

HW 6 Solutions

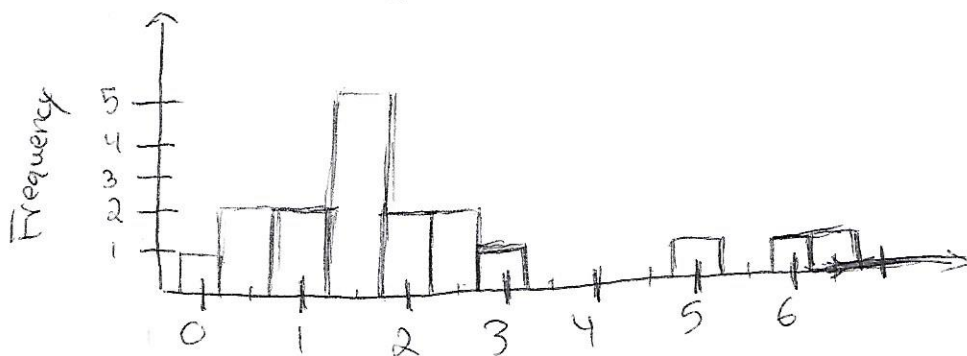
1. Sorted data

$\{0, 0.5, 0.5, 1, 1, 1.5, 1.5, 1.5, 1.5, 1.5,$
 $2, 2, 2.5, 2.5, 3, 5, 6, 6.5\}$

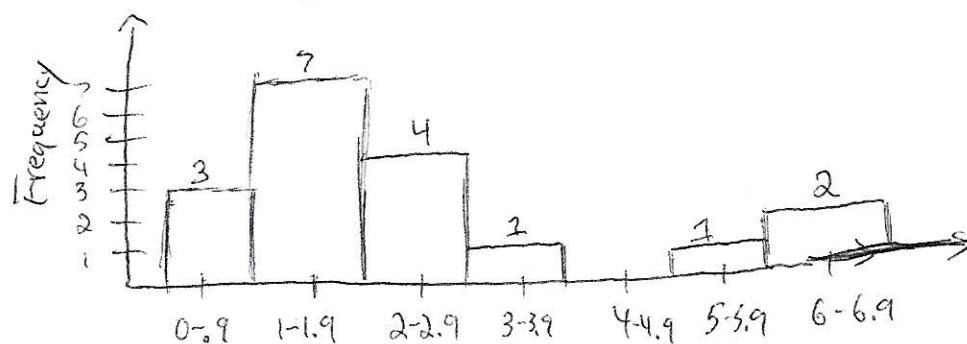
a) Frequency Table

0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5
1	2	2	5	2	2	1	0	0	0	1	0	1	1

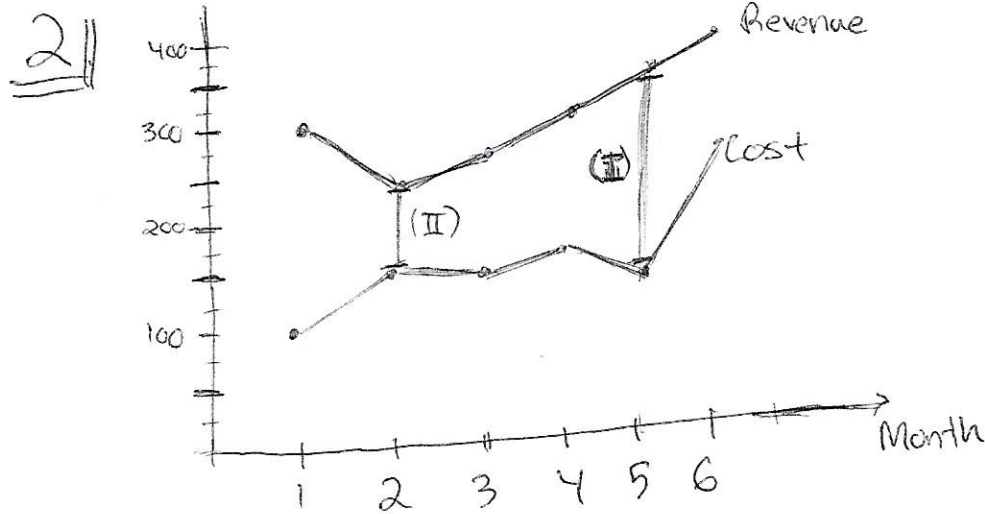
b) Bar graph



c) Histogram



- d) This was a subjective question so full points were given provided you offered a logical justification
- I would use a Bar graph since all the data is spread out by exactly 0.5 units — giving the data a discrete "feel".

