Measure of Variation
Lecture 5

Lecture objectives

- Describe data, using measures of variation, such as the range, variance, and standard deviation

1 Measures of Variation

In statistics, to describe the data set accurately, statisticians must know more than the measures of central tendency.

Example 1. A testing lab wishes to test two experimental brands of outdoor paint to see how long each will last before fading. The testing lab makes 6 gallons of each paint to test. The results (in months) are shown. Find the mean of each group.

Brand A	Brand B
10	35
60	45
50	30
30	35
40	40
20	25

Solution

The mean for brand A is
$\mu = \frac{\sum X}{N} = 35$
The mean for brand B is
$\mu = \frac{\sum X}{N} = 35$
Even though the means are the same for both brands, the spread, or variation, is quite different. brand B performs more consistently;it is less variable than Brand A.

1.1 Range

The range is the highest value minus the lowest value. The symbol R is used for the range.
$R = Highest\ value - Lowest\ value$

Example 2. Find the ranges for the paints in Example

Solution

For brand A, the range is
$R = 60 - 10 = 50$ months
For brand B, the range is
$R = 45 - 25 = 20$ months

The range for brand A shows that 50 months separate the largest data value from the smallest data value. For brand B, 20 months separate the largest data value from the smallest data value.

Example 3. The salaries for the staff of the some company . are shown here. Find the range.

Staff	Salary
Owner	100,000
Manager	40,000
Sales representative	30,000
Workers	25,000
	15,000
	18,000

Solution

The range is $R = 100,000 - 15,000 = 85,000$.

One extremely high or one extremely low data value can affect the range markedly,to have a more meaningful statistic to measure the variability, statisticians use measures called the *variance* and *standard deviation*.

1.2 Population Variance and Standard Deviation

1.2.1 The variance

The variance is the average of the squares of the distance each value is from the mean. The symbol for the population variance is σ^2.

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

where
X=individual value
μ=population mean
N=population size

1.2.2 The standard deviation

The standard deviation is the square root of the variance. The symbol for the population standard deviation is s. The corresponding formula for the population standard deviation is
$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$

Example 4. Find the variance and standard deviation for the data set for brand A paint in previous Example.

Solution

First	Find the mean for the data. $\mu = 35$
Second	Subtract the mean from each data value
Third	Square each of the previous result.
Fourth	Find the sum of the squares.Then Divide the sum by N to get the variance
Fifth	Take the square root of the variance to get the standard deviation.

X	$(X-\mu)$	$(X-\mu)^2$
10	-25	625
60	25	625
50	15	225
30	-5	25
40	5	25
20	-15	225
		Total=1750

so the variance $\sigma^2 = \frac{1750}{6} = 291.7$
the standard deviation $s = \sqrt{291.7} = 17.1$

1.2.3 Assignment

Find the variance and standard deviation for brand B in the paint example and compare it with the Brand A's, and comment on the result

1.3 Sample Variance and Standard Deviation

The formula for the sample variance, denoted by s^2, is
$s^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$
where
n=sample size
$\overline{x}$=sample mean
and the sample standard deviation is $s = \sqrt{s^2}$.

The shortcut formulas for computing the variance and standard deviation for data obtained from samples are as follows.

Variance	standard deviation
$s^2 = \frac{n(\sum x^2)-(\sum x)^2}{n(n-1)}$	$s = \sqrt{\frac{n(\sum x^2)-(\sum x)^2}{n(n-1)}}$

1.4 Variance and Standard Deviation for Grouped Data

Example 5. Find the variance and the standard deviation for the frequency distribution of the following data.

Class	Frequency	Midpoint
5.5 – 10.5	1	8
10.5 – 15.5	2	13
15.5 – 20.5	3	18
20.5 – 25.5	5	23
25.5 – 30.5	4	28
30.5 – 35.5	3	33
35.5 – 40.5	2	38

Solution

First	Multiply the frequency by the midpoint for each class, and find it's sum
Second	Multiply the frequency by the square of the midpoint, and find it's sum
Third	Substitute in the formula and solve for s^2 to get the variance

Class	Frequency	Midpoint	$f * X_m$	$f * X_m^2$
5.5 – 10.5	1	8	8	64
10.5 – 15.5	2	13	26	338
15.5 – 20.5	3	18	54	972
20.5 – 25.5	5	23	115	2645
25.5 – 30.5	4	28	112	3136
30.5 – 35.5	3	33	99	3267
35.5 – 40.5	2	38	76	2888
	Total=20		$\sum fX_m = 490$	$\sum fX_m^2 = 13310$

$s^2 = \frac{n(\sum f.X_m^2)-(\sum f.X_m)^2}{n(n-1)}$

$s^2 = \frac{20(13310)-490^2}{20(20-1)} = 68.7$

to get the standard deviation take the square root of the variance

$s = \sqrt{68.7} = 8.3$

1.5 Uses of the Variance and Standard Deviation

- Variances and standard deviations can be used to determine the spread of the data. If the variance or standard deviation is large, the data are more dispersed. This information is useful in comparing two (or more) data sets to determine which is more (most) variable.
- The measures of variance and standard deviation are used to determine the consistency of a variable.
- Finally, the variance and standard deviation are used quite often in inferential statistics.

1.6 Coefficient of Variation

A statistic that allows you to compare standard deviations when the units are different.
The coefficient of variation, denoted by CVar, is the standard deviation divided by the mean. The result is expressed as a percentage.

Sample	Population
$CVar = \frac{s}{\bar{x}}.100$	$CVar = \frac{\sigma}{\mu}.100$

Example 6. The mean of the number of sales of cars over a 3-month period is 87, and the standard deviation is 5. The mean of the commissions is 5225, and the standard deviation is 773. Compare the variations of the two.

Solution

The coefficients of variation are.
$CVar = \frac{s}{\bar{x}}.100 = \frac{5}{87}.100 = 5.7\%$ sales
$CVar = \frac{s}{\bar{x}}.100 = \frac{773}{5225}.100 = 14.8\%$ commissions.

Since the coefficient of variation is larger for commissions, the commissions are more variable than the sales.

1.7 Range Rule of Thumb

The range can be used to approximate the standard deviation.the approximation is called **range rule of thumb**.

$$s \approx \frac{range}{4}$$

The range rule of thumb is only an approximation and should be used when the distribution of data values is unimodal and roughly symmetric,also the rule can be used to estimate the largest and the smallest values of a data set.the smallest data value will be approximately 2 standard deviation below the mean, and the largest data value will be approximately 2 standard deviations above the mean of the data set.