Sample exam 1

Please note:

- For the actual exam, these questions will be delivered to you in Quiz format via the LMS.
- You will be required to submit handwritten answers to the short answer questions, and you will be required to upload these at the end of the exam.
- Late submissions cannot be accepted.

Because the exam is open book, detailed solutions to the sample exams will not be provided. Answers to selected questions will be provided so that you can check your work.

Question 1. **12 marks**

Consider a simple communication system consisting of a transmitter and a receiver. The transmitter can send a signal consisting of either a 0 or a 1. Because of radio interference, occasionally the transmitter will receive the wrong signal. For example, the transmitter may send a 0, but the transmitter receives a 1.

It is known that, regardless of which signal is sent, the opposite signal will be received 2% of the time. That is, if a 0 is sent, then there is a 2% chance that a 1 will be received instead, and similarly, if a 1 is sent, then there is a 2% chance that a 0 will be received instead. Otherwise, the correct signal is received.

Assume the transmitter sends out signals at random. Let A denote the event that the transmitter sends a 0, and let B denote the event that the receiver receives a 0. After one day of receiving transmissions, 65% of all transmissions received were 0s.

(a) **(2 marks)** Describe, in words, what the probabilities $P(B \mid A)$ and $P(B^c \mid A^c)$ represent.

(b) **(1 mark)** From the information above, state the probabilities $P(B)$, $P(B \mid A)$ and $P(B^c \mid A^c)$.

(c) **(2 marks)** Use the Law of Total Probability to express $P(B)$ in terms of $P(A)$.

(d) **(2 marks)** Use your answer to (c) to show that $P(A) = 0.656$.

(e) **(3 marks)** Determine $P(A^c \mid B)$ and $P(A^c \mid B^c)$ using Bayes' theorem.

Use the value of $P(A)$ stated in part (d) to answer this question.

(f) **(2 marks)** Suppose that the transmitter sends out fifty 1 signals in a row. Let X denote the number of correct signals that are received by the receiver. What probability distribution does X follow? Give your answer in the form $X \sim$ ____.

Show all working. Give your answers to at least 3 decimal places.

Question 2. **13 marks**

(a) Let $X \sim \text{NB}(5, 0.35)$. Calculate each of the following.

(i) **(2 marks)** $P(X = 3)$

(ii) **(3 marks)** $P(X \geqslant 2)$

(iii) **(3 marks)** $P(X = 3 \mid X \geqslant 2)$

(iv) **(1 mark)** $E(X)$

(v) **(1 mark)** $\text{Var}(X)$

(b) **(3 marks)** In your own words, give a specific example of a scenario where a geometric distribution could be applied. Your answer should explain what is being counted by the geometric distribution, and why the geometric distribution is suitable.

You do not need to write a lot: two or three sentences addressing the above will be sufficient for full marks.

Show all working. Give your final answers to at least 3 decimal places. You ***must*** *show details for the calculation of any relevant binomial coefficients.*

Question 3. **12 marks**

Consider the function f given by

$$f(x) = \begin{cases} \frac{3-x}{2}, & \text{if } 0 \leqslant x \leqslant 4, \\ \frac{x}{8} - 1, & \text{if } 4 < x \leqslant 8, \\ 0, & \text{otherwise.} \end{cases}$$

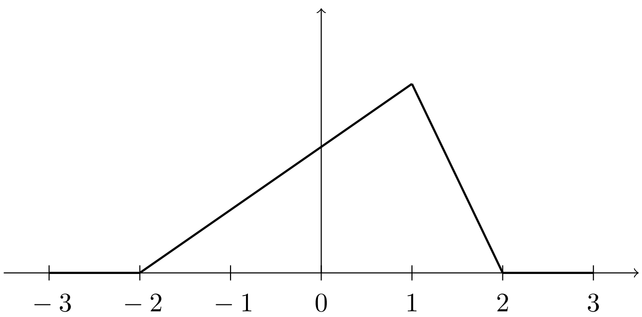
(a) **(2 marks)** Write code to implement this function in R. Take care to use correct syntax and symbols that would be used when typing code.

(b) **(4 marks)** Decide whether or not f is a probability density function. Justify your answer.

Now let g be the function given by

$$f_X(x) = \begin{cases} \frac{x+2}{6}, & \text{if } -2 \leqslant x < 1, \\ \frac{2-x}{2}, & \text{if } 1 \leqslant x \leqslant 2, \\ 0, & \text{otherwise.} \end{cases}$$

Its graph is shown below on the interval $[-3, 3]$. You must use the formula to determine the vertical values.



