The answers to the midterm exam

Nikos Apostolakis

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

$$84 \quad 12 \quad 27 \quad 15 \quad 40 \quad 18 \quad 33 \quad 33 \quad 14 \quad 4$$

   You may use the following table:

x	$x - \bar{x}$	$(x - \bar{x})^2$
84	56	3136
12	-16	256
27	-1	1
15	-13	169
40	12	144
18	-10	100
33	5	25
33	5	25
14	-14	196
4	-24	576
280	0	4628

   We have:

$$\text{Range} = 84 - 4 = 80$$

   We can then calculate the mean

$$\bar{x} = \frac{\sum x}{n} = \frac{280}{10} = 28$$

   the variance

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = \frac{4628}{9} = 514.22$$

   and the standard deviation

$$s = \sqrt{s^2} = 22.68$$

2. The mean price of houses in a certain neighborhood is \$175,000 and the standard deviation is \$38,000. Find the price range for which at least 75% of the houses will sell.

   Answer. According to Chebyshev's theorem 75% of the houses will sell at a price within two standard deviations from the mean. So we get the price range

$$[\mu - 2\sigma, \mu + 2\sigma] = [\$99000, \$251000]$$

□

3. In a survey of 20 patients who smoked the following data was obtained. Each value represents the number of cigarettes the patient smoked per day.

10	8	6	14
22	13	17	19
11	9	18	14
13	12	15	15
5	11	16	11

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

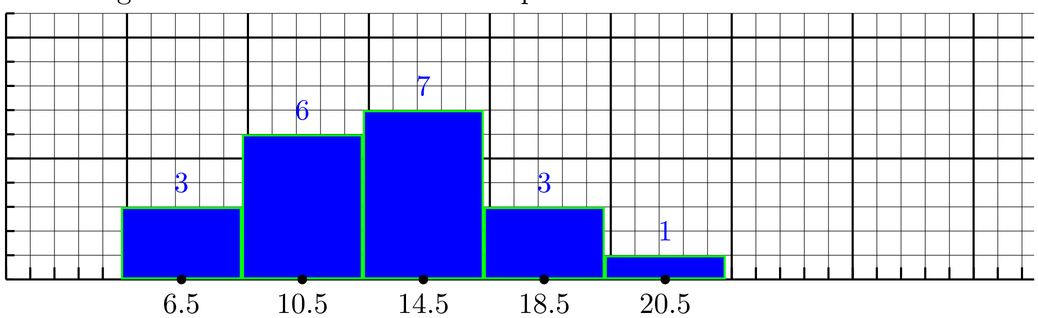
Classes	**Class Boundaries**	**Tally**	**Class Midpoints**
5 – 8	4.5 – 8.5	3	6.5
9 – 12	8.5 – 12.5	6	10.5
13 – 16	12.5 – 16.5	7	14.5
17 – 20	16.5 – 20.5	3	18.5
21 – 24	20.5 – 24.5	1	22.5

The smallest value is 5 and the larger is 22 so that the class width has to be the smaller integer larger than

$$\frac{22-5}{5} = 3.4$$

so the class width is $w = 4$

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



4. A store manager recorded the number of customers that enter the store between 12:00 PM and 2:00 PM for a 14-day period. The data are shown bellow.

33 38 43 30 29 40 51 27 42 23 31 25 35 30

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

Answer. We first have to order the data:

23 25 27 29 30 30 31 | 33 35 38 40 42 43 51

We have an even nuber of data values so the median will be the mean of the two middle values. So we have:

$$Q_2 = \frac{31+33}{2} = 32$$

The first quartile is then the median of the first half of the values so

$$Q_1 = 29$$

and the third quartile is the median of the second half of the values so

$$Q_3 = 40.$$

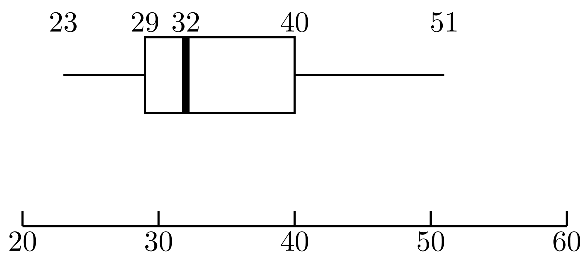
We can then calcualat the interquartile range:

$$IQR = Q_3 - Q_1 = 11$$

Finally the mode is 30. □

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

Answer.



□

5. The following set of paired data represents the number of hours that a student studied for a statistics exam x and the grade y that they received in the exam, for a random sample of 6 students. Find the correlation coefficient r.

x	y	x^2	y^2	xy
6	82	36	6724	492
2	63	4	3969	126
1	57	1	3249	57
5	88	25	7744	440
2	68	4	4624	136
3	75	9	5625	225
19	433	79	31935	1476

So we can calculate:

$$\begin{aligned} SS_x &= \sum x^2 - \frac{\left(\sum x\right)^2}{n} \\ &= 79 - \frac{19^2}{6} \\ &= 18.8333333333 \end{aligned}$$

$$\begin{aligned} SS_y &= \sum y^2 - \frac{\left(\sum y\right)^2}{n} \\ &= 31935 - \frac{433^2}{6} \\ &= 686.8333333 \end{aligned}$$

$$\begin{aligned} SS_{xy} &= \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n} \\ &= 1476 - \frac{19 \cdot 433}{6} \\ &= 104.83333333 \end{aligned}$$

So that

$$\begin{aligned} r &= \frac{SS_{xy}}{\sqrt{SS_x}\sqrt{SS_y}} \\ &= \frac{104.83333333}{\sqrt{18.8333333333}\sqrt{686.8333333}} \\ &\approx 0.92 \end{aligned}$$