GHULAM ISHAQ KHAN INSITITUTE
(GIKI)

NAME:- MUHAMMAD HARIS. SECTION:- BSCS.
REG NO:- 2023428. COURSE:- DS-221

Task no.1

Prove in general that std. dataset $\{\hat{x_i}\}$ derived from $\{x_i\}$ has a mean of 0 and st. deviation of 1.

Standardizing the dataset involves the derivation of z score using the formula below.

$$\hat{x_i} / z_i = \frac{x_i - \mu}{\sigma} \quad \text{--- (i)}$$

where - x_i is the point in the dataset.
- μ is the mea of given data set
- σ is the std. deviation of the given dataset.

We need to prove that $\{\hat{x_i}\}$ has a;

mean = 0 —— (a)
Std. deviation = 1. —— (b)

* Proving (a)

$$\mu_{\hat{x_i}} = \frac{1}{n} \sum_{i=1}^{n} z_i / \hat{x_i} \quad \text{--- (ii)}$$

Substitute

(i) into (ii)

$$\mu_{\hat{x_i}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)$$

Since σ is const.

$$= \frac{1}{\sigma} \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)$$

By appling addition law on summation
/as a linear operator:-

$$= \frac{1}{\sigma} \left(\frac{1}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} \mu \right)$$

Since μ is the mean of the original data set

$$\therefore \Rightarrow \frac{1}{\sigma} (\mu - \mu) \Rightarrow \frac{1}{\sigma} \cdot 0 \Rightarrow 0$$

—— (iii)

hence proved!

* proving (b)

Std. deviation of a standized set is given by:-

$$\sigma_z = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (z_i - \mu_z)^2}$$

Now as proved in (iii)

$$\sigma_z = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (z_i)^2}$$

also by —— (i) $\Rightarrow \sigma_z = \sqrt{\frac{1}{n} \left(\frac{1}{\sigma}\right)^2 \sum_{i=1}^{n} (x_i - \mu)^2}$

$$\sigma_z = \frac{1}{\sigma}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}$$

↳ equal to std of original set.

$$\sigma_z = \frac{1}{\sigma}\cdot\sigma = 1 \quad \text{---(iv)}$$

hence proved!

TASK NO.2

1. FIND SAMPLE MEAN & MEDIAN.

MEAN

Sample mean = $\frac{\text{Sum of all value of a dataset}}{\text{no. of values of a dataset}}$.

$$\Rightarrow \mu = \frac{\sum x_i}{N} \quad \text{---(i)}$$

By inserting values.

$$\Rightarrow \frac{572+572+573+568+569+575+565+570}{8}$$

$\Rightarrow$ 570.5 mm.

MEDIAN

Step 1: arrange the values in ascending order.

$\Rightarrow \{565, 568, 569, 570, 572, 572, 573, 575\}$

No. of values = 8 (even) $\therefore$ picking two middle values.

$$\text{Median}_{\text{Even!}} = \frac{\text{Sum of 2 middle values}}{02} \quad \text{——(ii)}$$

$$= \frac{570 + 572}{2} \Rightarrow 571\,mm.$$

2. RANGE, SAMPLE VARIANCE & STD DEVIATION

Range

$\Rightarrow$ Max − Min ——(i)

$575 - 565 = 10\,mm$

(Answer).

SAMPLE VARIANCE:-

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{——(i)}$$

where

$\bar{x}$ = sample mean.

∴ Solving first

$$\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$\Rightarrow (572 - 570.5)^2 + (572 - 570.5)^2 + (573 - 570.5)^2 + (568 - 570.5)^2 + (569 - 570.5)^2 + (575 - 570.5)^2 + (565 - 570.5)^2 + (570 - 570.5)^2$

$\Rightarrow 70.$

$$S^2 = \frac{70}{8-1} = \frac{70}{7} = 10\,mm^2$$

STd. deviation:

Square root of variance.

$S = \sqrt{S^2} = \sqrt{70} \Rightarrow$

$\approx 3.16\,mm$

TASK : 3

$\longrightarrow$ P.T.O

Mean & Median:-
$\bar{x}$

The difference b/w mean (570.5mm) & Median (571mm) is not too much but rather very little, this too can be (attributed to **the presence of outliers** / **small skew** in data, like in our case it is 565). —(a)

Mean & Ideal Mean:-
$\bar{x}$

Same reason as stated in statement (a) can be attributed to the difference b/w mean & ideal mean.