You are given that g is the probability density function for a random variable X.

(c) **(3 marks)** Determine $P(X \leqslant 0)$.

(d) **(3 marks)** Determine a formula for $P(X \leqslant x)$ for values of x such that $-2 \leqslant x < 1$.

Show all working. Give numeric answers ***exactly, as fractions****, in all cases.*

Question 4. **5 marks**

Consider the following function

$$f(x) = \begin{cases} \frac{1}{3}e^{\frac{x}{3}}, & \text{if } x \leqslant 0, \\ 0, & \text{otherwise} \end{cases}$$

The code below implements the function as `f`, as well as some other functions for later use.

```
f <- function(x) { ifelse(x <= 0, exp(x/3)/3, 0) }
helper <- function(x) { x*f(x) }
mystery.1 <- function() { integrate(helper, 3, Inf)$value }
mystery.2 <- function() { integrate(helper, -Inf, 0)$value }
mystery.3 <- function() { integrate(helper, 0, Inf)$value }
mystery.4 <- function() { integrate(helper, -Inf, 3)$value }
```

Suppose that X is a random variable whose probability density function is f.

(a) **(2 marks)** Which of the four mystery functions correctly calculates $E(X)$?

(b) **(3 marks)** Using R's `integrate` function and one of the mystery functions, write code that outputs the value of $\text{Var}(X)$.

Question 5. **12 marks**

Cars arrive at a drive through restaurant according to a Poisson process, with a rate of 19 cars per hour. From the moment an individual begins placing their order, it takes an average time of 2.5 minutes before they pick up and pay for their order from the next window.

Assume that this time taken follows an exponential distribution. There is just one service window; hence we can model this scenario as an M/M/1 queue, as depicted below.

Assume that the queue has been running for long enough that its steady state properties apply.

(a) **(1 mark)** Write down the service rate μ and the arrival rate λ for this queue (including units).

(b) **(2 marks)** What is the long term average number of customers in the system?

(c) **(2 marks)** On average, after arriving at the queue, how long will a customer need to wait before they reach the drive through speaker?

(d) **(2 marks)** On average, after arriving at the queue, how long will a customer need to wait before they leave the restaurant?

(e) Suppose that, once the drive through contains 3 cars or more, the queue will extend beyond the restaurant and start impeding traffic. The council has received several complaints from locals, and has informed the restaurant that fines will be imposed if traffic is impeded for more than 30% of the time the drive through is in operation.

Let N denote the number of cars in the drive through at any one time.

  (a) **(2 marks)** Determine the probability that there are at least 3 cars in the system.

  (b) **(3 marks)** Determine the value of ρ required so that, on average, they will not be fined.

Show all working. Give your answers to at least 3 decimal places.

Question 6. **8 marks**

A software developer has designed two different user interfaces for a program. They want to choose the interface which is easiest to use, and the way that the developer has decided to measure this is by the amount of time it takes a user to find an important button.

The first option, interface A, was tested on $n_1 = 31$ randomly chosen individuals. The average time taken to find the button in that case was 4.307 seconds, with a sample standard deviation of 0.368. The second option, interface B, was tested on $n_2 = 32$ randomly chosen individuals. The average time taken to find the button in that case was 6.446 seconds, with a sample standard deviation of 0.570.

Let μ_1 and μ_2 denote the true mean time to find the button for interface A and interface B, respectively. For this question, the degrees of freedom is $\text{df} = 53$.

Some R console input and output is shown below. You will need to use some of these numbers in some parts of this question.

```
> qt(0.90, df = 53)
[1] 1.29773
> pt(0.90, df = 53)
[1] 0.8139043
> qt(0.95, df = 53)
[1] 1.674116
> pt(0.95, df = 53)
[1] 0.8267878
> pt(0.975, df = 53)
[1] 0.8330052
> qt(0.975, df = 53)
[1] 2.005746
```

(a) **(1 mark)** What is the estimated mean difference in the time taken to find the button?

(b) **(2 marks)** Calculate the standard error for this estimated mean difference.

(c) **(2 marks)** Calculate a 95% confidence interval for the difference in the two means.

(d) **(3 marks)** Write an appropriate conclusion on the basis of your confidence interval.

Show all working. Give your answers to at least 3 decimal places.

