

AFE135 – Computing Assignment

Ryan Fitzpatrick - s5192584

Question 1

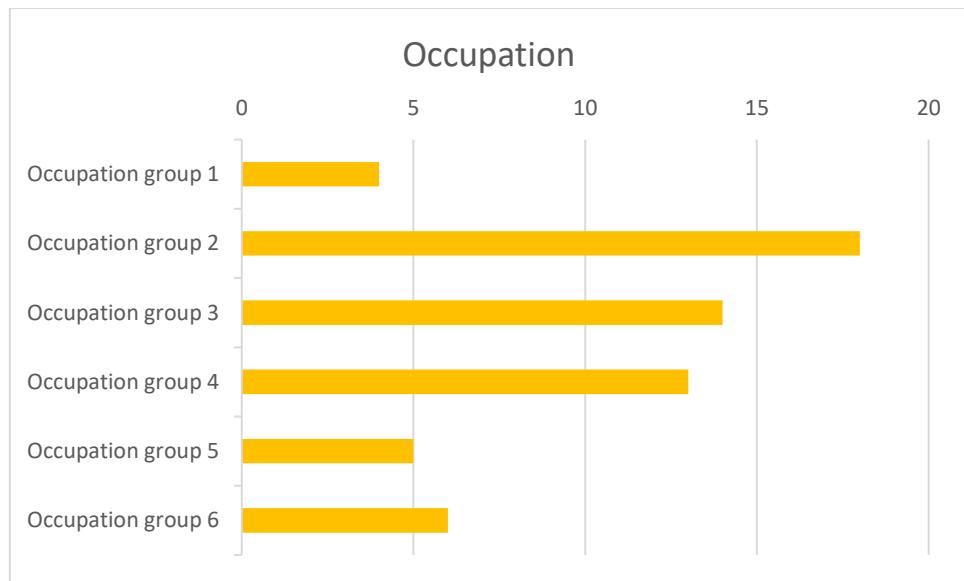
- a) The researcher would have to record specific information about a large sample of people. A questionnaire would be the most appropriate survey method for the researcher as it is easier, and cheaper than phone/personal interviews, while extrapolating the same data.
- b) Stratified Random Sampling – separating the population into non overlapping groups i.e. (Australian men and women), then select a random sample from each stratum.
- c) The researcher should include hours worked per week and the income of men/women as the two main variables for this analysis. This is because the data is quantitative and continuous, which can be used the smaller intervals such as income/time).
- d) The researcher may encounter a lower response rate from the self-administered questionnaire as well as potential misrepresentation of data due to the subject matter of income.

Question 2

Which graphical technique or chart should be used if the researcher is interested in comparing the number of individuals in each occupation category?

- a) The researcher's data is nominal because the 'values' of the variable, the occupations of the individuals, are the six categories.
A bar chart should be used by drawing a rectangle representing each category. This is because the height of the bar most accurately represents the frequency of the individuals in the occupation category.

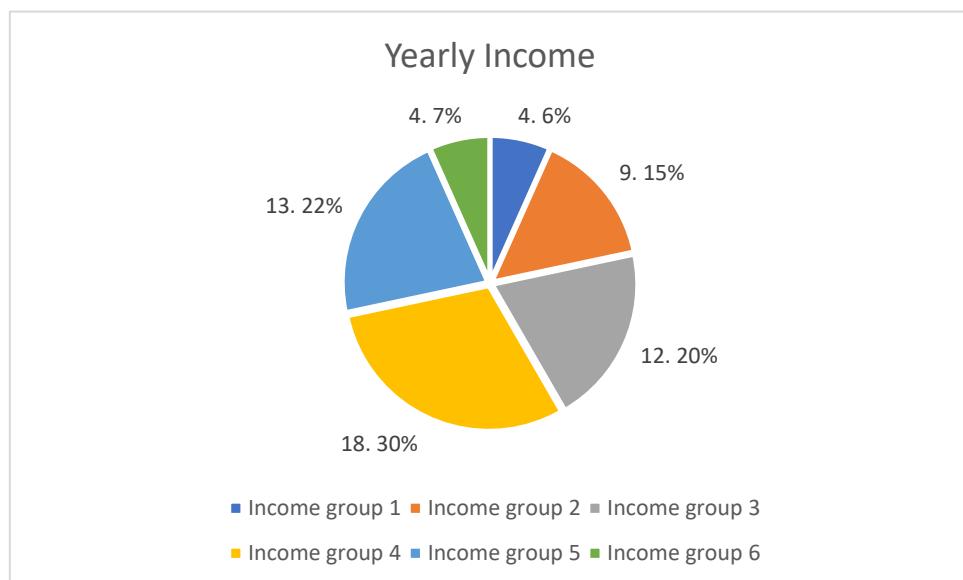
Occupation	
Occupation category	Frequency
Occupation group 1	4
Occupation group 2	18
Occupation group 3	14
Occupation group 4	13
Occupation group 5	5
Occupation group 6	6



