Chapter 6: Working with Differences

Extra questions

1. You have recently been given the expense claims of five sales representatives making the same journey:

£78.00	£56.84	£78.00	£64.76	£68.98

Determine the mean, the lowest values, the highest value, range and standard deviation.

2. A group of seven employees have been asked how much they typically spend each day in the canteen:

£2.40	£2.50	£0	£0	£2.50	£0	£1.60

Determine the mean, range and standard deviation. Comment on your results.

3. You have been given the following list of numbers.

18	20	19	19	21	20	20	22
21	20	18	21	22	19	20	21
20	21	20	18	20	21	18	19
22	18	21	19	20	21	22	20
19	22	19	20	21	21	19	20

Determine the mean standard deviation.

4. A supermarket has now decided to look at the amount requested in the form of 'cash back'.   They have the following records on the last 20 customers:

0	0	£30	0	0
0	0	0	0	£30
0	£40	0	0	0
0	0	0	£30	£40

How would you describe the differences in the data given in this table?

5. You have been given the following table giving the number of defective items found in boxes of components. Determine the mean and standard deviation.

Number of defective components	**Number of boxes**
0	60
1	4
2	2
3	1
4	1

6. A sample of motorists were asked what they had paid for a recent car service. Their replies were summarised in the following table

Cost of car service	**Number of motorists**
under £100	8
£100 but under £150	14
£150 but under £200	17
£200 but under £300	12
£300 and over	5

Determine the mean, median, quartiles and standard deviation.

7. You have been given the following survey results on the average age of customer for a particular product:

Age in years	**Number**
Under 20	4
20 but under 30	23
30 but under 40	45
40 but under 50	47
50 or more	1

Determine the mean and standard deviation.

8. Respond to the comment ‘that you don’t really need to work out the standard deviation as you can take the range and divide by 6 to get a reasonable approximation’.

Extra answers

1. Given below is an extract from a spreadsheet:

Expense claims £'s	**(*x*-mean)**	***(x-mean)***^{2}
78.00	8.684	75.4119
56.84	-12.476	155.6506
78.00	8.684	75.4119
64.76	-4.556	20.7571
68.98	-0.336	0.1129
346.58		327.3443
mean =	69.316	

The mean we have already calculated in the extra questions for chapter 5.

$$\bar{x} = \frac{78.00 + 56.84 + 78.00 + 64.76 + 68.98}{5} = \frac{346.58}{5} = £69.316$$

The lowest value is £56.84 and the highest value is £78.00. The range (the difference of these values is £21.16.

The standard deviation

$$= \sqrt{\frac{327.3441}{5}} = £8.091$$

2. Using a spreadsheet extract

Spend £'s	**(x-mean)**	**(x-mean)**^{2}
2.40	1.1143	1.24163
2.50	1.2143	1.47449
0.00	-1.2857	1.65306
0.00	-1.2857	1.65306
2.50	1.2143	1.47449
0.00	-1.2857	1.65306
1.60	0.3143	0.09878
9.00		9.24857
mean =	1.2857	
Standard deviation =		1.1494

The range = £2.50 – £0 = £2.50

Typically, measures of spread will be larger if you have the two clusters of number like this. There is more variation in the data overall because you have those that spend no money and then those that spend similar amounts of money.

3. Using a spreadsheet extract

Number	**Frequency** (f)	fx	(x - mean)	$(x - mean)^2$	$f(x - mean)^2$
18	5	90	-2.05	4.2025	21.0125
19	8	152	-1.05	1.1025	8.8200
20	12	240	-0.05	0.0025	0.0300
21	10	210	0.95	0.9025	9.0250
22	5	110	1.95	3.8025	19.0125
	40	802			57.9000

mean = 20.05

Standard Deviation = 1.2031

4. Including those that did not want cash back gives:

Cash back £'s	**Requests**	***fx***	**(*x*-mean)**	***(x-mean)*^{2}**	**f(x-mean)2**	**Cum Freq**
0	15	0	-8.500	72.2500	1083.7500	15
10	0	0	1.500	2.2500	0.0000	15
20	0	0	11.500	132.2500	0.0000	15
30	3	90	21.500	462.2500	1386.7500	18
40	2	80	31.500	992.2500	1984.5000	20
	20	170		1661.2500	4455.0000	

mean = 8.50

The standard deviation

$$= \sqrt{\frac{4455.00}{20}} = 14.9248$$

However, we would need to acknowledge that we have an increased variability because we have included those that want and do not want cash back.

If we considered only those 5 customers that wanted cash back:

£30 £30 £40 £30 £30

Then the mean is £32 and the standard deviation £4. Given data we can always calculate some statistics but what is in important is the clarification of purpose. We must be sure what situation we want to describe with these descriptive statistics.

