**University of Tunku Abdul Rahman**
**Group Assignment (294 marks)**

**UCCM2233 STATISTIC**

| No. | Name | Student ID | Programme |
|---|---|---|---|
| 1 | Richard Ting Li Zeng | 2301973 | IA |
| 2 | Leong Wei Feng | 2106573 | IA |
| 3 | Leong Sin Yu | 2202026 | IA |
| 4 | Toh Weng Kee | 2202415 | IA |
| 5 | Timothy Lai Yang Yi | 2301240 | IA |
| 6 | Wang YuChen | 2300802 | IA |

Answer **ALL** the questions. Marks allocated for each part of the questions are indicated in bracket at the end of it.
Read the following before you start:
1. It is a grouped assignment. Each group consists of **six** students from the **same faculty**.
2. Fill in all your particulars above.
3. Complete all problems and submit them before the deadline.
The marks allocated are indicated in blankets at the end of each question.
***WARNING!***
· 1. It is a **group assignment** that consists of **six** students.
· 2. **Marks will be deducted** for late submission.
· 2. **You will get ZERO** if you are caught being dishonest or doing any kind of cheating in the assignment.

| | | |
|---|---|---|
| Part A | Question 1 | /25 |
| | Question 2 | /25 |
| | Question 3 | /25 |
| | Question 4 | /25 |
| | Question 5 | /25 |
| | Question 6 | /25 |
| Part B | | |
| Data Set 1 - Employee | Job Category | /6 |
| | Years of Education | /33 |
| | Annual Salary | /33 |
| Data Set 2 - Voter | Gender | /6 |
| | Year of Education | /33 |
| | Age | /33 |
| | Grand Total | /294 |

Part A

Date:

1) a) (i)

| Stem | Leaf |
|---|---|
| 2.00 | 12 31 32 35 40 40 46 48 |
| 2.50 | 02 06 21 30 33 45 45 |
| 3.00 | 11 14 16 17 18 |

key : 2.00 | 31 = 2.00 + 0.31
= 2.31

(ii) $\Sigma x = 2.56 + 2.4 + 2.32 + 2.4 + 2.31 + 2.71 + 2.35 + 2.48 + 2.46 + 2.52 + 3.17 + 2.12 + 3.14 + 2.95 + 3.18 + 2.8 + 2.95 + 3.11 + 3.16 + 2.83$

$= 53.92$ #

$\Sigma x^2 = (2.56)^2 + (2.4)^2 + (2.32)^2 + (2.4)^2 + (2.31)^2 + (2.71) + (2.35)^2 + (2.48)^2 + (2.46)^2 + (2.52)^2 + (3.17)^2 + (2.12)^2 + (3.14)^2 + (2.95)^2 + (2.18)^2 + (2.8)^2 + (2.95)^2 + (3.11)^2 + (3.16)^2 + (2.83)^2$

$= 147.638$ #

∴ standard deviation :

$$s = \sqrt{\frac{1}{n-1}\left(\Sigma x^2 - \frac{1}{n}[\Sigma x]^2\right)}$$

$$= \sqrt{\frac{1}{20-1}\left(147.63 - \frac{1}{20}(53.92)^2\right)}$$

$$= \sqrt{\frac{1}{19}(2.26)}$$

$$= \sqrt{0.12}$$

$$= 0.3456$$

Date:

continued (ii) mean $= \frac{\Sigma x}{N}$

$= \frac{53.92}{20}$

$= 2.696$

$\Sigma |x - \bar{x}| = |2.56 - 2.696| + |2.4 - 2.696| + |2.32 - 2.696|$
$+ |2.4 - 2.696| + |2.51 - 2.696| + |2.71 - 2.696|$
$+ |2.35 - 2.696| + |2.48 - 2.696| + |2.46 - 2.696|$
$+ |2.52 - 2.696| + |3.17 - 2.696| + |2.12 - 2.696|$
$+ |3.14 - 2.696| + |2.95 - 2.696| + |3.18 - 2.696|$
$+ |2.8 - 2.696| + |2.95 - 2.696| + |3.11 - 2.696|$
$+ |3.16 - 2.696| + |2.83 - 2.696|$

$= 0.136 + 0.296 + 0.376 + 0.296 + 0.386$
$+ 0.014 + 0.346 + 0.576 + 0.444 + 0.254$
$+ 0.484 + 0.104 + 0.254 + 0.216 +$
$+ 0.236 + 0.176 + 0.474 + 0.414 + 0.464$
$+ 0.134$

$= 6.08$

$\therefore MAD = \frac{\Sigma |x - \bar{x}|}{n}$

$= \frac{6.08}{20}$

$= 0.304$ #

(iii) The distribution is right skewed.
This is because the long tail of distribution is on the right-hand side.
Moreover, on the left-hand side, there is no tail distribution.

Date:

Continued (iii) $sk = \frac{3(\bar{x} - x)}{s}$

$\because \bar{x} = 2.696$

$s = 0.3456$

$\text{median} = \frac{2.5612 - 71}{2}$

$= 2.635$ #

$\therefore sk = \frac{3(\bar{x} - x)}{s}$

$= \frac{3(2.696 - 2.635)}{0.3456}$

$\approx 0.53$ #

∴ Since $sk > 0$, then there will be positively skewed distribution and there is a right-skewed distribution.

(iv) 10% for 20 students : $20 \times 10\% = 2$

Cut-off value for 90th percentile:

$= \frac{3.16 + 3.17}{2}$

$= 3.165$ #

∴ 2 students graduate with distinction degree and the cut-off value is 3.165.