* Suggest overall room for improvement.

4.

Mean:-

The mean is 570.58mm that is very close to the ideal mean ∴ suggesting a good production quality with desired outcomes.

Std. deviation:-

Being 3.16mm which show(s) a moderate variability in the tire diameter. While the avg is close to ideal, there are fluctuations is size of the tires.

Range: It being 10 mm shows that the diametre lies b/w 565 mm to 575 mm being a noticeable difference.

* Over all quality of tires is good on avg however there is still room for improvement to reduce variability in size, achieve consiste and accurate results.

TASK NO. 3

CONSTRUCTING BOX PLOTS:-

Step no.1. Arrange the dataset in ascending order.

{702, 765, 785, 811, 832, 855, 896, 902, 905, 918, 919, 920, 923, 929, 936, 938, 948, 950, 956, 958, 958, 970, 972, 978, 1009, 1009, 1022, 1035, 1037, 1045, 1067, 1085, 1092, 1102, 1122, 1126, 1151, 1156, 1157, 1162, 1170, 1195, 1195, 1196, 1217, 1237, 1311, 1333, 1390}

⇒ Considering the data to be ranked accordingly.

TASK NO. 3
CNTD.

Step no. 2.
Finding Q_1 & Q_3

$Q_1 = 25^{th}$ percentile.
$Q_2 = 75^{th}$ percentile.

$$Q_1 = \frac{50.25}{100} = 12.5$$

$$Q_2 = \frac{50.75}{100} \Rightarrow 37.5$$

Step no. 3
Using intepolation to find exact values.

$$Q_1 = X_{12} + (0.5)(X_{13} - X_{12})$$
$$\Rightarrow 920 + (0.5)(923 - 920)$$
$$\Rightarrow 921.5$$

$$Q_3 \Rightarrow X_{37} + (0.5)(X_{38} - X_{37})$$
$$\Rightarrow 1156 + (0.5)(1157 - 1156)$$
$$\Rightarrow 1156.5$$

Step 4:-
$$IQR \Rightarrow Q_3 - Q_1 \Rightarrow 1156.5 - 921.5$$
$$\Rightarrow 235.$$

Step no. 6.