As seen in the above bar graph, the most individuals belong to occupation group 2.

b) A pie chart should be used as it exhibits the proportion or percentage of individuals of each yearly income group. This is because the data is categorical size of each slice of a circle is proportional to the percentage corresponding to the category.

Yearly income	
Income Category	Frequency
Income group 1	4
Income group 2	9
Income group 3	12
Income group 4	18
Income group 5	13
Income group 6	4



As shown in the above pie chart, the largest proportion of individuals belonged to income group 4.

Question 3

a) By applying the below Sturges formula, it can be determined that the appropriate number of class intervals is 7.

$$K = 1 + 3.3 \log 10 n$$

$$n = 60 \text{ and } K = 1 + 3.3 \log 60 = 1 + 3.3(1.77) = 6.86$$

Being that K is closer to 7, the researcher decides to use 7 classes.

b) The class width is calculated by taking the difference between the largest and smallest observations and dividing it by the number of classes.

Thus: Approximate class width = Largest value – Smallest value / Number of classes

$$\text{Hours per week: } 42-10 = 32/7 = 4.57$$

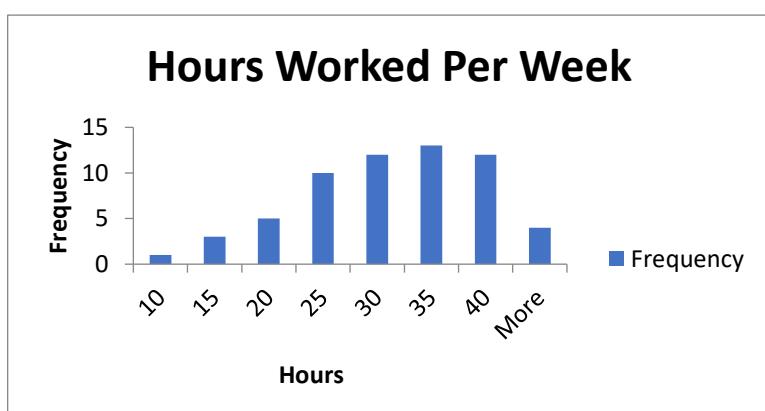
As the approximate value of the number of classes is closer to 5, the researcher decides to use 5 as the most appropriate class width.

$$\text{Yearly Income ('000's): } 79.2 - 48.1/7 = 4.42$$

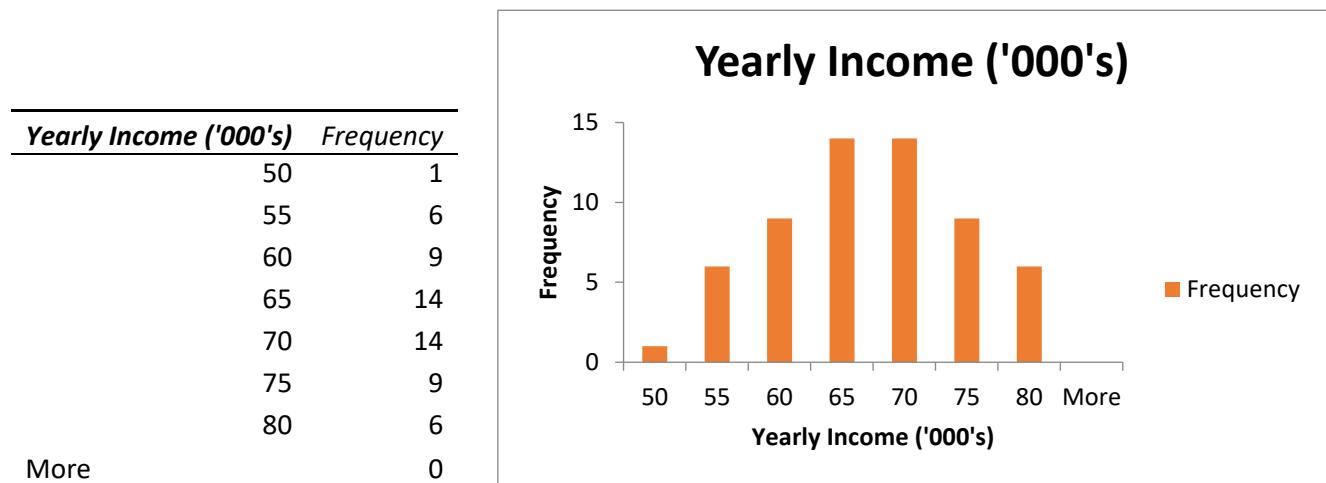
For convenience, the number is rounded to 5 – an acceptable action because there is no fixed rule about the number of class intervals, which is how determines the class width.