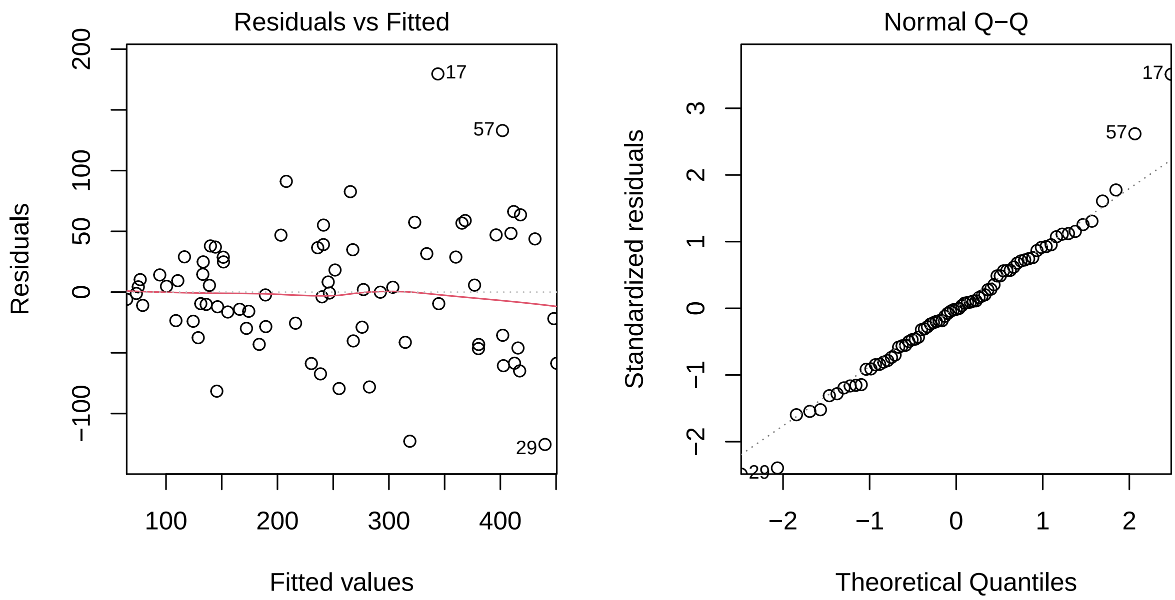
Question 7. **18 marks**

Suppose you were applying linear regression to model the amount of toxic substance in a poisonous mushroom, on the basis of its physical attributes. For this question, the explanatory variable is the diameter of the cap of the mushroom (denoted by `diameter`), which is measured in millimetres (mm). The response variable is the amount of toxic substance in the mushroom, measured in micrograms (µg).

Simple linear regression has been carried out in R using the data from 77 poisonous mushrooms. The output is shown below, along with the residuals versus fits plot and a Q-Q plot of the residuals.

```
Coefficients:
   Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.68511   14.03993   0.262    0.794
diameter    4.9864     0.2574   19.372   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 51.73 on 75 degrees of freedom
Multiple R-squared:  0.8334,    Adjusted R-squared:  0.8312
F-statistic: 375.3 on 1 and 75 DF,  p-value: < 2.2e-16
```



(a) **(4 marks)** Are there any clear violations in the residuals versus fits plot or the Q-Q plot to be concerned with? Justify your answer clearly with references to both plots.

NOTE: Regardless of your answer to (a), for the remainder of this question, assume that there are no linear regression model violations.

(b) **(2 marks)** Does the R output suggest that the regression model fits the data well? Explain.

(c) **(3 marks)** What is the estimate of the explanatory variable coefficient? Give a sentence interpreting this estimated coefficient, including any relevant units.

(d) **(2 marks)** Let β_1 denote the true coefficient for the explanatory variable and consider the hypotheses

$$H_0\colon \beta_1 = 0 \ \text{ versus } \ H_1\colon \beta_1 \neq 0.$$

Do you reject the null hypothesis at the $\alpha = 0.05$ significance level? Explain.

(e) **(2 marks)** Using the fact that $t_{75,0.975} = 1.992$, construct a 95% confidence interval for β_1.

(f) **(2 marks)** In the context of the problem, interpret the confidence interval you calculated in part (e).

(g) **(3 marks)** Suppose the following code was run, with associated output shown. The variable `model` contains the linear model created for this question.

```
> df <- data.frame(diameter=50)
> predict(model, df, interval="prediction")
   fit      lwr      upr
1 502.327 395.4237 609.2304
```

Provide a simple, in-context statement that interprets this output.

