, if unknown, quantity and is either in the interval or not, we don't know which.

Exercise 15(b)

This statement is true. You might think of this as entitling you to make the claim that the population mean is in the interval $(1.25, 2.46)$, with the proviso that 5% of statements made on a similar basis will turn out to be false.

Exercise 15(c)

This statement is false. The valid interpretation of this confidence interval is a statement about the true (but unknown) value of the population mean; there is no reason to assume that the population mean is zero (or the midpoint of the interval or any other value within the interval).

Exercise 15(d)

This statement is false. The 95% confidence interval gives a range of values for the population mean which are plausible at the 95% confidence level. Since this interval contains zero, it is plausible that the population mean is zero.

Exercise 16

The value of the test statistic does not lie in the rejection region of this test. Therefore the null hypothesis is not rejected, and there is little or no evidence that the average tumour sizes are not the same. Thus statement (A) is correct and statement (B) is incorrect.

Exercise 17

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 18

The fitted regression line tells us that on average, the more miles run in training the less time taken to run a marathon. The intercept (the value when $x = 0$) is 303 minutes suggesting (perhaps optimistically) that without training, on average, a male could run a marathon in just over 6 hours. The slope (the coefficient of x) tells us how quickly males improve with training: for every extra mile run per week, the marathon time reduces, on average, by 1.59 minutes. (Why can this relationship not go on forever?)