

Statistics (ABC260S) Test 1 Memo 2025

■ represents one mark; ▲ represents half a mark

Questions

1. [10 marks] Answer the following multiple choice questions (two marks each).

- a. [2 marks] What is the capital city of Morocco?
 - i. Rabat **Correct** ■ ■
 - ii. Tunis
 - iii. Casablanca
 - iv. Los Angeles
- b. [2 marks] What is the deepest lake in the world?
 - i. Tanganyika
 - ii. Superior
 - iii. Baikal **Correct** ■ ■
 - iv. Argentino
- c. [2 marks] What is $3 + 8$?
 - i. 5
 - ii. 11 **Correct** ■ ■
 - iii. 24
 - iv. 38
- d. [2 marks] Which of the below countries has the largest population?
 - i. Mexico
 - ii. Nigeria
 - iii. Brazil
 - iv. Indonesia **Correct** ■ ■
- e. [2 marks] The most viewed YouTube video of all time is...
 - i. Despacito by Luis Fonsi
 - ii. Gangnam Style by Psy
 - iii. Baby Shark Dance by Pinkfong **Correct** ■ ■
 - iv. See You Again by Wiz Khalifa

2. [11 marks] Below are given the final marks of a class of twelve students.

47 40 58 45 70 55 54 47 35 47 31 94

- a. [3 marks] Find the mean of the final marks.

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \blacksquare \\ &= \frac{1}{12} [47 + 40 + 58 + 45 + 70 + 55 + 54 + 47 + 35 + 47 + 31 + 94] \blacksquare \\ &= \frac{623}{12} = 51.9167 \blacksquare\end{aligned}$$

- b. [3 marks] Find the median of the final marks.

First we put the data in increasing order: 31 35 40 45 47 47 47 54 55 58 70 94

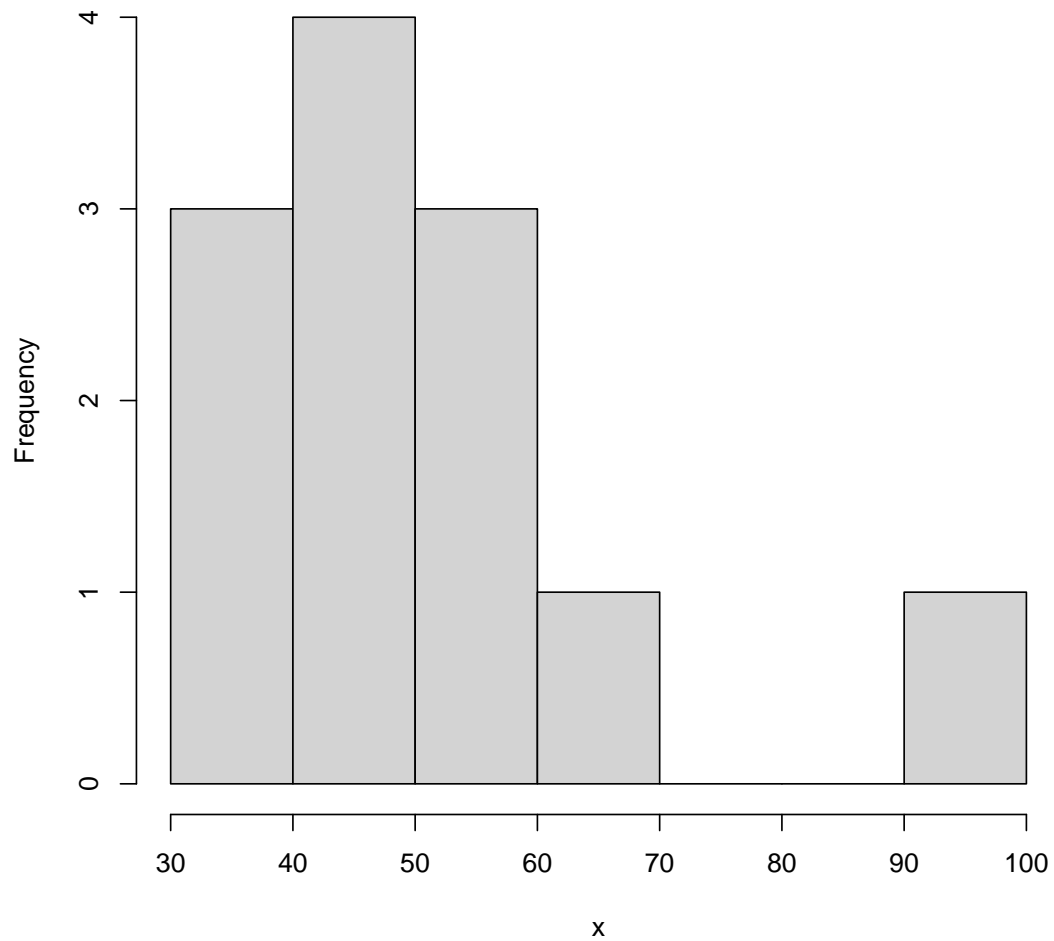
■ Then we take the average of the 6th and 7th values ■. Hence, the median is $\frac{47 + 47}{2} = 47$ ■

