

# Final BTM 310 Marking Scheme

John Karuitha

7/11/2021

Table 1: Age Groups of Customers of ABC Hotel

age_group	number
18 and less	45
18-25	83
25-35	154
35-50	18
Over 50	170

Table 2: Age Groups of Customers of ABC Hotel

age_group	prop
18 and less	0.0957447
18-25	0.1765957
25-35	0.3276596
35-50	0.0382979
Over 50	0.3617021

## QUESTION ONE

### PART A

As part of the marketing group at ABC Hotels LTD, you are asked to find out the age distribution of ABC's clients. With the help of 10 of your colleagues, you conduct exit interviews by randomly selecting people to question at 20 different days. You ask them to tell you if they are younger than 18 years old, 18 to 25 years old, 25 to 35 years old, 35 to 50 years old, or older than 50.

From 470 responses, you find out that 45 are younger than 18, 83 are 18 to 25 years old, 154 are 25 to 35, 18 are 35 to 50, and 170 are older than 50. For the age distribution:

- a) Make a frequency table with two columns; Age group and Number. (2 marks)

*The frequency table should be as in Table 1 above (2 marks)*

- b) Make a relative frequency table. (2 Marks)

*The relative frequency should be as in Table 2 above; can either be fractions, decimals or percentages. Accept any of these formats for representing proportions.*

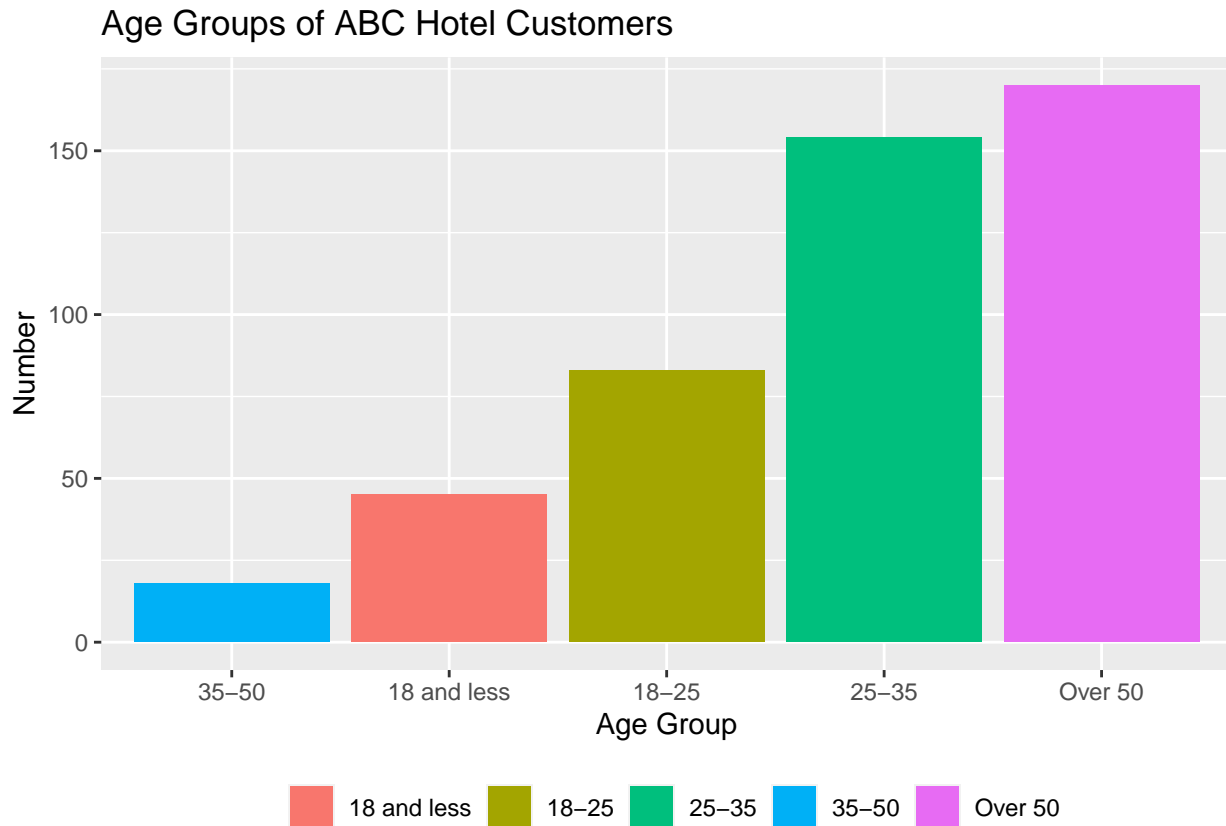
- c) Is Age group Level a quantitative or qualitative variable? Explain. (1 mark)

*Age group is qualitative or categorical as it puts the people into bins or categories and has no units.*

- d) Is Number a quantitative or qualitative variable? Explain. (1 marks)

*Number is quantitative and discrete as it has units but cannot take on fractions: NB: The student need NOT say it is discrete, quantitative is enough for full marks.*

- e) Draw a bar-graph to represent the above information. (3 marks).