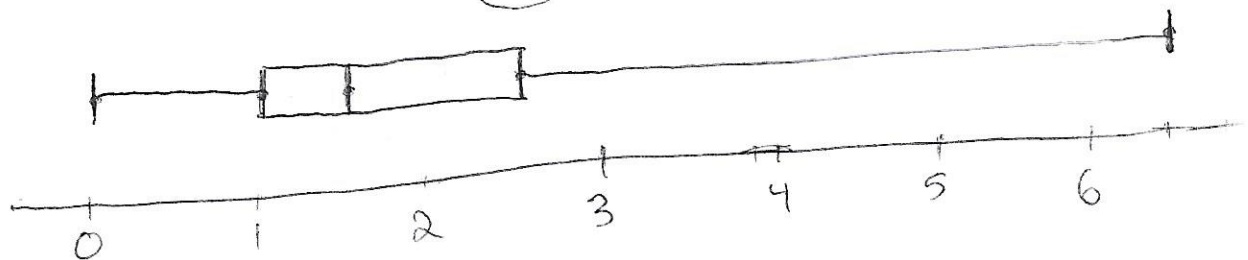


$$\text{Profit} = \text{Revenue} - \text{Cost}$$

- The most profit was made in month 5 (indicated by I being the largest gap)
- The least profit was made in month 2 (indicated by II being the smallest gap)

3 a)

Min	Q_1	Median	Q_3	Max
0	$L = \frac{25}{100} \cdot 18 = 4.5$	$L = \frac{50}{100} \cdot 18 = 9$	$L = \frac{75}{100} \cdot 18 = 13.5$	6.5
	$Q_1 = d_5 = 1$	$M = \frac{d_9 + d_{10}}{2} = 1.5$	$Q_3 = d_{14} = 2.5$	



b) Range = $6.5 - 0 = 6.5$, IQR = $2.5 - 1 = 1.5$

c) $sd^2 = \frac{1}{18} ((0-2.22)^2 + 2(0.5-2.22)^2 + 2(1-2.22)^2 + 5(1.5-2.22)^2 + 2(2-2.22)^2 + 2(2.5-2.22)^2 + (3-2.22)^2 + (5-2.22)^2 + (6-2.22)^2 + (6.5-2.22)^2)$

mean = 2.22

= 3.20

$sd = 1.789$

4 a) Recall $Q_1 = 1$, $Q_3 = 2.5$.

The outliers are the datapoints ≤ 1 or ≥ 2.5
(note that there are other definitions for outliers
such as $\leq Q_1 - 1.5 \cdot IQR$ or $\geq Q_3 + 1.5 \cdot IQR$,
but this is a simplified version).

Thus, the outliers are

$$\{0, 0.5, 0.5, 1, 1, 2.5, 2.5, 3, 5, 6, 6.5\}.$$

b) New dataset is $\{1.5, 1.5, 1.5, 1.5, 1.5, 2, 2\}$

$$\frac{\text{Min}}{1.5}$$

$$\frac{Q_1}{L = \frac{25}{100} \cdot 7 = 1.75}$$

$$Q_1 = d_2 = 1.5$$

$$\frac{\text{Median}}{L = \frac{50}{100} \cdot 7 = 3.5}$$

$$M = d_4 = 1.5$$

$$\frac{Q_3}{L = \frac{75}{100} \cdot 7 = 5.25}$$

$$Q_3 = d_6 = 2$$

$$\frac{\text{Max}}{2}$$

c) mean = 1.64

$$sd^2 = \frac{1}{7} (5(1.5 - 1.64)^2 + 2(2 - 1.64)^2) = 0.051$$

$$\boxed{sd = 0.226}$$

The standard deviation is far smaller than with outliers included. This makes sense, removing "far-out" datapoints reduces the spread.