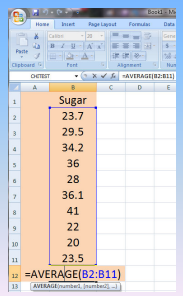
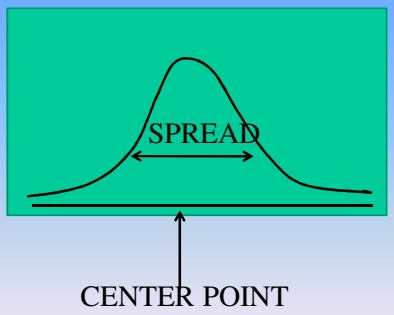
STATS 101 - Introductory Statistics

Lecture 2

Descriptive Statistics



NEED TO ESTIMATE TWO FEATURES OF A PROBABILITY DISTRIBUTION



SAMPLE MEAN is a measure of center point of data distribution

= total of all observations / n

Computing the SAMPLE MEAN using the STATISTICAL FUNCTION built in EXCEL

Go to a blank cell (say A(15)), and click on icon f_x, then choose FUNCTION CATEGORY *statistical*, click on FUNCTION NAME *average*, and hit OK.

BASIC EXCEL-continued

SAMPLE VARIANCE = measure of spread or variability in the sample = average squared distance of observations from the mean

1. Calculate distance of each observation in sample from the sample mean, and square each distance.
2. SAMPLE VARIANCE = sum of squared distance / (n-1)

NOTE: Dividing by (n-1) instead of n gives an UNBIASED ESTIMATE of the POPULATION VARIANCE.

This will pop-up another window. In the NUMBER 1 window, type in

A1:A10 (address of cells containing the given sample of size 10)

and hit OK. This will give the mean of 29.4 in cell A(15) – same as before.

As in the case of the SAMPLE MEAN, we will calculate the SAMPLE VARIANCE in two ways:

1. By using EXCEL worksheet functions.
2. By using STATISTICAL FUNCTIONS built in EXCEL.

Data for Example 1 – sugar content in 10 brands of candy bars:
23.7, 29.5, 34.2, 36, 28, 36.1, 41, 22, 20, 23.5

sample mean = $\bar{x} = 29.4$

The SAMPLE VARIANCE is defined as :

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1}$$

$x_i =$ $i-th$ sample value

$\bar{x} =$ sample mean

The numerator of s^2 is:

$$(23.7-29.4)^2 + (29.5-29.4)^2 +$$
$$(34.2-29.4)^2 + (36-29.4)^2 +$$
$$(28-29.4)^2 + (36.1-29.4)^2 +$$
$$(41-29.4)^2 + (22-29.4)^2 +$$
$$(20-29.4)^2 + (23.5-29.4)^2$$

To calculate the above expression, return to the EXCEL worksheet, and do the following:

1. In cell B(1), type

= A1-29.4

This will result in -5.7 in cell B(1).

2) Stay in cell B(1), hit COPY icon. Highlight cells B(2) to B(10), hit PASTE.

3) Go to cell C(1) and type:

= B(1)^2

This will result in 32.49 in cell C(1).

4) Go to cell C(1), hit COPY icon.

5) Highlight cells C(2) to (10), hit PASTE.
This will result in values of $(x\text{-mean})^2$ in Column C, as shown on the next slide.

Example 1

	A	B	C
1	23.70	-5.7	32.49
2	29.50	0.1	0.01
3	34.20	4.8	23.04
4	36.00	6.6	43.56
5	28.00	-1.4	1.96
6	36.10	6.7	44.89
7	41.00	11.6	134.56
8	22.00	-7.4	54.76
9	20.00	-9.4	88.36
10	23.50	-5.9	34.81

6) Go to cell C(11), and type

= sum(C1:C10)

This will give you the numerator of 458.44 in cell C(11).

7) Go to cell C(12), and type

= C11/9

(Note: n-1 = 9 for this example)

This will give you the sample variance of 50.93778 in cell C(12).

SAMPLE STANDARD DEVIATION (sd) =√(sample variance) = √(50.93778) = 7.137

Calculating the sample variance using the STATISTICAL FUNCTION

Go to a blank cell (say A(16)), and click on the f_x icon , then choose FUNCTION CATEGORY *statistical*, click on FUNCTION NAME *var (for variance) or stdev (for sd)*, and hit OK.

This will give you the sample variance of 50.93778 in the cell A(16) or sd of 7.137 – same as before.