c) Histogram for hours worked per week:

Hours Per Week	Frequency
10	1
15	3
20	5
25	10
30	12
35	13
40	12
More	4



The histogram is negatively skewed as it has a long tail to the left.

Histogram for Yearly Income ('000's):

This histogram is symmetrical.

Question 4:

a)	Numerical Summary Report (Hours Worked Per Week)	Numerical Summary Report (Yearly Income '000's)
Mean:	29.19167	Mean: 64.8
Median:	28	Median: 64.8
Mode:	34.5	Mode: 59
Variance:	61.9584	Variance: 57.45206
Standard Deviation:	7.871366	Standard Deviation: 7.579714
Smallest value:	10	Smallest value: 48.6
Largest value:	42	Largest value: 79.2
Quartile 1:	24.5	Quartile 1: 59.4
Quartile 2:	28	Quartile 2: 64.8
Quartile 3:	36	Quartile 3: 71.82

b) The coefficient of correlation informs us that there is a strong positive relationship.

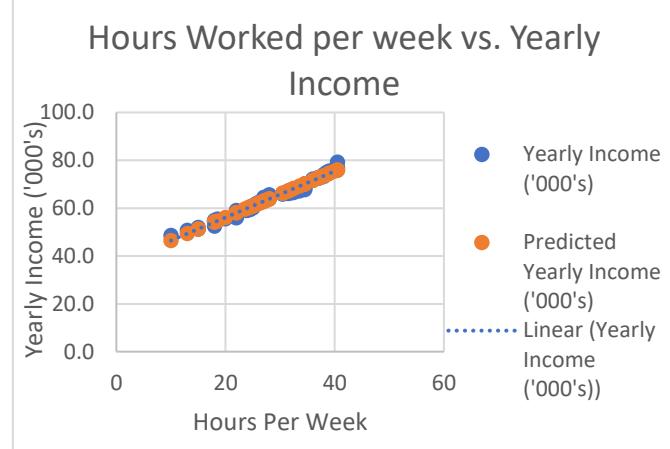
	Hours Per Week	Yearly Income ('000's)
Hours Per Week	1	
Yearly Income ('000's)	0.988941138	1

Question 5:

a) The dependant variable for this analysis is yearly income, with the independent variable being hours worked per week, as the yearly income of the individuals varies depending on the hours worked.

b) Linear equation:
 $y = \beta_0 + \beta_1 x + \epsilon$

c) Standard error is
0.575519349

**SUMMARY OUTPUT****Regression Statistics**

Multiple R	0.988941138
R Square	0.978004574
Adjusted R Square	0.977618689
Standard Error	1.133954302
Observations	59

ANOVA

	df	SS	MS	F	Significance F
Regression	1	3258.925771	3258.925771	2534.448	6.07E-49
Residual	57	73.29358446	1.285852359		
Total	58	3332.219356			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	36.78385598	0.575519349	63.91419512	9.48E-55	35.6314	37.93631	35.6314	37.93631
Hours Per Week	0.966506287	0.01919831	50.34330014	6.07E-49	0.928062	1.00495	0.928062	1.00495

d) Co coefficient of determination: R Square = 0.978004574

END OF ASSIGNMENT