OKAN ÜNİVERSİTESİ
İSTANBUL

# BBA182 Applied Statistics
# Week 4 (2) Measures of variation

DR SUSANNE HANSEN SARAL

EMAIL: SUSANNE.SARAL@OKAN.EDU.TR

HTTPS://PIAZZA.COM/CLASS/IXRJ5MMOX1U2T8?CID=4#

WWW.KHANACADEMY.ORG

OKAN ÜNİVERSİTESİ
İSTANBUL

# ***Average*** **distance to the mean:**

## Standard deviation

Most **commonly used measure of variability**

Measures the standard (average) distance of all data points from the mean.

OKAN ÜNİVERSİTESİ
İSTANBUL

# Using Microsoft Excel

Descriptive Statistics can be obtained from Microsoft® Excel

- Select:

  data / data analysis / descriptive statistics

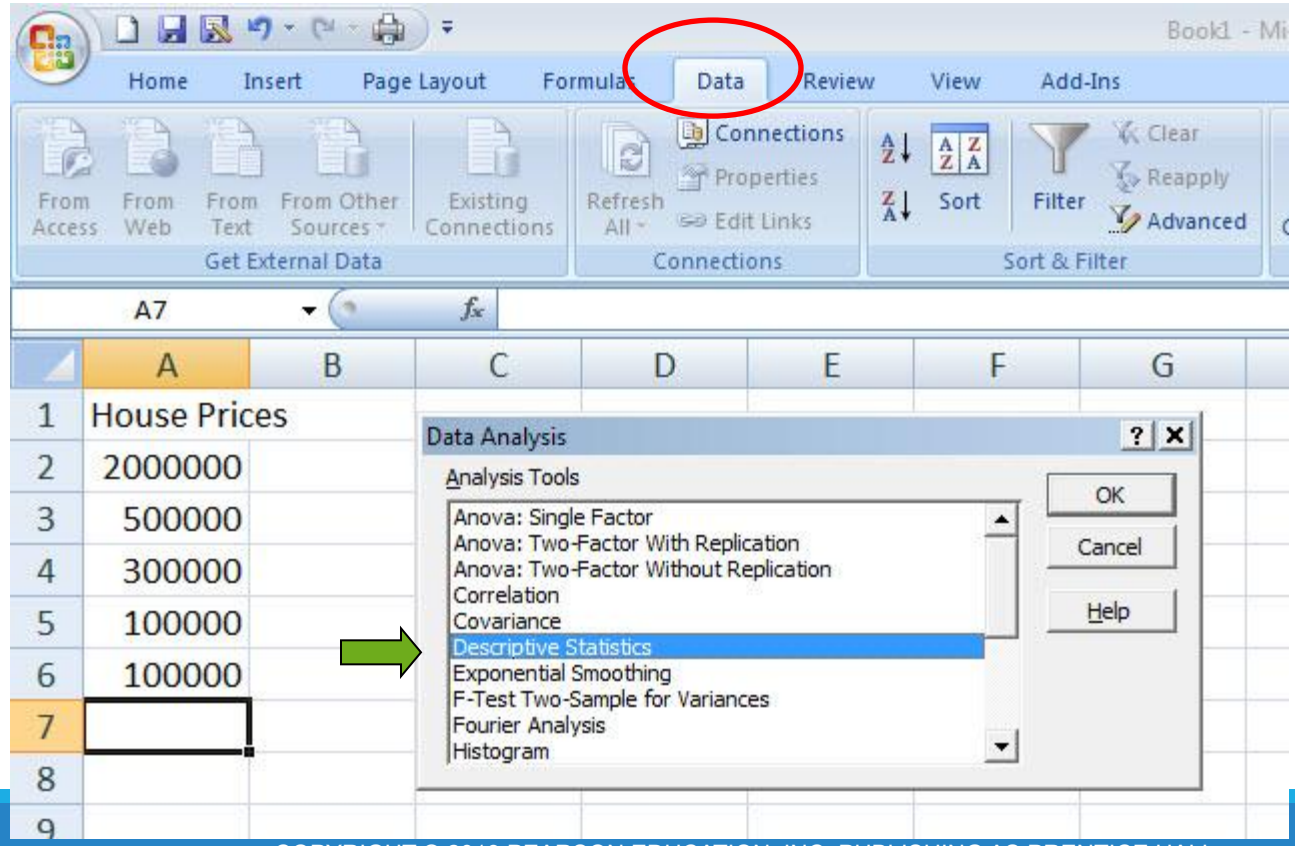
- Enter details in dialog box

COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

OKAN ÜNİVERSİTESİ
İSTANBUL

# sing Excel to find Descriptive Statistics

- Select data / data analysis / descriptive statistics




COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

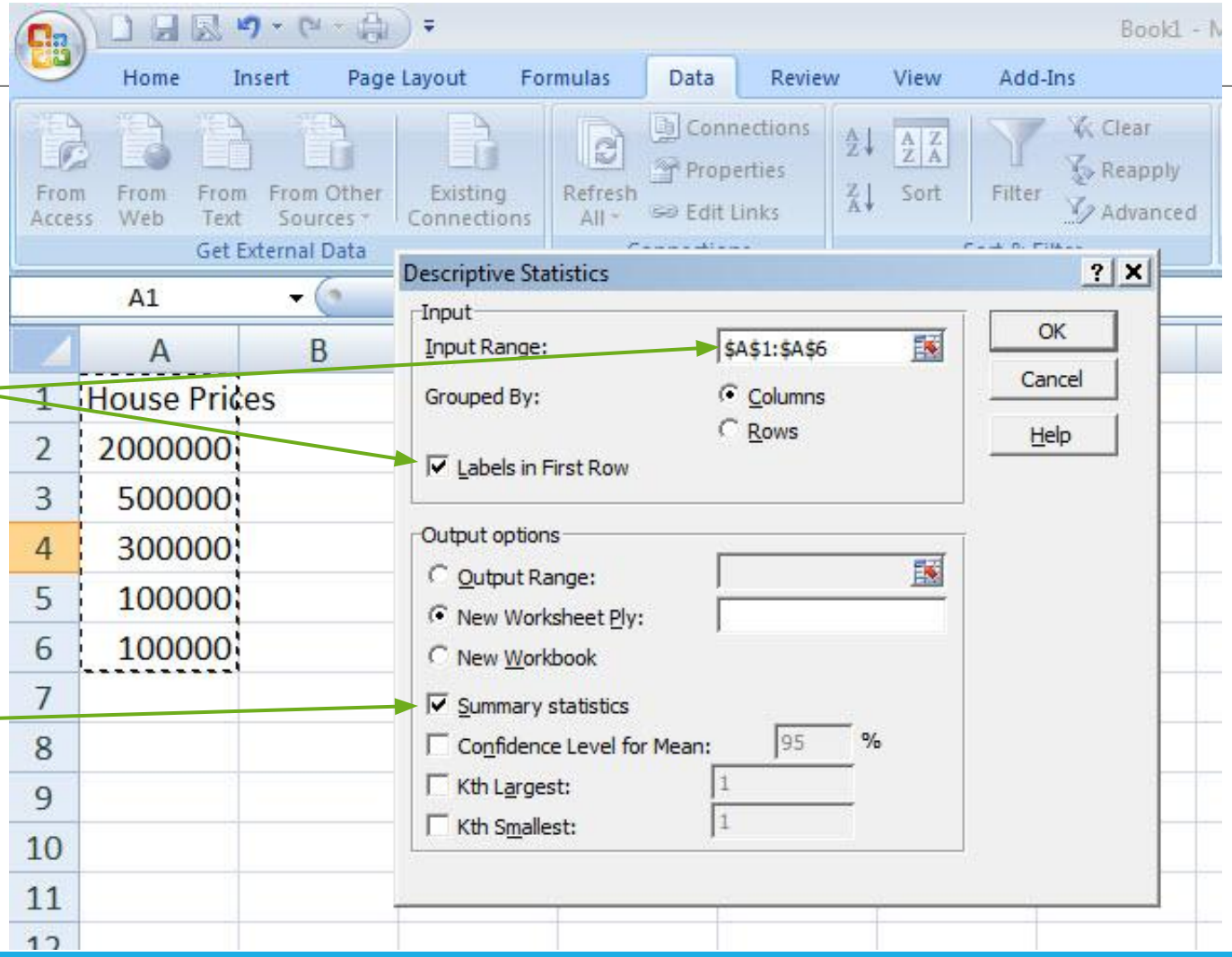
OKAN ÜNİVERSİTESİ
İSTANBUL

# Using Excel to find Descriptive Statistics

Enter input range details

Check box for summary statistics

Click OK




COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

