Answers to the first Exam

1. Consider the following data, describing the number of wins for each of the 33 teams of a baseball league.

1	2	3	3	4	4
5	5	5	6	6	7
7	7	8	8	9	9
9	9	10	10	10	10
10	11	11	11	11	12
12	12	13			

(a) Fill in the following frequency table (use five classes)

Answer. The range is $13 - 1 = 12$ and we have $n = 5$ classes. So the class width will be the smallest integer larger than $\frac{12}{5} = 2.4$. Thus the class width is $w = 3$.

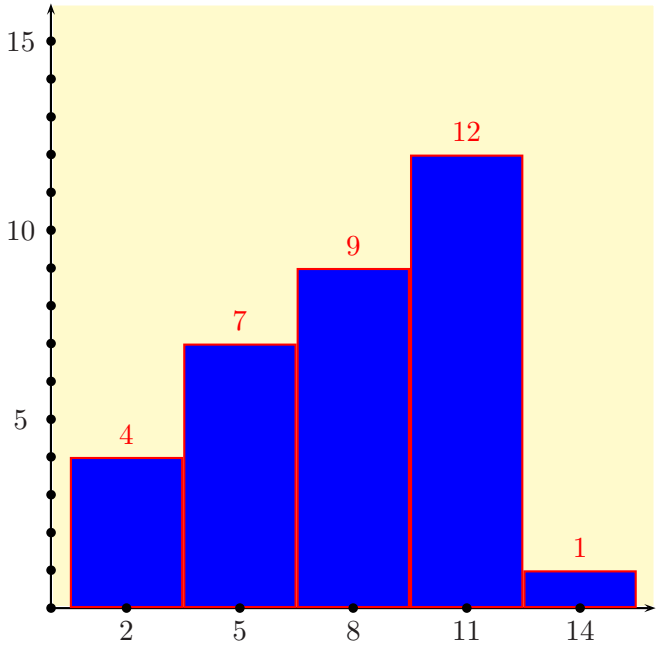
So we have the following table:

Classes	Class Boundaries	f	Class Midpoints
1 – 3	.5 – 3.5	4	2
4 – 6	3.5 – 6.5	7	5
7 – 9	6.5 – 9.5	9	8
10 – 12	9.5 – 12.5	12	11
13 – 15	12.5 – 15.5	1	14

□

(b) Make a histogram from the data in the first part:

Answer. We have the following histogram:



□

2. Find the mean, the range, the variance, and the standard deviation of the following sample data. Round your answers to two decimal digits:

$$11 \quad 10 \quad 8 \quad 4 \quad 6 \quad 7 \quad 11 \quad 6 \quad 11 \quad 7$$

You may use the following table. **You may use any method you want in your calculations.**

Answer. We have:

x	x^2
11	121
10	100
8	64
4	16
6	36
7	49
11	121
6	36
11	121
7	49
81	713

$$\text{Range} = 11 - 4 = 7$$

We can then calculate the mean

$$\bar{x} = \frac{\sum x}{n} = \frac{81}{10} = 8.1$$

the variance

$$s^2 = \frac{\sum x^2 - (\sum x)^2 / n}{n-1} = \frac{713 - 656.1}{9} \approx 6.32$$

and the standard deviation

$$s = \sqrt{s^2} \approx 2.52$$

□

3. The average score in an exam was 82 with a standard deviation of 5.6. Find the range of scores which contains at least 88.9% of the scores.

Answer. By Chebyshev's theorem 88.9% of the scores will lie within three standard deviations from the mean. It follows that at least 88.9% of the scores will lie in the range $[82 - 3 \cdot 5.6, 82 + 3 \cdot 5.6]$ or

$$[65.2, 98.8]$$

□

4. Consider the following data:

$$4 \quad 7 \quad 7 \quad 5 \quad 2 \quad 9 \quad 7 \quad 6 \quad 8 \quad 4 \quad 1 \quad 5 \quad 2 \quad 8 \quad 7 \quad 6 \quad 6 \quad 9$$

(a) Find the mode, the median, the first and third quartiles and the interquartile range.

Answer. We first sort the data from the smallest to the largest:

$$1 \quad 2 \quad 2 \quad 4 \quad 4 \quad 5 \quad 5 \quad 6 \quad 6 \quad 6 \quad 7 \quad 7 \quad 7 \quad 7 \quad 8 \quad 8 \quad 9 \quad 9$$

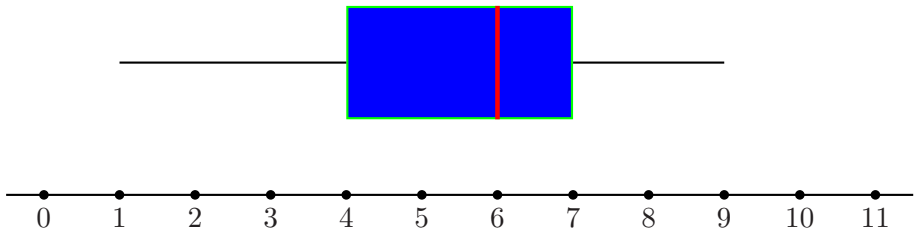
By inspection we see that the mode is 7. There are $n = 18$ values. Therefore the median Q_2 will be the mean of the ninth and tenth value. Therefore $Q_2 = 6$. The first quartile will then be the median of the first nine values, so we have $Q_1 = 4$ and the third quartile will be the median of the last nine values, so we have that $Q_3 = 7$. The lowest value is $Q_0 = 1$ and the highest value is $Q_4 = 9$. In sum we have the following five-number description of the data:

$$\begin{aligned} Q_0 &= 1 \\ Q_1 &= 4 \\ Q_2 &= 6 \\ Q_3 &= 7 \\ Q_4 &= 9 \end{aligned}$$

Finally the interquartile range is $7 - 4 = 3$. □

(b) Draw a box and whisker plot of the above data.

Answer. We have the following plot:



□

5. The following set of paired data represents x, the height in feet, and y, diameter in inches of five trees. Find the correlation coefficient r and the equation of the regression line.

Answer. We have:

x	y	x^2	y^2	xy
70	8.3	4900	68.89	581.
72	10.5	5184	110.25	756.
75	11.0	5625	121.	825.
76	11.4	5776	129.96	866.4
85	12.9	7225	166.41	1096.5
378	54.1	28710	596.51	4124.9

For the correlation coefficient we have:

$$\begin{aligned} r &= \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} \\ &= \frac{5 \cdot 4124.9 - 378 \cdot 54.1}{\sqrt{5 \cdot 28710 - 378^2}\sqrt{5 \cdot 596.51 - 54.1^2}} \\ &= \frac{174.7}{\sqrt{666}\sqrt{55.74}} \\ &\approx 0.91 \end{aligned}$$

To find the regression line we first calculate the slope b:

$$\begin{aligned} b &= \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} \\ &= \frac{174.7}{666} \\ &\approx 0.26 \end{aligned}$$

To find the intercept a we need the mean values of x and y:

$$\bar{x} = \frac{\sum x}{n} = \frac{378}{5} = 75.6$$

and

$$\bar{y} = \frac{\sum y}{n} = \frac{54.1}{5} = 10.82$$

Therefore the intercept is:

$$a = \bar{y} - b\bar{x} \approx 10.82 - 0.26 \cdot 75.6 = -8.84$$

Thus the equation of the regression line is:

$$\hat{y} = 0.26x - 8.84$$

□