Show all working. Give your answers to at least 3 decimal places.

Question 8 **4 marks**

From the following, which would be suitably described by a discrete random variable?

(a) The time you spend talking on your next phonecall.

(b) The number of weddings in the next six months.

(c) The genre of a randomly chosen movie.

(d) The average temperature on Earth on a randomly chosen day.

(e) The number of different places visisted in one day by a random person.

(f) The amount of distance travelled in one day by a random person.

There is at least one correct answer and there may be more than one. Select the correct answers.

For Question 8, you will be marked on the following basis:

- *If you select every correct answer and no incorrect answers, then you will get 4 marks.*
- *If you select no incorrect answers, but not every correct answer, you will get 2 marks.*
- *If you select at least one incorrect answer, and select more correct answers than incorrect answers, you will get 1 mark.*
- *Otherwise, you will get 0 marks.*

Question 9 **4 marks**

Suppose the following hypotheses were to be tested, for unknown means μ_1 and μ_2:

$$H_0\colon \mu_1 = \mu_2 \text{ versus } H_1\colon \mu_1 \neq \mu_2$$

To test the hypotheses, an unpaired two-sample t-test was performed in R, and the following output was obtained:

```
data:  x and y
t = 1.3949, df = 138.54, p-value = 0.1653
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.6929317  4.0128124
sample estimates:
mean of x mean of y
 40.82541  39.16547
```

Assume a 5% level of significance. Which of the following statements are valid conclusions?

(a) Because $p > 0.05$, we know that the means are not equal.

(b) Because $p > 0.05$, we reject the null hypothesis.

(c) Because $p > 0.05$, we accept the null hypothesis.

(d) The evidence suggests that there is a difference in the two means.

(e) The evidence does not suggest that there is a difference in the two means.

(f) To two decimal places, the 95% confidence interval is $(-0.69, 4.01)$.

There is at least one correct answer and there may be more than one. Select the correct answers.

For Question 9, you will be marked on the following basis:

- *If you select every correct answer and no incorrect answers, then you will get 4 marks.*
- *If you select no incorrect answers, but not every correct answer, you will get 2 marks.*
- *If you select at least one incorrect answer, and select more correct answers than incorrect answers, you will get 1 mark.*
- *Otherwise, you will get 0 marks.*

Question 10 **3 marks**

Let $Y \sim \text{Exp}(\frac{1}{4})$. To two decimal places, $P(Y \geqslant 3)$ is:

(a) 0.118

(b) 0.393

(c) 0.472

(d) 0.528

(e) 0.607

(f) 1.00

There is one correct answer. Select the correct answer.

Question 11 **3 marks**

Which of the following R calls would output the value $P(X \leqslant 1)$, where $X \sim \text{Gamma}(3, 2)$?

(a) `dgamma(0, shape = 3, scale = 2) + dgamma(1, shape = 3, scale = 2)`

(b) `dgamma(1, shape = 3, scale = 2)`

(c) `pgamma(0, shape = 3, scale = 2)`

(d) `pgamma(1, shape = 3, scale = 2)`

(e) `1 - dgamma(1, shape = 3, scale = 2)`

(f) `1 - pgamma(1, shape = 3, scale = 2)`

There is one correct answer. Select the correct answer.

Question 12 **3 marks**

A probability mass function for a random variable Z is given to you below in table form.

z	-2	-1	3	5	6
$P(Z = z)$	0.05	0.012	0.512	0.152	0.274

The value of $P(Z \geqslant 3)$ is:

(a) 0.2

(b) 0.4

(c) 0.426

(d) 0.512

(e) 0.6

(f) 0.938

There is one correct answer. Select the correct answer.

Question 13 **3 marks**

Let $X \sim \text{Bin}(18, 0.39)$. The exact value of $P(X \geqslant 1)$ is:

(a) $1 - 0.61^{18}$

(b) $\binom{18}{0} \times 0.39^0 \times (1 - 0.39)^{18}$

(c) $1 - 0.39^{18}$

(d) $\binom{18}{1} \times 0.39 \times 0.61^{17}$

(e) $1 - \binom{18}{1} \times 0.39 \times 0.61^{17}$

(f) $\binom{18}{0} \times 0.61^0 \times (1 - 0.61)^{18}$

There is one correct answer. Select the correct answer.