(v) ~~mean of 19 students~~ $= \frac{53.92 - 2.12}{19}$

$= 2.726$

mean of new 11 student $= 2.65$

$\therefore \frac{\Sigma x}{n} = \frac{(2.726 \times 19) + (2.65 \times 11)}{30}$

$= \frac{80.944}{30}$

$= 2.698$ #

Date:

b) Chebyshev's Theorem:

$$1 - \frac{1}{k^2} = \frac{85}{100}$$

$$-\frac{1}{k^2} = -0.15$$

$$\frac{1}{k^2} = 0.15$$

$$k^2 = \frac{100}{15}$$

$$k = 2.582$$

$$-2.582 = \frac{x - 45}{7} \qquad 2.582 = \frac{y - 45}{7}$$

$$-18.074 = x - 45 \qquad 18.074 = y - 45$$

$$x = 26.926 \qquad y = 18.074 + 45$$

$$= 63.074$$

$\therefore$ Range $= 63.074 - 26.926$

$= 36.148$ //

2)

a) i)

- Female users are mainly between the ages of 20 and 40
- Male users are mainly between the age of 40 and 60
- Members are all between the ages of 18 and 75
- Male members are generally older than female members.

ii)

- Onion Club Gold members are more male-dominated
- Onion Club members are more female-dominated.
- Club members have the highest proportion
- There are about the same number of women in Onion Club Gold members as there
- Are in Onion Club members.
- Onion Club Gold members are the only club with more female members than male members.
- Internet catalog Customers have the lowest proportion.

iii)

76+154+397=627

154/627 *100%=24%

b) i)

stratified random sampling, as she is going to collect the sample from the population that differs widely in the possession of a characteristics, is divided into different strata and the members in each stratum have the similar characteristic.

ii)

- Get a list of all baby records in the clinic database.
- Assign a unique identifier to each record.
- Divide the records into two layers: boys and girls.
- Random sampling from all levels.
- The selected boys and girls were recorded.
- Analyse the selected record.
- Calculate the average time it takes for the baby.
- Compare averages with statistics.
- Draw a chart and write the final report.

No.

Date

3. (a) (i) $\frac{RM230.25 + RM682.75}{2} = RM456.50$

$k_L = \frac{RM230.25 - RM456.50}{90.5} = -2.5$ $\quad k_H = \frac{RM682.75 - RM456.50}{90.5} = 2.5$

$1 - \frac{1}{(2.5)^2} = 0.84$

$0.84 \times 100\% = 84\%$

∴ At least 84% of employees receive an increment of RM230.25 to RM682.75.

(ii) The theorem used in part (i) is Chebyshev's Theorem, this theorem helps to determine where most of your data fall within a distribution of values because the skewness of the data distribution is unknown.

(iii) $CV_A = \frac{90.5}{456.5} \times 100\%$ $\quad CV_B = \frac{20.2}{380.4} \times 100\%$

$= 19.82\%$ $\quad = 5.31\%$

∴, Company A has higher variation of salaries increment.

(b) Mean : The value of mean is the largest. This is because the mean is effect by the extreme values in the right tail, so it cause the lengthen of the distribution direction. The mean will get higher values.

Mode : The value of mode is the smallest. This is because mode is the value that occur with the highest frequency in data set, and the highest frequency will be focus on left side of the distribution, the summit of the distribution will be pulled towards left, so that the value of mode will be less than mean and median in positive skewed distribution.

Median : Value of median is between mean and mode. The value of median in positively skewed distribution will larger than mode less than mean.

A set of values with a single peak is skewed to right or positively skewed :

Mode < Median < Mean

No.

Date :

(c) Stratified Sampling :

- Population is subdivided into at least two subgroups / strata that share the same characteristics / homogenous.
- Samples are randomly selected from each of these strata.
- The collection of all samples from all strata gives the stratified random sample.
- Size of the sample selected from different strata are proportional to size of the strata in the population.

Cluster Sampling :

- Population is divide into groups called cluster.
- Each clusters is representative of the population.
- Clusters are non-overlapped and homogenous to each other.
- Selected a random sample of clusters. Finally, selected a random sample or all of elements from each of the selected cluster.
- These samples from different clusters will be grouped to form a cluster random sampling.

4.

**A. Descriptive Statistics:**

--Consists of methods for organizing, presenting, and summarizing data by using tables, graphs, and descriptive measures.
--State facts and proven outcomes from a population.
--It deals with describing and analyzing raw data collected from the fieldwork.
--It presents information in a convenient, usable, and comprehensible form.
--Focus on describing the visible characteristics of a dataset.

**Inferential Statistics:**

--Consists of methods that use sample results to perform estimations and hypothesis tests about a population of the study.
--It deals with the problems of making inferences or drawing conclusions about a population based on information obtained from the samples.
--Analyse samplings to make predictions about larger populations.
--Focus on making predictions or generalizations about a larger dataset, based on a sample of those data.

**B. I.** Cluster Sampling. It divided the population into groups that call cluster so that each cluster is representative of the population. The clusters are non-overlapped and homogeneous to each other, and it will select a random sample of clusters. These samples that had been selected will be grouped to form a cluster random sample. So that this sampling technique is suitable for the researcher to select the samples required from various departments.

**II.** First, the researcher needed to define the cluster. The clusters will be the various departments within the multinational corporation. Each department will be considered a separate cluster. The researcher needed to confirm that each cluster is homogeneous. The researcher may randomly select a specific number of departments from the corporation. After selecting the clusters, the researcher must confirm that each cluster is homogenous and can collect the data. After collecting the data, the researcher can analyse the result to compare the job productivity of workers who work from home and at the office during the Covid-19 pandemic.

**III. Quantitative variable**—Continuous variables. The studies need to know the job productivity of workers, so it means that the data should take in any value within a range, continuous variables can take on an uncountably infinite number of values within a given range, so it is suitable for the research.

**Qualitative variable**—Ordinal variables. Ordinal variables can compare the productivity of workers who work from home and workers who work at the office, and it also can be the order and ranking of the workers.

**IV. Continuous variable**—Mean: Calculate the average weekly hours of workers that work from home and workers that work at the office to compare the overall time spent on work.
Histogram: Create a histogram to visualize both two groups weekly working hours, and it also can reveal patterns or clusters in the working hour data.