NB: The x-axis should represent age group with number on the y-axis.

*Marks distribution: Getting the chart x and y axis right 1/2 mks Getting the bars right 1 mk. NB: The bars need NOT be ordered in ascending or descending order. The graph need NOT be colored. Titles and axis labels = 1/2 marks*

- f) What is the probability that a client of ABC hotel picked at random would be a over 50 years old? (2 mark)

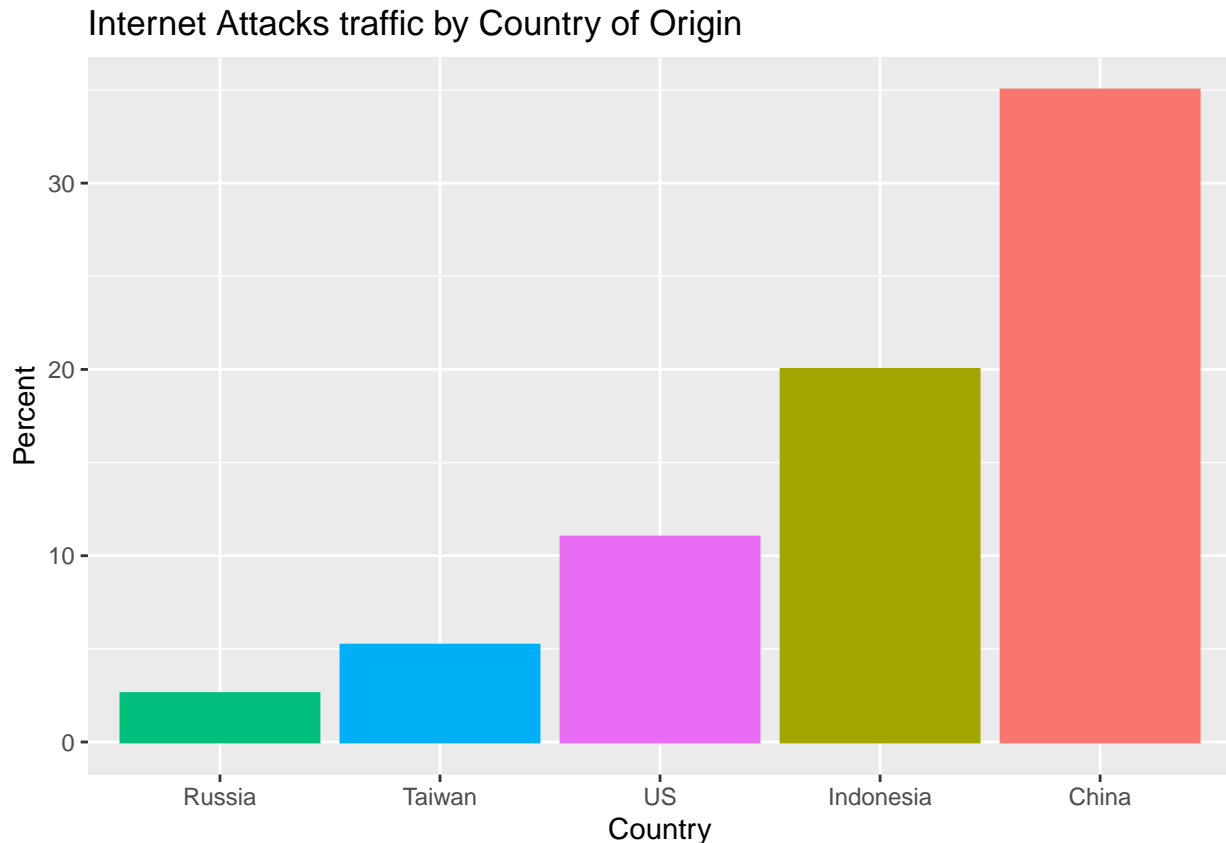
$\frac{170}{470} = \frac{5}{90} = 0.361$ . Fraction also acceptable.

## PART B

According to Akamai's 2013 State of the Internet report ([www.akamai.com/stateoftheinternet/](http://www.akamai.com/stateoftheinternet/)), 35% of observed attack traffic originated from China, 20% from Indonesia, 11% from the US, 5.2% from Taiwan and 2.6% from Russia.