# Excel output

**Microsoft Excel descriptive statistics output, using the house price data:**

House Prices:

$2,000,000
500,000
300,000
100,000
100,000

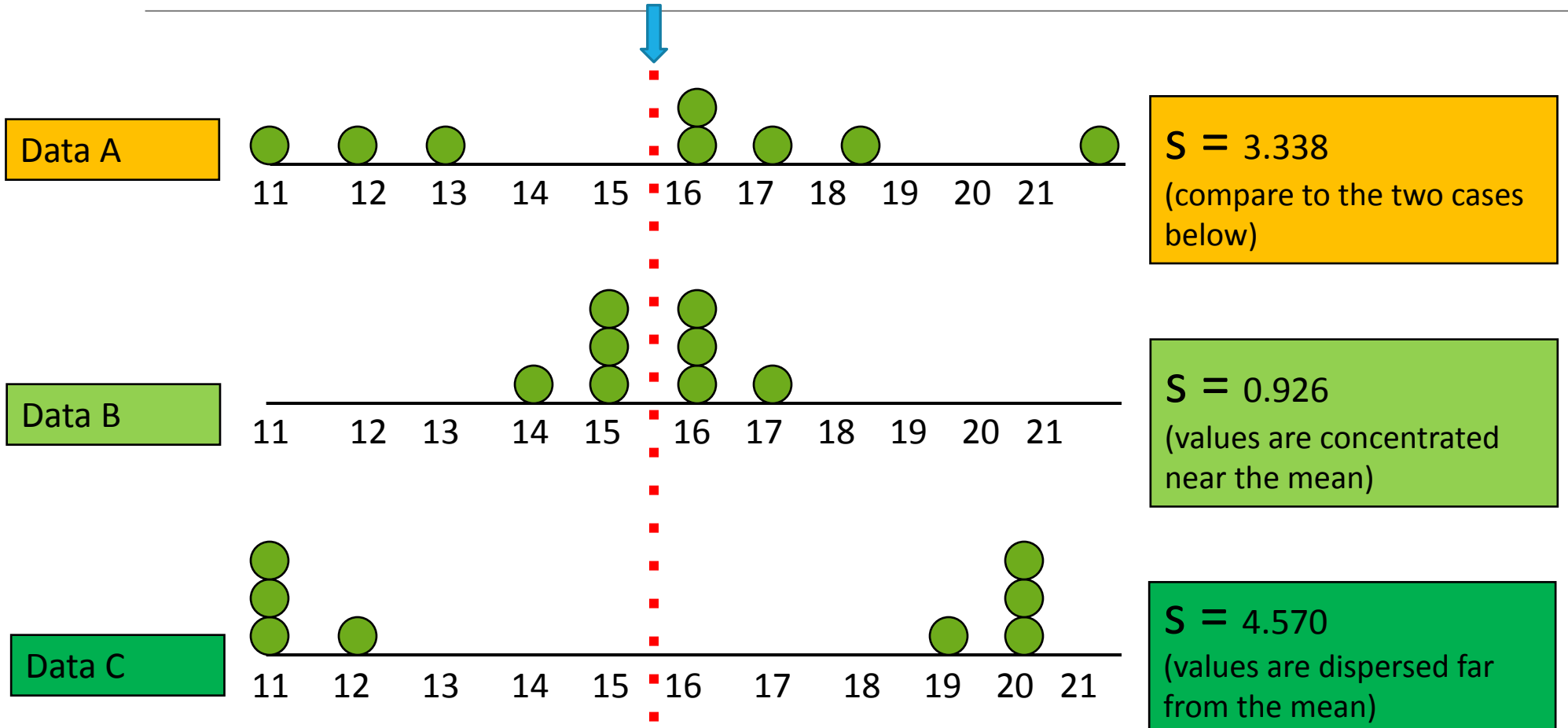
| | A | B |
|---|---|---|
| 1 | *House Prices* | |
| 2 | | |
| 3 | Mean | 600000 |
| 4 | Standard Error | 357770.8764 |
| 5 | Median | 300000 |
| 6 | Mode | 100000 |
| 7 | Standard Deviation | 800000 |
| 8 | Sample Variance | 6.4E+11 |
| 9 | Kurtosis | 4.130126953 |
| 10 | Skewness | 2.006835938 |
| 11 | Range | 1900000 |
| 12 | Minimum | 100000 |
| 13 | Maximum | 2000000 |
| 14 | Sum | 3000000 |
| 15 | Count | 5 |

COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

# Comparing Standard Deviations of 3 different data

OKAN ÜNİVERSİTESİ
İSTANBUL

**Mean = 15.5 for each data set**




COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

# Comparing Standard Deviations of 2 data sets

1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,5

1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,120

**Without calculating, which of the two data sets do you *expect* to have the highest variation and standard deviation? Why?**

# Describing distributions – what to pay attention to!

Pay attention to:

- ✔ its' shape (symmetric, right or left skewed)
- ✔ its' center (mean, median, mode)
- ✔ Its' spread (variance, standard deviation)

OKAN ÜNİVERSİTESİ
İSTANBUL

# Effect of the size of the standard on on the shape of a distribution

**The standard deviation affects the shape of a distribution:**

**When there are small** distances between the data points, most of the scores in the data set will be close to the mean and the resulting **standard deviation will be small**. The distribution will be narrow.

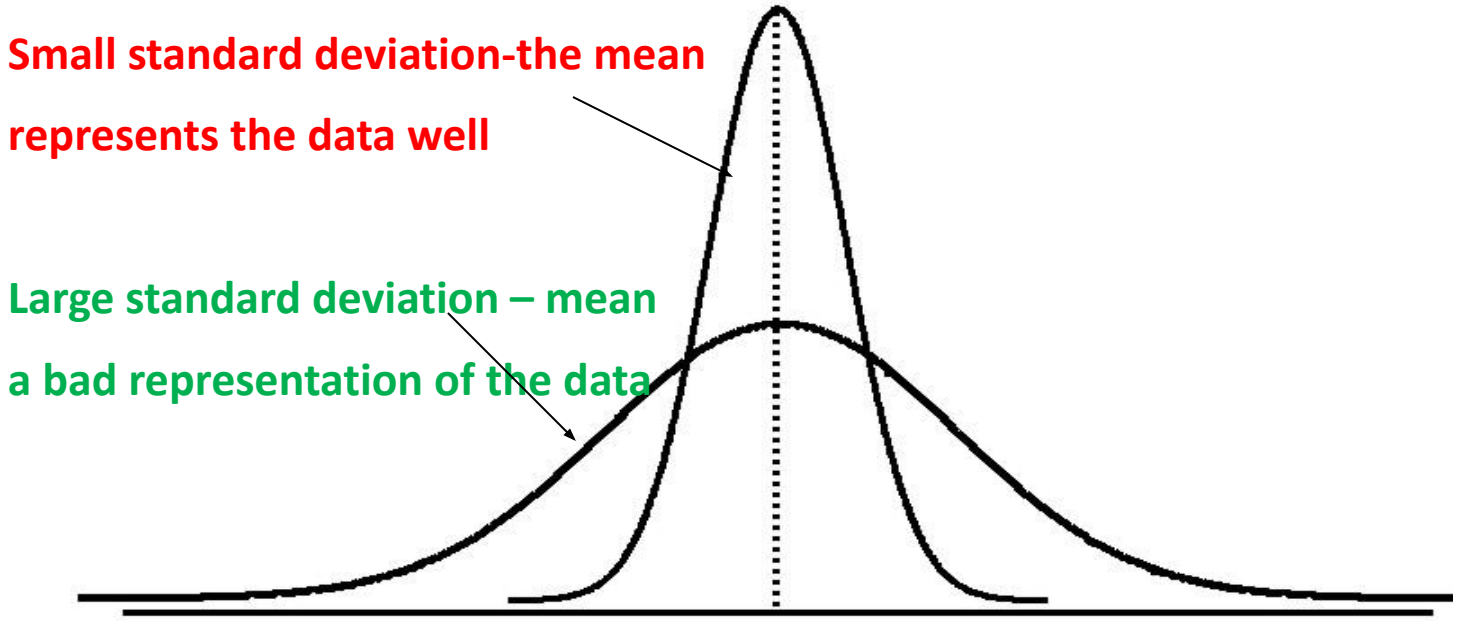
**When there are** large distances between data points, the scores will be further away from the mean and the **standard deviation is larger**. The distribution will be wide.

