MAT 1372 Statistics with Probability classwk 4 Fall 2012

# Measures of spread or variability and empirical rule

## 3.5 Sample variance = (sum of the squares of the deviations)/(n-1)

In symbols:

 

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| val | dev | freq | dev square | prod of dev square & freq |
| -1 | -1.2 | 3 | 1.44 | 4.32 |
| 0 | -.2 | 2 | .04 | .08 |
| 1 | .8 | 5 | .64 | 3.2 |
|  |  | sum |  | 7.6 |

To get the sample variance s2, we divide this sum by n-1=10-1=9 to get .84.

Exercise: using the hmwk 2 excel file in the openlab and the definition find the sample variance for the data sets of problems 2.2.1, 2.2.2, 2.3.1, 2.3.3

**11.** Compute the sample variance of the following data sets by hand:

**(a)** 1, 2, 3, 4, 5

**(b)** 6, 7, 8, 9, 10

**(c)** 11, 12, 13, 14, 15

**(d)** 2, 4, 6, 8, 10

**(e)** 10, 20, 30, 40, 50

**16.** Find the sample standard deviation of the data set given by the following frequency table:

**Val Freq Val Freq**

3 1 5 3

4 2 6 2

**17.** The following data represent the acidity of 40 successive rainfalls in the state of Minnesota. The acidity is measured on a pH scale, which varies from 1 (very acidic) to 7 (neutral).

3.71, 4.23, 4.16, 2.98, 3.23, 4.67, 3.99, 5.04, 4.55, 3.24, 2.80, 3.44,

3.27, 2.66, 2.95, 4.70, 5.12, 3.77, 3.12, 2.38, 4.57, 3.88, 2.97, 3.70, 2.53, 2.67,

4.12, 4.80, 3.55, 3.86, 2.51, 3.33, 3.85, 2.35, 3.12, 4.39, 5.09, 3.38, 2.73, 3.07

**(a)** Find the sample standard deviation.

**(b)** Find the range.

**(c)** Find the interquartile range.



**3.6 NORMAL DATA SETS AND THE EMPIRICAL RULE**

Many of the large data sets one encounters in practice have histograms that are similar in shape. These histograms are often symmetric about their point of highest frequency and then decrease on both sides of this point in a bell-shaped fashion. Such data sets are said to be *normal*, and their histograms are called *normal* *histograms*.

**Definition** *A data set is said to be* normal *if a histogram describing it has the following properties:*

**1.** *It is highest at the middle interval.*

**2.** *Moving from the middle interval in either direction, the height decreases in such a way that the entire histogram is bell-shaped.*

**3.** *The histogram is symmetric about its middle interval.*

Figure 3.2 shows the histogram of a normal data set.





It follows from the symmetry of the normal histogram that a data set that is approximately normal will have its sample mean and sample median approximately equal.

Suppose that $\overbar{x}$and *s* are the sample mean and sample standard deviation, respectively, of an approximately normal data set. The following rule, known as the *empirical rule*, specifies the approximate proportions of the data observations that are within *s*, 2*s*, and 3*s* of the sample mean $\overbar{x}$.



**3.5.17. (addition)** The following data represent the acidity of 40 successive rainfalls in the state of Minnesota. The acidity is measured on a pH scale, which varies from 1 (very acidic) to 7 (neutral).

3.71, 4.23, 4.16, 2.98, 3.23, 4.67, 3.99, 5.04, 4.55, 3.24, 2.80, 3.44,

3.27, 2.66, 2.95, 4.70, 5.12, 3.77, 3.12, 2.38, 4.57, 3.88, 2.97, 3.70, 2.53, 2.67,

4.12, 4.80, 3.55, 3.86, 2.51, 3.33, 3.85, 2.35, 3.12, 4.39, 5.09, 3.38, 2.73, 3.07

**(a)** Draw a histogram for the given data.

**(b)** Do the data appear to be approximately normal?

**(c)** How do the mean and median compare? Does this verify your answer to (b)?

**(d)** Pretend that the answer to (b) is yes, give an interval that you would expect to contain approximately 95 percent of the data observations.

**(e)** What percentage of the data lies in the interval given in part (g)?