Chapter 1: Why is my evil lecturer forcing me to learn statistics?

Labcoat Leni's Real Research

Is Friday the 13th Unlucky?

Problem

Scanlon, T. J., et al. (1993). *British Medical Journal, 307*, 1584–1586.

Many of us are superstitious, and a common superstition is that Friday the 13th is unlucky. Most of us don't literally think that someone in a hockey mask is going to kill us, but many people are wary. Scanlon and colleagues, in a tongue-in-cheek study (Scanlon, Luben, Scanlon, & Singleton, 1993), looked at accident statistics at hospitals in the South West Thames region of the UK. They took statistics both for Friday the 13th and Friday the 6th (the week before) in different months in 1989, 1990, 1991 and 1992. They looked at both emergency admissions of accidents and poisoning, and also transport accidents.

Date	Accidents and Poisoning		Traffic Accidents	
	Friday 6th	Friday 13th	Friday 6th	Friday 13th
October 1989	4	7	9	13
July 1990	6	6	6	12
September 1991	1	5	11	14
December 1991	9	5	11	10
March 1992	9	7	3	4
November 1992	1	6	5	12

Calculate the mean, median, standard deviation and interquartile range for each type of accident and on each date.

Solution

Let's begin with accidents and poisoning on Friday the 6th.

First, arrange the scores in ascending order: 1, 1, 4, 6, 9, 9.

The median: The median will be the (n + 1)/2th score. There are 6 scores, so this will be the 7/2 = 3.5th. The 3.5th score in our ordered list is half way between the 3rd and 4th scores which is (4+6)/2= 5 accidents.

The mean: The mean is 5 accidents:

$$\bar{x} = \frac{\sum x_i}{n}$$
$$= \frac{1 + 1 + 4 + 6 + 9 + 9}{6}$$
$$= \frac{30}{6}$$
$$= 5$$

The lower quartile: This is the median of the lower half of scores. If we split the data in half, there will be 3 scores in the bottom half (lowest scores) and 3 in the top half (highest scores). The median of the bottom half will be the (3+1)/2 = 2nd score below the mean. Therefore, the lower quartile is 1 accident.

The upper quartile: This is the median of the upper half of scores. If we again split the data in half and take the highest 3 scores, the median will be the (3+1)/2 = 2nd score above the mean. Therefore, the upper quartile is 9 accidents.

The interquartile range: This is the difference between the upper and lower quartiles: 9 – 1 = 8 accidents.

To calculate the sum of squares, first take the mean from each score, then square this difference, and finally, add up these squared values:

Score	Error (Score – Mean)	Error Squared
1	–4	16
1	–4	16
4	–1	1
6	1	1
9	4	16
9	4	16

So, the sum of squared errors is: 16 + 16 + 1 + 1 + 16 + 16 = 66.

The variance is the sum of squared errors divided by the degrees of freedom ($N - 1$):

$$s^2 = \frac{\text{sum of squares}}{n - 1} = \frac{66}{5} = 13.20$$

The standard deviation is the square root of the variance:

$$s = \sqrt{\text{variance}} = \sqrt{13.20} = 3.63$$

Next let's look at accidents and poisoning on Friday the 13th.

First, arrange the scores in ascending order: 5, 5, 6, 6, 7, 7.

The median: The median will be the (n + 1)/2th score. There are 6 scores, so this will be the 7/2 = 3.5th. The 3.5th score in our ordered list is half way between the 3rd and 4th scores which is (6+6)/2 = 6 accidents.

The mean: The mean is 6 accidents:

$$\bar{x} = \frac{\sum x_i}{n}$$
$$= \frac{5 + 5 + 6 + 6 + 7 + 7}{6}$$
$$= \frac{36}{6}$$
$$= 6$$

The lower quartile: This is the median of the lower half of scores. If we split the data in half, there will be 3 scores in the bottom half (lowest scores) and 3 in the top half (highest scores). The median of the bottom half will be the (3+1)/2 = 2nd score below the mean. Therefore, the lower quartile is 5 accidents.

The upper quartile: This is the median of the upper half of scores. If we again split the data in half and take the highest 3 scores, the median will be the (3+1)/2 = 2nd score above the mean. Therefore, the upper quartile is 7 accidents.

The interquartile range: This is the difference between the upper and lower quartiles: 7 – 5 = 2 accidents.

To calculate the sum of squares, first take the mean from each score, then square this difference, finally, add up these squared values:

Score	Error (Score – Mean)	Error Squared
7	1	1
6	0	0
5	–1	1
5	–1	1
7	1	1
6	0	0

So, the sum of squared errors is: 1 + 0 + 1 + 1 + 1 + 0 = 4.