**Ordinal variables**—Frequency Distribution: Create a frequency table to show the distribution of both groups' productivity rates and provide an overview of the productivity distribution of both groups.

Bar Chart: It visualises the rate of productivity in both groups. A bar chart will display the number of staff members in each productivity category, making it easier to compare the rate of productivity between both groups.

Subject: _____ No: _____ Date: _____

# Assignment 1 (Part A)

5.)

| $x$ | $f$ | $fx$ | $x^2$ | $fx^2$ | $cf$ |
|---|---|---|---|---|---|
| 0 | 90 | 0 | 0 | 0 | 90 |
| 1 | 120 | 120 | 1 | 120 | 210 |
| 2 | 70 | 140 | 4 | 280 | 280 |
| 3 | 40 | 120 | 9 | 360 | 320 |
| 4 | 20 | 80 | 16 | 320 | 340 |
| 5 | 10 | 50 | 25 | 250 | 350 |
| $\Sigma x = 15$ | $\Sigma f = 350$ | $\Sigma fx = 510$ | $\Sigma x^2 = 55$ | $\Sigma fx^2 = 1330$ | |

a) mean $= \frac{510}{350} = 1.4571$

median $= \left(\frac{n+1}{2}\right)^{th}$ , $n = 350$

$= \frac{350+1}{2}$

$= 175.5^{th}$

$= 1$

b) $s^2 = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$

$= \frac{1330}{350} - \left(\frac{510}{350}\right)^2$

$= 1.6767$

$s = \sqrt{1.6767}$

$= 1.2949$ //

$Q_2 = \frac{\frac{350}{2} + \frac{350}{2} + 1}{2}$

$= \frac{175^{th} + 176^{th}}{2}$

$= 176.5^{th}$

$Q_1 = \frac{350}{4}$

$= 87.5^{th}$

$= 0$

$Q_3 = \frac{176^{th} + 1}{2}$

$= \frac{177^{th}}{2}$

$= 88.5^{th}$

$= 2$

Quartile deviation $= \frac{2-0}{2}$

$= 1$

c) The dispersion are more appropriate to be used to summarize the data set in the frequency distribution. It is because can provide context to central tendency. When measures of central tendency (mean, median, mode) give a single representative value, measures of dispersion complement them by providing additional context. It will show how the data points are distributed around the central value, offering a more complete picture of the dataset. Besides that, also can compare

Subject: No.
Date:

the data sets. When compare two or more datasets, measure of dispersion help determine which dataset has greater variability. Lastly, the another point is outlier detection. The measures of dispersion such as the range or standard deviation, can identify outliers. Outliers can impact statistical analyses. In conclusion, it can enrich data analysis and interpretation, understand the data and making decision.

6) a)

| x | f | m | $fm^2$ | fm | Cumulative Frequency |
|---|---|---|---|---|---|
| 1500-2999 | 13 | 2249.5 | 6579326325 | 29243.5 | 13 |
| 3000-4499 | 33 | 3749.5 | 463938758.3 | 123733.5 | 46 |
| 4500-5999 | 22 | 5249.5 | 606259505.5 | 115489 | 68 |
| 6000-7499 | 14 | 6749.5 | 637780503.5 | 94493 | 82 |
| 7500-8999 | 10 | 8249.5 | 680542502.5 | 82495 | 92 |
| 9000-10499 | 8 | 9749.5 | 760422002 | 77996 | 100 |
| Total | 100 | | 3214726525 | 523450 | |

$\because \Sigma f = 100$, $\Sigma fm^2 = 3214726525$, $\Sigma fm = 523450$

$\therefore$ mean $= \frac{\Sigma fm}{\Sigma f}$ median $= L + \left(\frac{\frac{1}{2}N - P_{f3}}{f_m}\right)C$

$= \frac{523450}{100}$ $= 4499.5 \left(\frac{\frac{100}{2} - 46}{22}\right) 1500$

$= 5234.5$ # $= 4772.23$ #

b) Skewed to right because mean is greater than median

c) sample standard deviation:

$$s = \sqrt{\frac{1}{n-1}\left(\Sigma fm^2 - \frac{1}{n}[\Sigma fm]^2\right)}$$

$$= \sqrt{\frac{1}{100-1}\left(3214726525 - \frac{1}{100}[523450]^2\right)}$$

$= 2189.8$ #

d) Let $CV_1$ as the CV of 10 years ago,
$CV_2$ as the CV of recent monthly,

(1) $CV = \frac{s}{\bar{x}}$ $CV_2 = \frac{2189.8}{5234.5} \times 100\%$

$= 41.83\%$ #

Date:

continue d) ii) $CV_1 = \frac{500}{1800} \times 100\%$

$= 27.78\%$.

∴ 10 years ago ($CV_1$) is smaller than recent monthly income ($CV_2$), that's mean 10 years ago is relatively more consistent in monthly income per household in a suburban town of Malaysia.