- c. [3 marks] Find the standard deviation of the final marks.

$$\begin{aligned}s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \blacksquare \\ &= \frac{1}{11} \left[\left(47 - \frac{623}{12}\right)^2 + \left(40 - \frac{623}{12}\right)^2 + \cdots + \left(94 - \frac{623}{12}\right)^2 \right] \blacksquare \\ &= 16.8817 \\ s &= \sqrt{16.8817} = 4.1087 \blacksquare\end{aligned}$$

- d. [2 marks] Draw a histogram of the final marks. Use a bin width of 10.

Histogram of Final Marks



■ for correctly labelled axes; ■ for correct drawn histogram

3. [9 marks] Consider the random variable X with probability density function as follows:

$$f_X(x) = \begin{cases} k(x^2 + x) & \text{for } 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

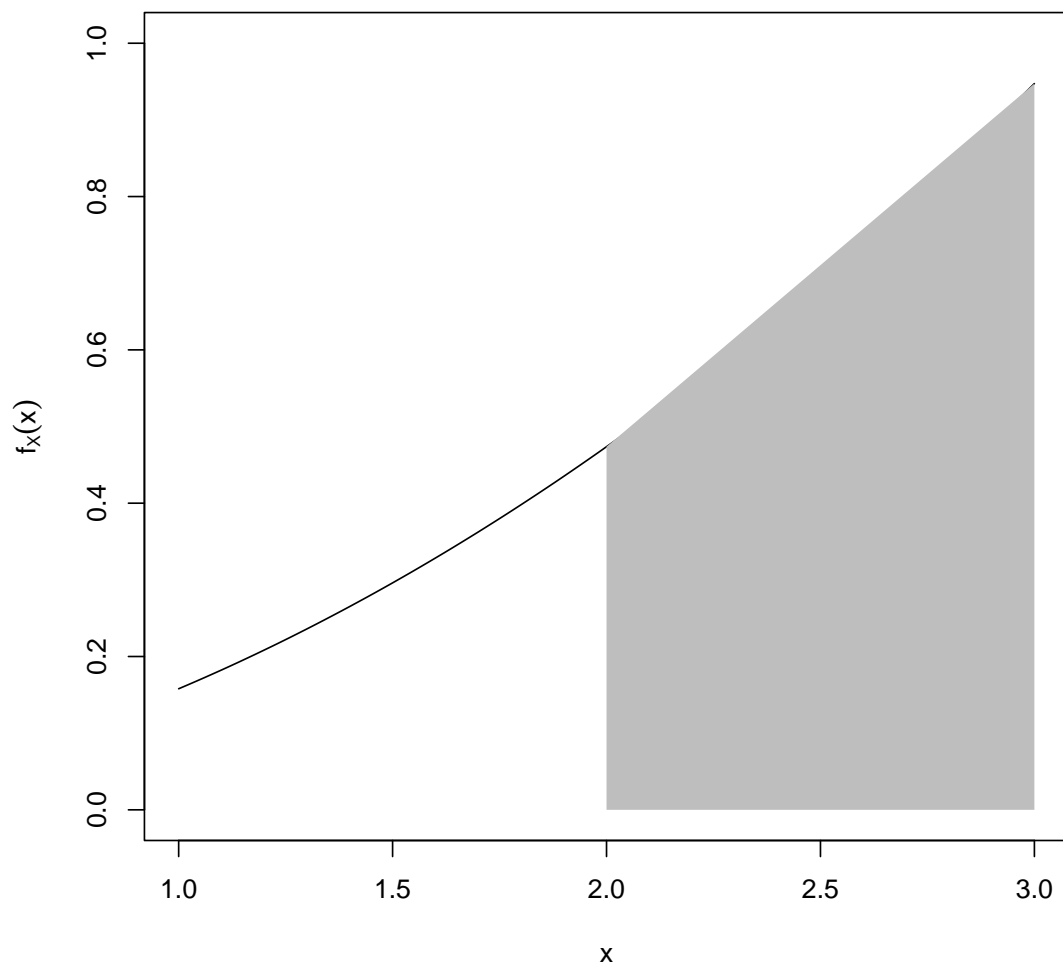
a. [3 marks] Find the value of k .

