Descriptive statistics of Excel Data Analysis Tools

It is convenient for us to use Excel to analyze our data. Indeed, Excel comes equipped with a Descriptive Statistics tool in the Data Analysis add-in package, termed Analysis ToolPak or ATP. With this tool, we get as many as 16 different descriptive statistical parameters without having to enter a single function on the worksheet. See the figure below.

	A	B	C	D	E	F
1	Data			*Data*		
2	54.8					
3	53.6			Mean	54.44166667	
4	55.5			Standard Error	0.218277809	
5	54.2			Median	54.3	
6	54.9			Mode	53.6	
7	53.6			Standard Deviation	0.756136512	
8	55.2			Sample Variance	0.571742424	
9	54.2			Kurtosis	-1.275987593	
10	55.6			Skewness	0.203888076	
11	53.4			Range	2.2	
12	53.9			Minimum	53.4	
13	54.4			Maximum	55.6	
14				Sum	653.3	
15				Count	12	
16				Largest(1)	55.6	
17				Smallest(1)	53.4	
18				Confidence Level(95.0%)	0.480426219	
19						
20						
21						
22						
23						

Sheet1

Ready

Most of us are quite familiar with most of the descriptive statistics listed in Figure 1 except probably the terms “skewness” and “Kurtosis”. Actually, both of these parameters give us an insight into the shape of data distribution. Let’s take a closer look at them to understand how their figures come about.

Skewness

Skewness is a measure of the symmetry in a distribution of data. An *asymmetric* distribution is said to be *skewed*. A simpler formula of calculating skewness is the average of the cubed estimated standard scores (i.e., z-scores):

$$\text{Skewness} = \sum_{i=1}^{n} \frac{{z_i}^3}{n} \qquad \text{.... [1]}$$

where $z_i = \frac{x_i - \bar{x}}{s}$

Upon calculation by equation [1]. skewness = 0.156.

However, the formula that Excel uses is a Fisher's skewness formula, which approaches the value of the average cubed z-scores as n increases:

$$\text{Skewness} = n \sum_{i=1}^{n} \frac{z^3}{(n-1)(n-2)} \qquad \text{.... [2]}$$

By using the equation [2], skewness = 0.204. A symmetric distribution is expected to have a skewness of zero (0).

Kurtosis

Kurtosis measures the degree to which a distribution of scores is taller or flatter with respect to its width. In other words, it is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution.

A simple definition of kurtosis is as shown in equation [3]:

$$\text{Kurtosis} = \frac{\sum_{i=1}^{n} z^4}{n} - 3 \qquad \text{....[3]}$$

A normal distribution curve would have an average z-score, raised to the 4th power, of 3. Therefore, 3 is subtracted in the formula so that a normal curve would have kurtosis of zero (0). If kurtosis by calculation is found to be positive, it implies the curve is narrow with fat tails, whilst negative kurtosis indicates the curve being broad with 'thin' tails.

By using the above equation, kurtosis = -1.542.

Again, Excel's formula for kurtosis (Fisher) is slightly different and it attempts to remove bias from the calculation of kurtosis based on a sample:

$$\text{Kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i-1}^{n} z^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \qquad \text{....[4]}$$

Using equation [4] gives kurtosis = -1.276. A perfect normal curve by this manner of calculation would have kurtosis of zero (0).