Part B

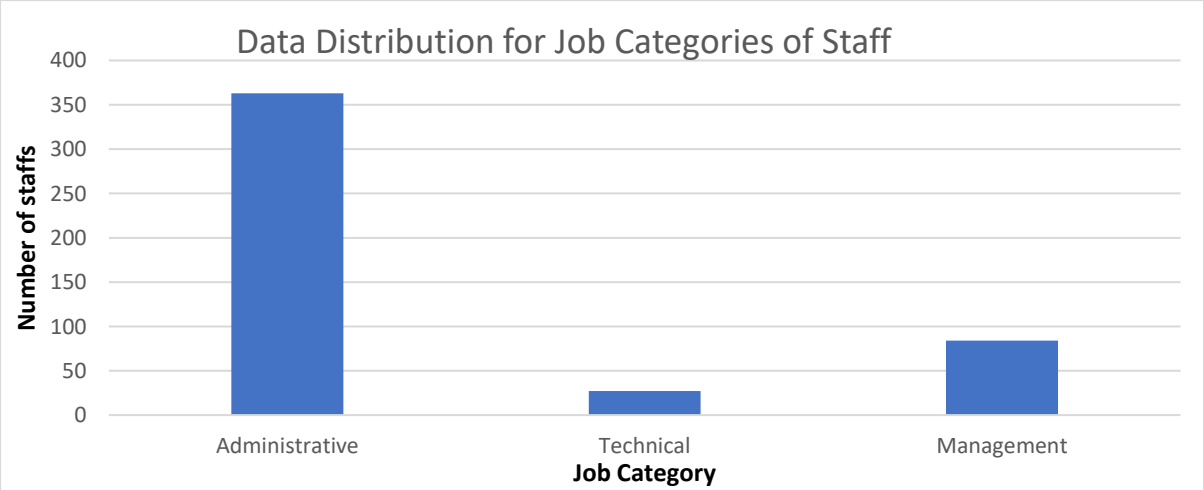
Data Set 1 – Employee

Job Category

i)

| Job Category | Number of staffs | Percentage |
|---|---|---|
| Administrative | 363 | 76.58% |
| Technical | 27 | 5.70% |
| Management | 84 | 17.72% |
| Total | 474 | 100.00% |

Comment: The job that has the highest number of staff is Administrative with a percentage of 76.58% and the job that has the lowest number of staff is Technical with a percentage of 5.70%.



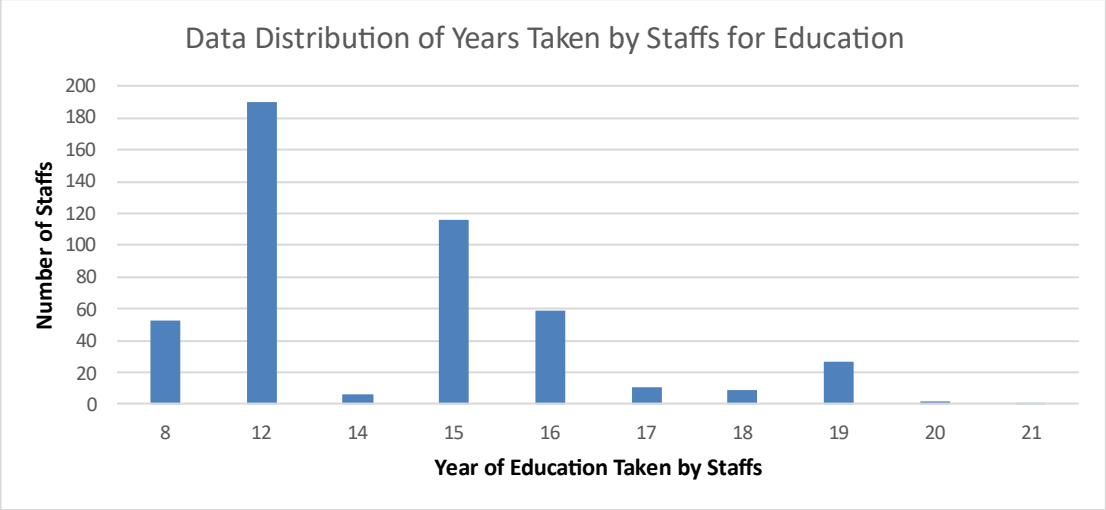

Comment: The data distribution for job categories of staff is skewed to the right because technical and management only have a few staff.

Years of education

i)

| Years of education taken by staff | Number of staffs | Percentage |
|---|---|---|
| 8 | 53 | 11.18% |
| 12 | 190 | 40.08% |
| 14 | 6 | 1.27% |
| 15 | 116 | 24.47% |
| 16 | 59 | 12.45% |
| 17 | 11 | 2.32% |
| 18 | 9 | 1.90% |
| 19 | 27 | 5.70% |
| 20 | 2 | 0.42% |
| 21 | 1 | 0.21% |
| Total | 474 | 100.00% |

Comment: There are 40.08% of the staff have 12 years of education while there are only 0.21% of the staff have 21 years of education.




Comment: The data distribution of years taken by staff for education is skewed to the right which means that there are fewer staff having education that more than 12 years.

ii)

| | |
|---|---|
| Mean = | 13.49 |
| Median = | 12.00 |
| Mode = | 12.00 |

Comment:

The average years taken by staff for Education is 13.49 years.

50% of all the years taken by staff for Education less than 12 years.

The most common years taken by staff for Education is 12 years.

Since mean > median, the data distribution for years taken by staff for Education is right skewed, ie there are a few staff who have a long time in Education.

iii)

| | |
|---|---|
| Standard deviation = | 2.88 |
| Mean absolute deviation = | 2.42 |
| Quartile deviation = | 1.50 |

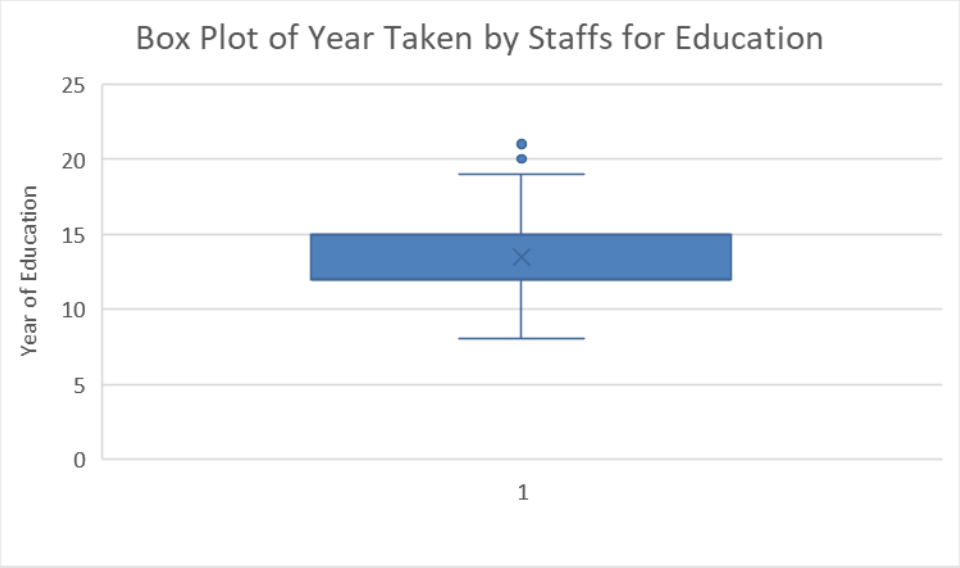
Comments:

The distribution of years taken by staff for education is scattered widely from the mean since the value for the standard deviation is large.