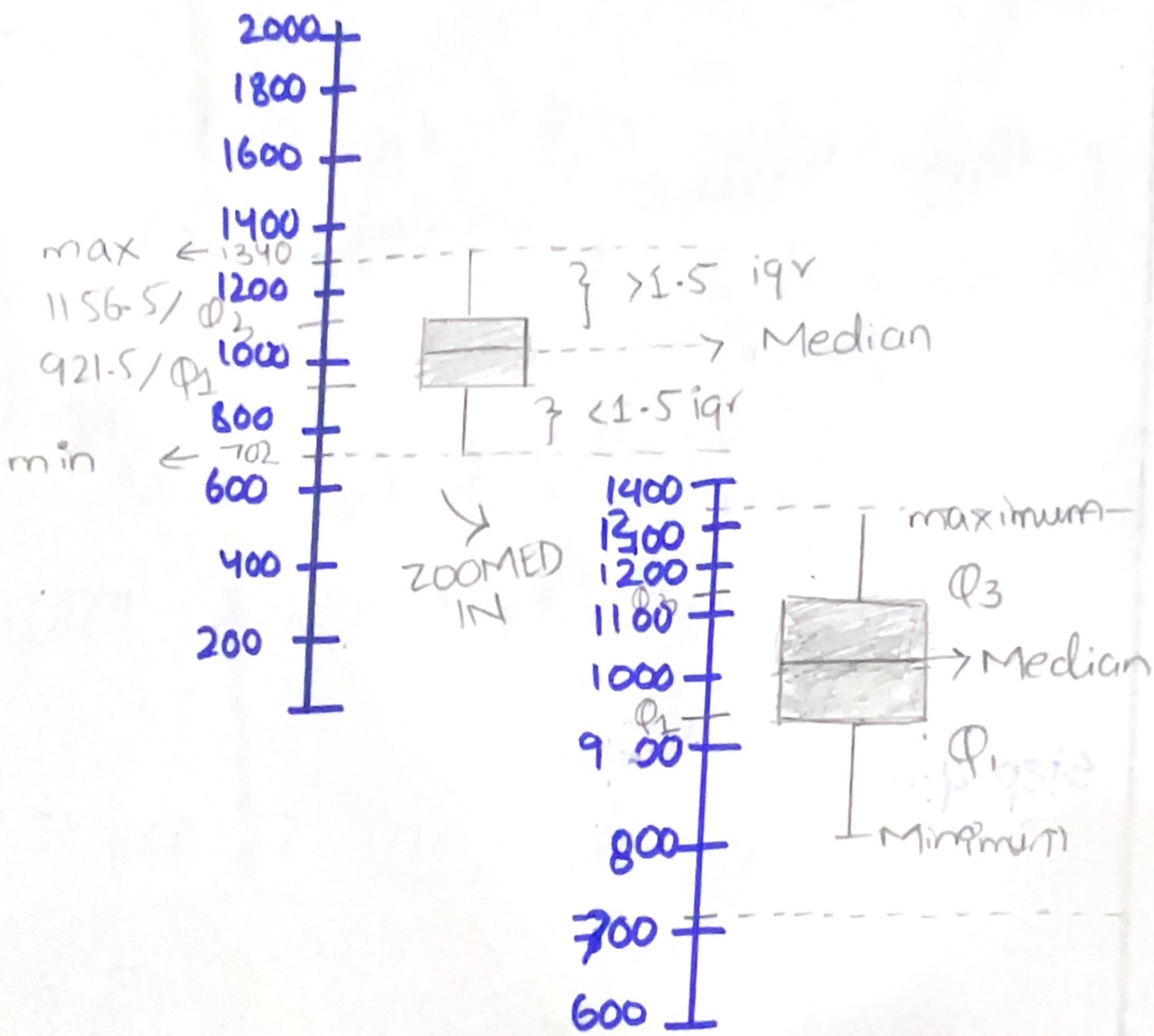
$$\text{Lower Bound / min} = Q_1 - 1.5(IQR)$$
$$= 925.5 - 1.5(235)$$
$$= 569$$

$$\text{Upper Bound / max} = Q_2 + 1.5(IQR)$$
$$= 1156.5 + 1.5(235)$$
$$= 1509.$$

As both of these valus donot exist in our datasets ∴

min ⇒ 702

max ⇒ 1340.



TASK NO. 4:-

1. Sample mean & Sample std. deviation

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{---} \quad (a)$$

$\sum x_i = 242.99$

$\Rightarrow \; x = \frac{242.99}{36} \Rightarrow 6.7497 \approx 6.75$

also.

$$s = \sqrt{\frac{1}{n-1} \sum_{n=i}^{n} (X_i - \bar{X})^2}$$

$$\text{Std. dev.} = \sqrt{\frac{(6.72-6.75)^2 + (6.77-6.75)^2 \ldots + (6.72-6.75)^2}{35}}$$

$\text{Std. dev.} = 0.051$ Ans!

2. Construct Relative frequency Histogram.

Step1 Group the data into intervals and calculate the frequency of each group.

Range of the data $= 6.62 - 6.82$.
spread $\Rightarrow 0.20$.

Creating intervals of 0.02

6.62 - 6.64, 6.64 - 6.66, 6.66 - 6.68, 6.68 - 6.70
6.70 - 6.72, 6.74 - 6.76, 6.76 - 6.78, 6.78 - 6.80
6.80 - 6.82

Step 2 Counting the no. of data points that fall in a particular frequency.

6.62 - 6.64 ⇒ 4 datapoints.

6.64 - 6.66 ⇒ 4

6.66 - 6.68 ⇒ 5

6.68 - 6.70 ⇒ 3

6.70 - 6.72 ⇒ 7

6.72 - 6.74 ⇒ 4

6.74 - 6.76 ⇒ 2

6.76 - 6.78 ⇒ 7

6.78 - 6.80 ⇒ 2

6.80 - 6.82 ⇒ 2

Step 3. Calculating relative frequencies and dividing each with total no. of data points.