5. Using a spreadsheet extract

Number of def. com (x)	**Number of boxes** (f)	fx	(x-mean)	$(x\text{-mean})^2$	$f(x\text{-mean})^2$	**Cum freq**
0	60	0	-0.2206	0.04866	2.9196	60
1	4	4	0.7794	0.60748	2.4299	64
2	2	4	1.7794	3.16631	6.3326	66
3	1	3	2.7794	7.72513	7.7251	67
4	1	4	3.7794	14.28395	14.2840	68
	68	15			33.6912	

mean = 0.2206

The standard deviation

$$= \sqrt{\frac{33.6912}{68}} = 0.7039$$

Given the majority of boxes had 0 defectives, the typical order statistics (the median and the quartiles) would all give 0.

6. Using a spreadsheet extract showing the columns needed for calculating the standard deviation for either of the formula given.

Cost of car service	**Number of motorists** (f)	**Mid-point** (x)	fx	(x - mean)	$(x - \text{mean})^2$	$f(x - \text{mean})^2$
under £100	8	50	400	-126.34	15961.615	127692.9209
£100 but under £150	14	125	1750	-51.34	2635.7223	36900.1116
£150 but under £200	17	175	2975	-1.34	1.7936862	30.4927
£200 but under £300	12	250	3000	73.66	5425.9008	65110.8099
£300 and over	5	350	1750	173.66	30158.044	150790.2184
	56		9875			380524.5536

mean = 176.34

Cost of car service	**Number of motorists** (f)	**Mid-point** (x)	fx	fx^2	**Cum Freq**
under £100	8	50	400	20000	8
£100 but under £150	14	125	1750	218750	22
£150 but under £200	17	175	2975	520625	39
£200 but under £300	12	250	3000	750000	51
£300 and over	5	350	1750	612500	56
	56		9875	2121875	

The standard deviation using

$$\sqrt{\frac{\sum f(x-\bar{x})^2}{n}}$$

$$\sqrt{\frac{380524.5536}{56}} = 82.4323$$

The standard deviation using

$$\sqrt{\left[\frac{\sum fx^2}{n} - \left(\frac{\sum x}{n}\right)^2\right]}$$

$$\sqrt{\left[\frac{2121875}{56} - \left(\frac{9875}{56}\right)^2\right]} = 82.4323$$

The median and the quartiles you can get directly from the ogive (plot cumulative frequency against the upper boundary of the interval groups). Here we just show the calculations:

$$median = 150 + 50\left(\frac{28-22}{17}\right) = £167.65$$

$$lower\ quartile = 100 + 50\left(\frac{14-8}{14}\right) = £121.43$$

$$upper\ quartile = 200 + 100\left(\frac{42-39}{12}\right) = £225.00$$

7. The calculations are shown for calculating the standard deviation using two formulae.

Age in years	**Number** (f)	**Mid-point** (x)	fx	(x - mean)	(x - mean)^2	f(x - mean)^2
under 20*	4	18.00	72.00	-18.60	345.96000	1383.84000

20 but under 30	23	25.00	575.00	-11.60	134.56000	3094.88000
30 but under 40	45	35.00	1575.00	-1.60	2.56000	115.20000
40 but under 50	47	45.00	2115.00	8.40	70.56000	3316.32000
50 or more*	1	55.00	55.00	18.40	338.56000	338.56000
	120		4392.00			8248.80000

mean = 36.60

Age in years	**Number** (f)	**Mid-point** (x)	fx	fx^2
under 20*	4	18.00	72	1296.000
20 but under 30	23	25.00	575	14375.000
30 but under 40	45	35.00	1575	55125.000
40 but under 50	47	45.00	2115	95175.000
50 or more*	1	55.00	55	3025.000
	120		4392	168996.000

* need to assume a mid-point value

$$\sqrt{\frac{\sum f(x-\bar{x})^2}{n}}$$

$$\sqrt{\frac{8248.8000}{120}} = 8.29$$

The standard deviation using

$$\sqrt{\left[\frac{\sum fx^2}{n} - \left(\frac{\sum x}{n}\right)^2\right]}$$

$$\sqrt{\left[\frac{168996}{120} - \left(\frac{4392}{120}\right)^2\right]} = 8.29$$

8. It is true that most data you will get a reasonable approximation to the standard deviation by dividing the range by 4, 5 or 6. The question remains on what is reasonable and to what extent your data allows this method of approximation to work. If you do use the range, you are working with just two observations (the highest and lowest) and not using all the information available from the other observations. These observations are also the most extreme and can be the most error prone. Whether the approximation is reasonable depends on the distribution of the data (these ideas are developed when you look at the normal distribution).

However, it is useful to have some idea what the value of the standard deviation should be. This will allow you to look at your calculated answers or those produced by computer software, and make a judgment as to whether they are correct or not. Given the availability of calculators to do these sums and software like Excel, the expectation is that you should have a means of calculation.