The average of the absolute values of the data of years taken by staff for education deviations from the mean is 2.42.

The distribution of years taken by staff for education is scattered widely from the median since the value for the quartile deviation is large.

iv)



| | |
|---|---|
| Pearson's measure of skewness | 1.55 |

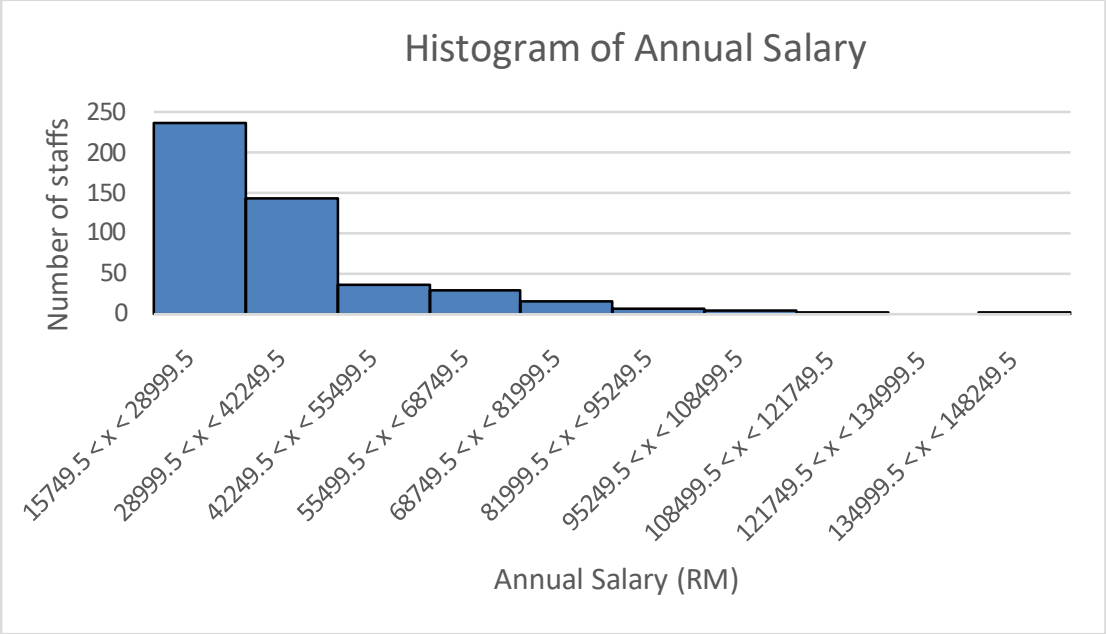
Comment: The Pearson skewness coefficient of this graph is greater than zero, the heavy tail is on the right, and the distribution is skewed to the right. The top edge of this graph is close to 20, and the bottom edge is about 7 to 8, with two outliers.

Annual Salary

i)

| Annual Salary | Number of staffs | Class Boundary | Percentage |
|---|---|---|---|
| 15750-28999 | 238 | 15749.5 < x < 28999.5 | 50.21% |
| 29000-42249 | 144 | 28999.5 < x < 42249.5 | 30.38% |
| 42250-55499 | 35 | 42249.5 < x < 55499.5 | 7.38% |
| 55500-68749 | 29 | 55499.5 < x < 68749.5 | 6.12% |
| 68750-81999 | 16 | 68749.5 < x < 81999.5 | 3.38% |
| 82000-95249 | 6 | 81999.5 < x < 95249.5 | 1.27% |
| 95250-108499 | 4 | 95249.5 < x < 108499.5 | 0.84% |
| 108500-121749 | 1 | 108499.5 < x < 121749.5 | 0.21% |
| 121750-134999 | 0 | 121749.5 < x < 134999.5 | 0.00% |
| 135000-148249 | 1 | 134999.5 < x < 148249.5 | 0.21% |
| Total | 474 | | 100.00% |

Comment: There are 50.21% of the staff have a salary that ranges from 15750-28999 and there are only 0.21% of the staff have a salary that range from 108500-121749 and 135000-148249.




Comment: The data on staff's annual salary is skewed to the right because there is only a few staff having a salary that is more than 28999.5.

ii)

| | |
|---|---|
| Mean = | 34419.57 |
| Median = | 28875.00 |
| Mode = | 30750.00 |

Comment:

The average annual salary is RM34419.57.

50% of all the staff have an annual salary of less than RM28875.

The most common staff annual salary is RM30750.