$$\begin{aligned}
 k \int_{-\infty}^{\infty} f_X(x) dx &= 1 \blacksquare \\
 k \int_1^3 x + x^2 dx &= 1 \\
 k \left[\frac{1}{2}x^2 + \frac{1}{3}x^3 \right]_1^3 &= 1 \blacksquare \\
 k \left[\frac{9}{2} + 9 - \left(\frac{1}{2} + \frac{1}{3} \right) \right] &= 1 \\
 k &= \frac{3}{38} = 0.07895 \blacksquare
 \end{aligned}$$

b. [3 marks] Find $\Pr(X > 2)$.

$$\begin{aligned}
 \Pr(X > 2) &= \int_2^3 f_X(x) dx \blacksquare \\
 &= k \left[\frac{1}{2}x^2 + \frac{1}{3}x^3 \right]_2^3 \blacksquare \\
 &= k \left[\frac{9}{2} + 9 - \left(2 + \frac{8}{3} \right) \right] \\
 &= \frac{53}{6}k = \frac{53}{76} = 0.6974 \blacksquare
 \end{aligned}$$

c. [3 marks] Show the probability just calculated on a graph of $f_X(x)$.



■ for correctly labeled axes, ■ for correctly plotted $f_X(x)$, ■ for correct region shaded

4. [7 marks] Consider the dataset provided in the Excel file `surf4cars_Toyota_Corolla.xlsx`. A description of the columns in the dataset is as follows. Answer the following questions using R code and output only.
- a. [4 marks] Fit a linear regression model to the data where Asking Price is the response variable and Year Model, Mileage, and Transmission Type are explanatory variables. Display a summary of the model output.

```
toyota <- read_excel("surf4cars_Toyota_Corolla.xlsx")
mylm <- lm(AskingPrice ~ YearModel + Mileage + TransmissionType,
           data = toyota)
summary(mylm)

##
## Call:
## lm(formula = AskingPrice ~ YearModel + Mileage + TransmissionType,
##     data = toyota)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -82483 -11141  -1945   4069 144598
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.841e+07  2.641e+06  -6.971 5.06e-11
## YearModel       9.252e+03  1.305e+03   7.091 2.56e-11
## Mileage        -3.748e-01  1.011e-01  -3.707 0.000275
## TransmissionTypeManual -1.711e+04  4.551e+03  -3.759 0.000227
##
## (Intercept)      ***
## YearModel         ***
## Mileage           ***
## TransmissionTypeManual ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30490 on 190 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.6562, Adjusted R-squared:  0.6508
## F-statistic: 120.9 on 3 and 190 DF,  p-value: < 2.2e-16
```

■ for importing data; ■ for calling `lm`; ■ for formula and data arguments; ■ for summary

- b. [3 marks] Predict the asking price of a Toyota Corolla with 2018 year model, 100 000 km driven, and a manual transmission. Include 95% prediction limits.

```
newdata <- data.frame(YearModel = 2018,
                      Mileage = 100000,
                      TransmissionType = "Manual")
predict(mylm, newdata = newdata,
        interval = "prediction")

##      fit      lwr      upr
## 1 202504.5 141723 263286.1
```

■ for creating new data frame; ■ for calling `predict`; ■ for `interval = "prediction"`