

Obtaining Descriptive Statistics in MS Excel

(1) If you do not see *Data Analysis* in the menu, this means you need to use the *Add-ins* and make sure that the box in front of Analysis ToolPak is checked.

(2) Select *Data Analysis > Descriptive Statistics*.

(3) Enter the *Input Range*, where the data is located, e.g., b4:b17.

Check the box in front of *Summary Statistics*.

You have to indicate where you want the output to appear. You will probably want the output to appear either on the same page or on another worksheet. If you want the output to appear on the same page, then check the circle in front of *Output Range* and indicate where the output should go. If the data appear in, say, b4: b17, then your output should not appear in the first 17 rows. If you check the circle in front of *New Worksheet Ply* then your output will appear on another worksheet. This may be a good idea if you are afraid that the output is too large to appear on the same page as the input data.

Example: A manager wants to know how long it takes to assemble a computer. She randomly selects 14 employees.

Times (in minutes):

100, 90, 45, 67, 80, 92, 70, 71, 77, 29, 89, 76, 80, 83

The data was input into cells b4 to b17 (there are 14 scores).

Here is the Excel output:

<i>Column1</i>	
Mean	74.92857143
Standard Error	5.013678308
Median	78.5
Mode	80
Standard Deviation	18.75946647
Sample Variance	351.9175824
Kurtosis	1.923164749
Skewness	-1.31355395
Range	71
Minimum	29
Maximum	100
Sum	1049
Count	14

- (a) *Mean*, \bar{X} , (average time) is 74.93 minutes (rounded).
- (b) *Standard Error* is $s / \sqrt{n} = 5.014$ (rounded). Divide the standard deviation by the square root of the sample size. This measure will be used in the second part of the course dealing with inference. Don't worry about it now.
- (c) *Median* is 78.5 minutes; half the employees did better than this and half did worse.
- (d) *Mode* is 80 minutes; two employees did the job in this time. All other scores had frequencies of 1.
- (e) *Standard Deviation* is a measure of dispersion. The standard deviation is 18.76 minutes (rounded). If the scores are normally distributed, then about 95.5% of scores should be within two standard deviations of the mean (see below). Another way of expressing this is that 95.5% of scores should be between the mean \pm two standard deviations (74.93 ± 37.52), or, 37.41 minutes and 112.45 minutes.
- (f) *Sample Variance* is the standard deviation squared.
- (g) *Kurtosis* is a measure of peakedness and is rarely used.
- (h) *Skewness* value is -1.31. If the data is symmetric, the value should be about 0. There is a negative skew to this data set: the mean is below the median.
- (i) *Range* is 71 minutes. The maximum value – minimum value, or 100 minutes – 29 minutes.
- (j) *Minimum* (lowest value) is 29 minutes. One employee was very quick and did the job in 29 minutes.
- (k) *Maximum* (highest value) is 100 minutes. This employee was relatively slow and did the job in 100 minutes.
- (l) *Sum* is the sum of all 14 observations, $\sum X_i$.
- (m) *Count* is the sample size.