Since mean > median, the data distribution for the staff annual salary is right skewed, ie there are a few staff having a very high annual salary.

iii)

| | |
|---|---|
| Standard deviation = | 17057.64 |
| Mean absolute deviation = | 12016.32 |
| Quartile deviation = | 6468.75 |

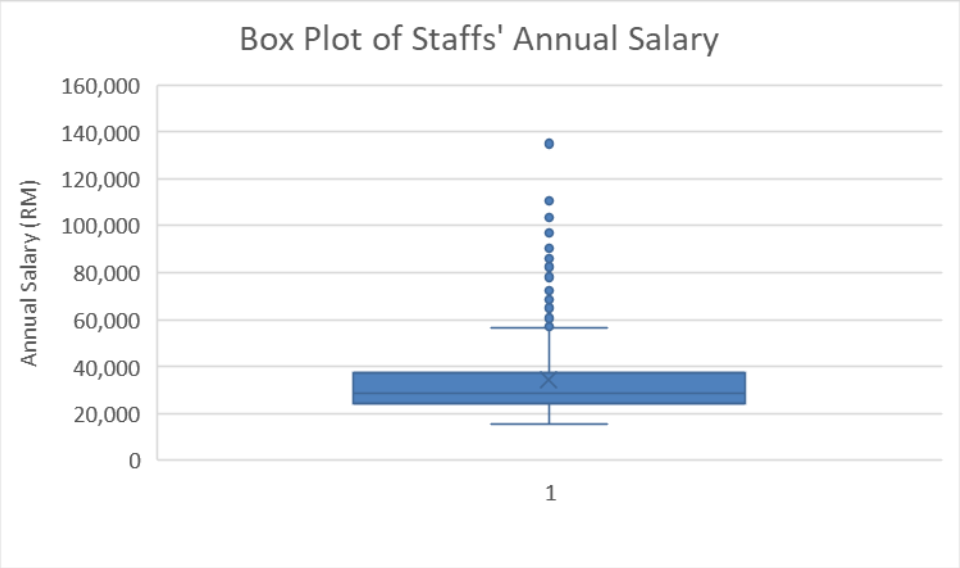
Comments:

The distribution for the staff's annual salary is scattered widely from the mean since the value for the standard deviation is large.

The average of the absolute values of the data for staff annual salary deviations from the mean is RM12016.32.

The distribution for the staff annual salary is scattered widely from the median since the value for the quartile deviation is large.

iv)



| | |
|---|---|
| Pearson's measure of skewness | 0.98 |

Comment: The Pearson skewness coefficient of this graph is greater than zero, the heavy tail is on the right, and the distribution is skewed to the right. The top edge of this graph is close to 60,000 and the bottom edge is close to 20,000, with multiple outliers.

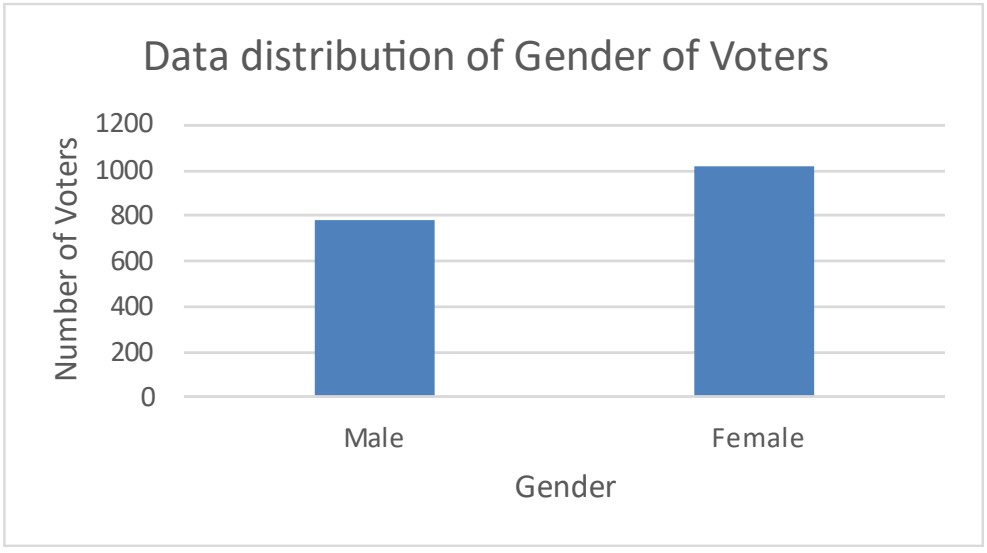
Data Set 2 - Voters

Gender

i)

| Gender | Number of Voters | Percentage |
|---|---|---|
| Male | 784 | 43.58% |
| Female | 1015 | 56.42% |
| Total | 1799 | 100.00% |

Comment: The gender that has a greater number of voters is female with a percentage of 56.42% and male-only have 43.58%.



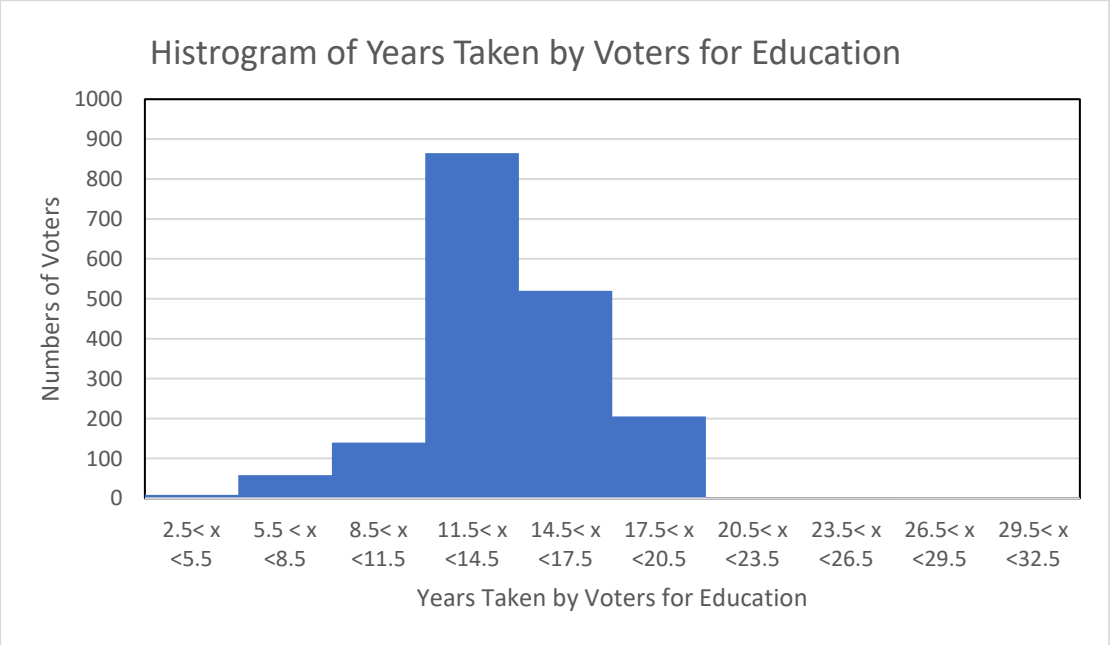

Comment: The data distribution of gender of voters is skewed to the left because the number of male voters is lesser than female voters.