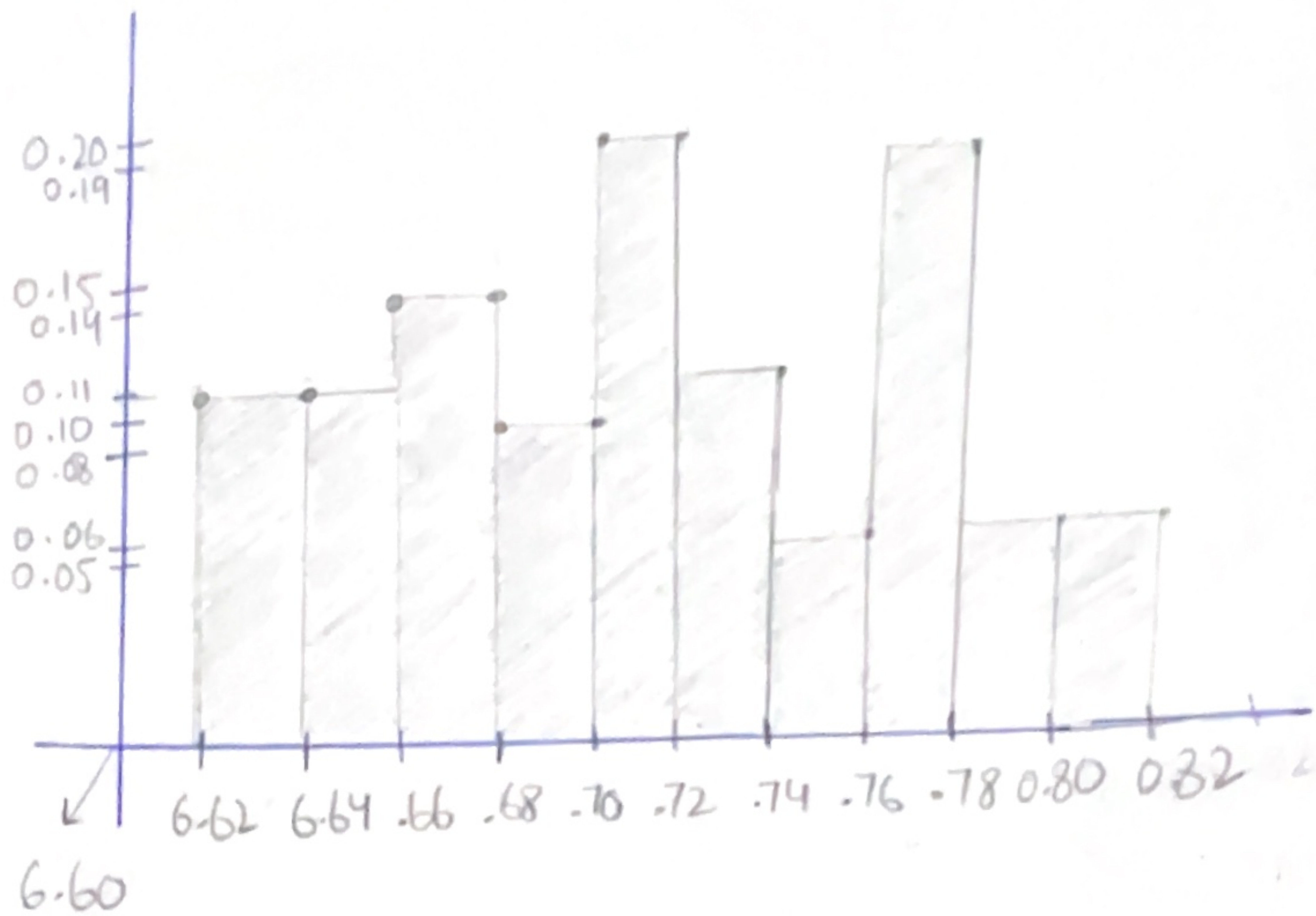
$$\left\{ \frac{4}{36}, \frac{4}{36}, \frac{5}{36}, \frac{3}{36}, \frac{7}{36}, \frac{4}{36}, \frac{2}{36}, \frac{7}{36}, \frac{2}{36}, \frac{2}{36} \right\}$$

{0.11, 0.11, 0.14, 0.08, 0.19, 0.11, 0.06, 0.19, 0.06, 0.06}.

Step 4:-

On the x-axis plotting frequencing.

On the Y-axis plotting relative frequencing.



Comments.

The histogram likely appears to be left-skewed this is because of the more occurance of higher values before the mean.

THE END!

10/12/24