- a) Make a bar chart to display the results and label it clearly. (5 marks)

*Marks distribution: Getting the chart x and y axis right 1 mks Getting the bars right 3 mk. NB: The bars need NOT be ordered in ascending or descending order. The graph need NOT be colored. Titles and axis labels = 1 marks*



b) Would a pie chart be an effective way of communicating this information? Why or why not? (4 marks)

*NO. The pie chart uses angles to represent data. The human mind is not good at distinguishing angles. The bar graph uses height which the mind easily discerns. The pie chart is especially messy when you have many categories with close values.*

## PART C

You are provided with the following observations regarding the income in a hotel for a week, beginning from a Monday to the following Sunday.

100, 80, 70, 50, 175, 184, 49

g) Write a sample R code that you would use to this data. (3 marks)

```
weekly_income <- c(100, 80, 70, 50, 175, 184, 49)
```

*NB: The variable name may differ and not just weekly\_income. The following is also right*

```
c(100, 80, 70, 50, 175, 184, 49)
```

h) Write a sample R code that you would use to compute the mean of the data. (2 mark)

```
mean(weekly_income) ##or whichever name they used for the variable, or
```

```
mean(c(100, 80, 70, 50, 175, 184, 49))
```

## QUESTION TWO

Suppose you take the average score in the statistics exam for a group of students. You find that the data is normally distributed with the average score at 60 marks with a standard deviation of 10 marks.

- a) Compute the z-score for a score of 70 marks. (3 marks)

$$z_x = \frac{x - \bar{x}}{\sigma_x} \text{ (1 mark)}$$

$$z = (70 - 60) / 10 = 1 \text{ (2 marks)}$$

- b) Approximately what fraction of students will score less than 40 marks? (6 marks)

$$z = (40 - 60) / 10 = -2$$

Scoring less than 40 marks has same area as scoring over 40 marks.

$$0.5 - 0.4772 = 0.0228$$

- c) Approximately what fraction of students will score between 50 and 60 marks? (6 marks)

$$(50-60) / 10 = -1$$

$$(60-60) / 10 = 0$$

Area from 0 to -1 is same as area from 0 to 1 = 0.3413

## QUESTION THREE

A recent study of Kenya Revenue Authority (KRA) audits showed that, for estates worth less than 15 million, about 1 out of 7 of all estate tax returns are audited, but that probability increases to 0.5 for estates worth over Ksh. 15 million. Suppose a tax accountant has three clients who have recently filed returns for estates worth more than Ksh. 15 million. What are the probabilities that:

- a) All three will be audited? (3 marks)

We use the multiplication rule assuming independence.

$$0.5 * 0.5 * 0.5 = 0.125$$

- b) None will be audited? (3 marks)

We use the complement rule

$$1 - 0.125 = 0.875$$

- c) At least one will be audited? (5 marks)

Here, we examine 3 probabilities

$$P(1 \text{ is audited}) + p(2 \text{ are audited}) + p(3 \text{ are audited}) \text{ (1 Mark)}$$

$$0.5 + (0.5 * 0.5) + (0.5 * 0.5 * 0.5) = 0.875 \text{ (4 Marks)}$$

- d) What did you assume in calculating these probabilities? (4 marks)

- Independence

## QUESTION FOUR

A regression of Total Revenue on Ticket Sales by the concert production company of Exercises 2 and 4 finds the model

$$\text{Revenue} = -14,228 + 36.87\text{TicketSales}$$

- a) What is the explanatory or independent variable? (2 mark)

*Ticket sales*

- b) What is the response or dependent variable? (2 mark)

*revenue*

- c) What does the slope mean in this context? (2 marks)

*A unit increase in sales corresponds to a 36.87 units increase in revenue*

- d) Management is considering adding a stadium-style venue that would seat 10,000. What does this model predict that revenue would be if the new venue were to sell out? (6 marks)

$$\text{revenue} = -14,228 + 36.87(10,000) = 354,472$$

- e) Why would it be a poor business decision to assume that this model accurately predicts revenue for the situation in part (d)? (3 marks)

*There is no guarantee the stadium would be full.*

## QUESTION FIVE

- a) Distinguish between a sample and population. (3 marks)

*Sample: Subset of a population. Population: Large collection of individuals or objects that are the main focus of scientific enquiry.*

*NB: language may vary:*

- b) Compare and contrast the arithmetic mean and the geometric mean, giving situations where each is a more appropriate measure of center. (4 marks).

*-The computation is different. Mean is the sum of the quantities of interest divided by the count of these subjects. (A formula definition is also fine).*

*- The geometric mean is the  $n$ th root of the product of the quantities of interest, where  $n$  is the count of these quantities.*

$$\text{Arithmetic}_{\text{mean}} = \frac{\sum x}{N}$$

$$\text{geometric}_{\text{mean}} = \sqrt[n]{X_1 * X_2 * \dots * X_n}$$

- c) List and discuss four sampling techniques that researchers use to select a sample from the population (8 marks). (1 mark for listing and one for explaining)

- random sampling
- Convenience
- Snowballing
- Stratification
- Clustering
- etc