Years of education

i)

| Years Taken by Voters for education | *Class Boundary* | *Numbers of Voters* | Percentage |
|---|---|---|---|
| 3~5 | 2.5< x <5.5 | 9 | 0.50% |
| 6~8 | 5.5 < x <8.5 | 58 | 3.22% |
| 9~11 | 8.5< x <11.5 | 140 | 7.78% |
| 12~14 | 11.5< x <14.5 | 865 | 48.08% |
| 15~17 | 14.5< x <17.5 | 520 | 28.90% |
| 18~20 | 17.5< x <20.5 | 205 | 11.40% |
| 21~23 | 20.5< x <23.5 | 0 | 0.00% |
| 24~26 | 23.5< x <26.5 | 0 | 0.00% |
| 27~29 | 26.5< x <29.5 | 1 | 0.06% |
| 30~32 | 29.5< x <32.5 | 1 | 0.06% |
| Total | | 1799 | 100.00% |

Comment: The most number of voters have years of education is within 12-14 years which is 48.08% of the total voters and the least number of voters have years of education is 27-32 years with a percentage of 0.06%.




Comment: The data distribution of years taken by voters for education is skewed to the right because only few voters have education that more than 20 years.

ii)

| | |
|---|---|
| Mean = | 13.97 |
| Median = | 14.00 |
| Mode = | 12.00 |

Comment:

The average years taken by voters for Education is 13.97 years.

50% of all the years taken by voters for Education less than 14 years.

The most common years taken by voters for Education is 12 years.

Since mean > median, the data distribution for years taken by staff for Education is right skewed, ie a few voters are having a long time in Education.

iii)

| | |
|---|---|
| Standard deviation = | 2.89 |
| Mean absolute deviation = | 2.30 |
| Quartile deviation = | 2.00 |

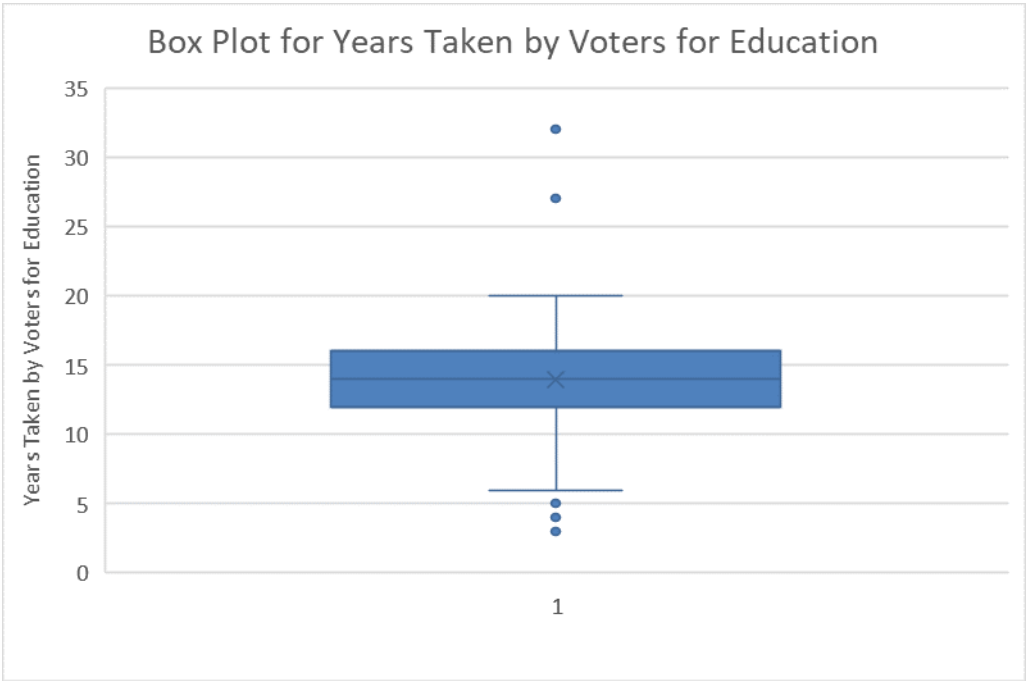
Comments:

The distribution of years taken by voters for education is scattered widely from the mean since the value for the standard deviation is large.

The average of the absolute values of the data of years taken by voters for education deviations from the mean is 2.30.