The variance is the sum of squared errors divided by the degrees of freedom ($N - 1$):

$$s^2 = \frac{\text{sum of squares}}{n - 1} = \frac{4}{5} = 0.8$$

The standard deviation is the square root of the variance:

$$s = \sqrt{\text{variance}} = \sqrt{0.8} = 0.894$$

Next, let's look at traffic accidents on Friday the 6th.

First, arrange the scores in ascending order: 3, 5, 6, 9, 11, 11.

The median: The median will be the (n + 1)/2th score. There are 6 scores, so this will be the 7/2 = 3.5th. The 3.5th score in our ordered list is half way between the 3rd and 4th scores. The 3rd score is 6 and the 4th score is 9. Therefore the 3.5th score is (6+9)/2 = 7.5 accidents.

The mean: The mean is 7.5 accidents:

$$\bar{x} = \frac{\sum x_i}{n}$$
$$= \frac{3 + 5 + 6 + 9 + 11 + 11}{6}$$
$$= \frac{45}{6}$$
$$= 7.5$$

The lower quartile: This is the median of the lower half of scores. If we split the data in half, there will be 3 scores in the bottom half (lowest scores) and 3 in the top half (highest scores). The median of the bottom half will be the (3+1)/2 = 2nd score below the mean. Therefore, the lower quartile is 5 accidents.

The upper quartile: This is the median of the upper half of scores. If we again split the data in half and take the highest 3 scores, the median will be the (3+1)/2 = 2nd score above the mean. Therefore, the upper quartile is 11 accidents.

The interquartile range: This is the difference between the upper and lower quartiles: 11 − 5 = 6 accidents.

To calculate the sum of squares, first take the mean from each score, then square this difference, finally, add up these squared values:

Score	Error (Score − Mean)	Error Squared
9	1.5	2.25
6	–1.5	2.25
11	3.5	12.25
11	3.5	12.25
3	–4.5	20.25
5	–2.5	6.25

So, the sum of squared errors is: 2.25 + 2.25 + 12.25 + 12.25 + 20.25 + 6.25 = 55.5.

The variance is the sum of squared errors divided by the degrees of freedom ($N - 1$):

$$s^2 = \frac{\text{sum of squares}}{n - 1} = \frac{55.5}{5} = 11.10$$

The standard deviation is the square root of the variance:

$$s = \sqrt{\text{variance}} = \sqrt{11.10} = 3.33$$

Finally, let's look at traffic accidents on Friday the 13th.

First, arrange the scores in ascending order: 4, 10, 12, 12, 13, 14.

The median: The median will be the (n + 1)/2th score. There are 6 scores, so this will be the 7/2 = 3.5th. The 3.5th score in our ordered list is half way between the 3rd and 4th scores. The 3rd score is 12 and the 4th score is 12. Therefore the 3.5th score is (12+12)/2= 12 accidents.

The mean: The mean is 10.83 accidents:

$$\bar{x} = \frac{\sum x_i}{n}$$
$$= \frac{4 + 10 + 12 + 12 + 13 + 14}{6}$$
$$= \frac{65}{6}$$
$$= 10.83$$

The lower quartile: This is the median of the lower half of scores. If we split the data in half, there will be 3 scores in the bottom half (lowest scores) and 3 in the top half (highest scores). The median of the bottom half will be the (3+1)/2 = 2nd score below the mean. Therefore, the lower quartile is 10 accidents.

The upper quartile: This is the median of the upper half of scores. If we again split the data in half and take the highest 3 scores, the median will be the (3+1)/2 = 2nd score above the mean. Therefore, the upper quartile is 13 accidents.

The interquartile range: This is the difference between the upper and lower quartile: 13 − 10 = 3 accidents.

To calculate the sum of squares, first take the mean from each score, then square this difference, finally, add up these squared values:

Score	Error (Score − Mean)	Error Squared
4	−6.83	46.65
10	−0.83	0.69
12	1.17	1.37
12	1.17	1.37
13	2.17	4.71
14	3.17	10.05

So, the sum of squared errors is: 46.65 + 0.69 + 1.37 + 1.37 + 4.71 + 10.05 = 64.84.

The variance is the sum of squared errors divided by the degrees of freedom ($N - 1$):

$$\square^2 = \frac{\text{sum of squares}}{\square - 1} = \frac{64.84}{5} = 12.97$$

The standard deviation is the square root of the variance:

$$s = \sqrt{\text{variance}} = \sqrt{12.97} = 3.6$$