The distribution of years taken by voters for education is scattered widely from the median since the value for the quartile deviation is large.

iv)



| | |
|---|---|
| Pearson's measure of skewness | -0.03 |

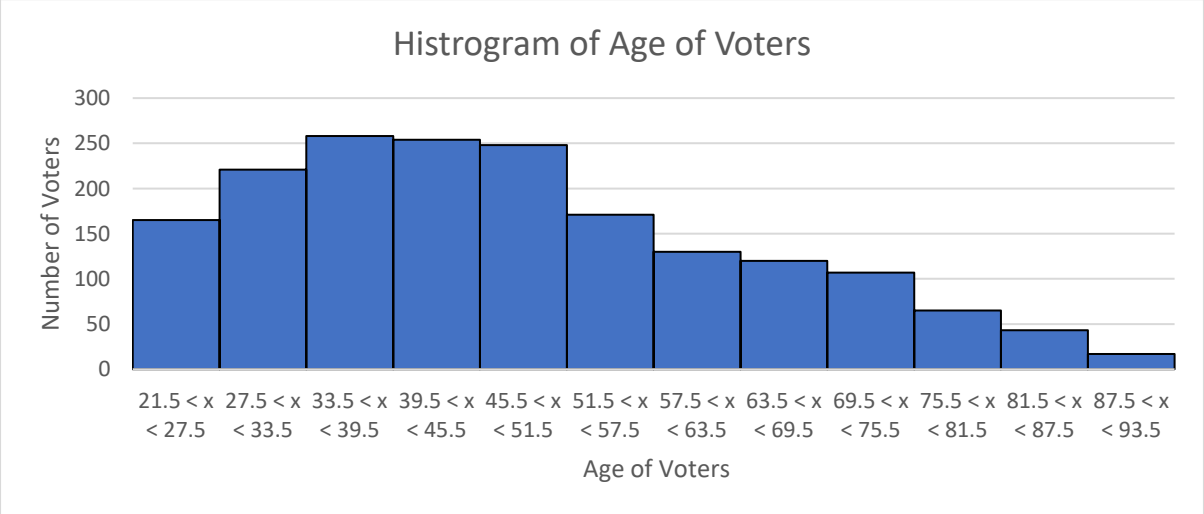
Comment: The value of Pearson's coefficient of skewness displays that the distribution is negatively skewed, so it means that the tail of the distribution extended to the lower value. The box plot is in a median skewness, so it displays symmetry.

Ages of Voters

i)

| Age of Voters | Class boundary | Number of Voters | Percentage |
|---|---|---|---|
| 22-27 | 21.5 < x < 27.5 | 165 | 9.17% |
| 28-33 | 27.5 < x < 33.5 | 221 | 12.28% |
| 34-39 | 33.5 < x < 39.5 | 258 | 14.34% |
| 40-45 | 39.5 < x < 45.5 | 254 | 14.12% |
| 46-51 | 45.5 < x < 51.5 | 248 | 13.79% |
| 52-57 | 51.5 < x < 57.5 | 171 | 9.51% |
| 58-63 | 57.5 < x < 63.5 | 130 | 7.23% |
| 64-69 | 63.5 < x < 69.5 | 120 | 6.67% |
| 70-75 | 69.5 < x < 75.5 | 107 | 5.95% |
| 76-81 | 75.5 < x < 81.5 | 65 | 3.61% |
| 82-87 | 81.5 < x < 87.5 | 43 | 2.39% |
| 88-93 | 87.5 < x < 93.5 | 17 | 0.94% |
| Total | | 1799 | 100.00% |

Comment: There are 14.34% of the voters are in the age of 34-39 and there are 0.94% of the voters are in the age of 0.94%




Comment: The data on the age of voters is skewed to the right because, after the age of 34-39, voters become fewer and lesser.

ii)

| | |
|---|---|
| Mean = | 47.88 |
| Median = | 46.00 |
| Mode = | 48.00 |

Comment:

The average age of voters is 47.88 years old.

50% of all the voters have aged less than 46 years old.

The most common staff age is 48 years old.

Since mean > median, the data distribution for the voters' age is right skewed, ie a few voters are having very high age.

iii)

| | |
|---|---|
| Standard deviation = | 16.31 |
| Mean absolute deviation = | 13.41 |
| Quartile deviation = | 12.00 |

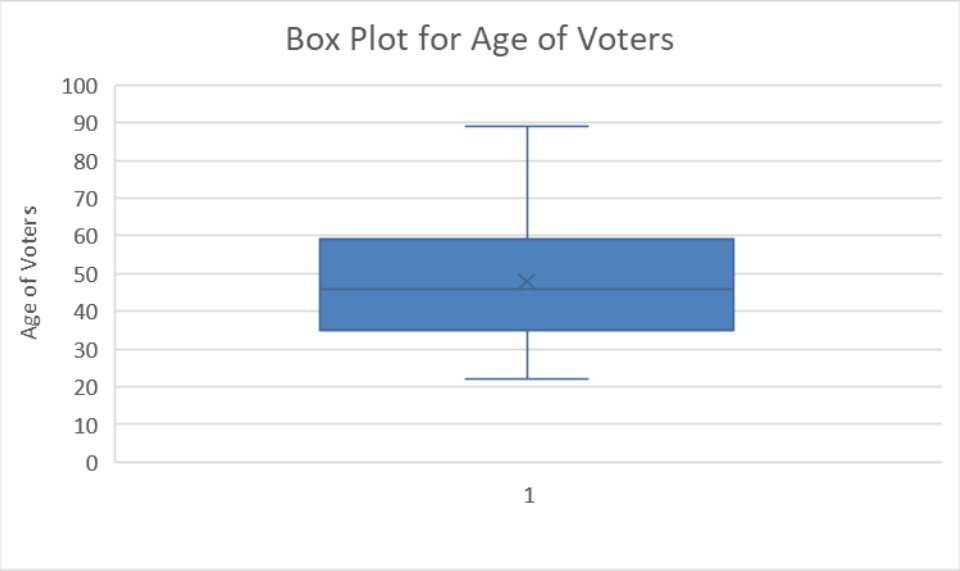
Comments:

The distribution of the age of voters is scattered widely from the mean since the value for the standard deviation is large.

The average of the absolute values of the data for voters' age deviations from the mean is 13.41.

The distribution of age of voters is scattered widely from the median since the value for the quartile deviation is large.

iv)




| | |
|---|---|
| Pearson's measure of skewness | 0.35 |

Comment: The value of Pearson's coefficient of skewness displays that the distribution is positively skewed, so it means that the tail of the distribution extended to the higher value. The box plot has a positive skewness, so it displays right-skewed.