As illustrated in the following slide:

COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

OKAN ÜNİVERSİTESİ
İSTANBUL

# Effect of the size of the standard deviation on the shape of a distribution




COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

# Examples of applications of the standard deviation in business

**Logistics:**

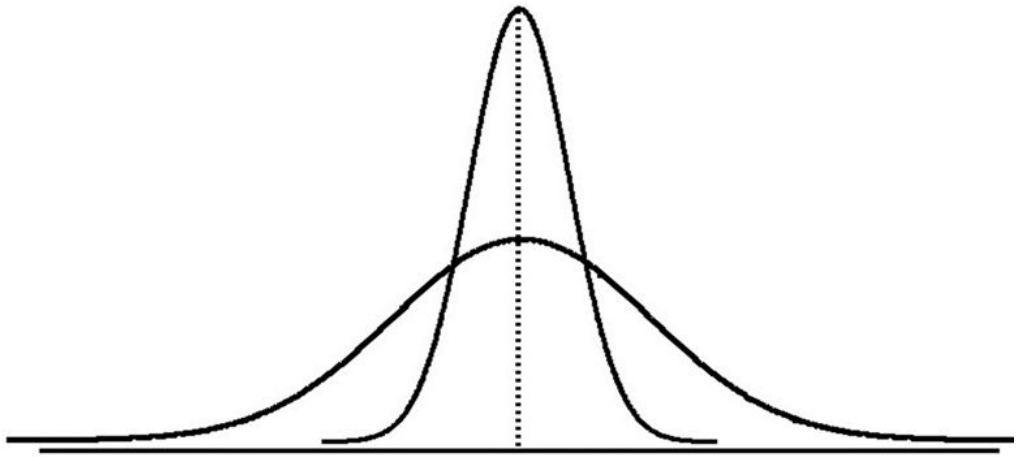
Measurement of timeliness/reliability/consistency

**Financial sector:**

Measurement of risk (difference between actual rate of return and the expected rate of return)

**Production**:

Quality control management. Measurement of consistency and reliability of manufacturing processes



OKAN ÜNİVERSİTESİ
İSTANBUL

# Standard deviation a measure for risk in Finance

Comparing 2 different assets, asset A and asset B with the same mean:

**Table 2.5** Rates of Return: Asset A and Asset B

| | *Asset A* | *Asset B* |
|---|---|---|
| Mean Rate of Return | 12.2% | 12.2% |
| Standard Deviation in Rate of Return | 0.63 | 3.12 |

Copyright ©2013 Pearson Education, publishing as Prentice Hall

# Standard deviation a measure for consistency in quality control

(Consistency in Turkish: Tutarlılık)

Comparing two manufacturing processes for number of defects in a sample, with similar means of defects:

| *Process 1:* | | | *Process 2* | |
|---|---|---|---|---|
| | | | | |
| Mean | 10.4 | | Mean | 10 |
| Median | 11 | | Median | 10 |
| Mode | #N/A | | Mode | #N/A |
| Standard Deviation | 4.393177 | | Standard Deviation | 1.581139 |
| Sample Variance | 19.3 | | Sample Variance | 2.5 |
| Count | 5 | | Count | 5 |

OKAN ÜNİVERSİTESİ
İSTANBUL

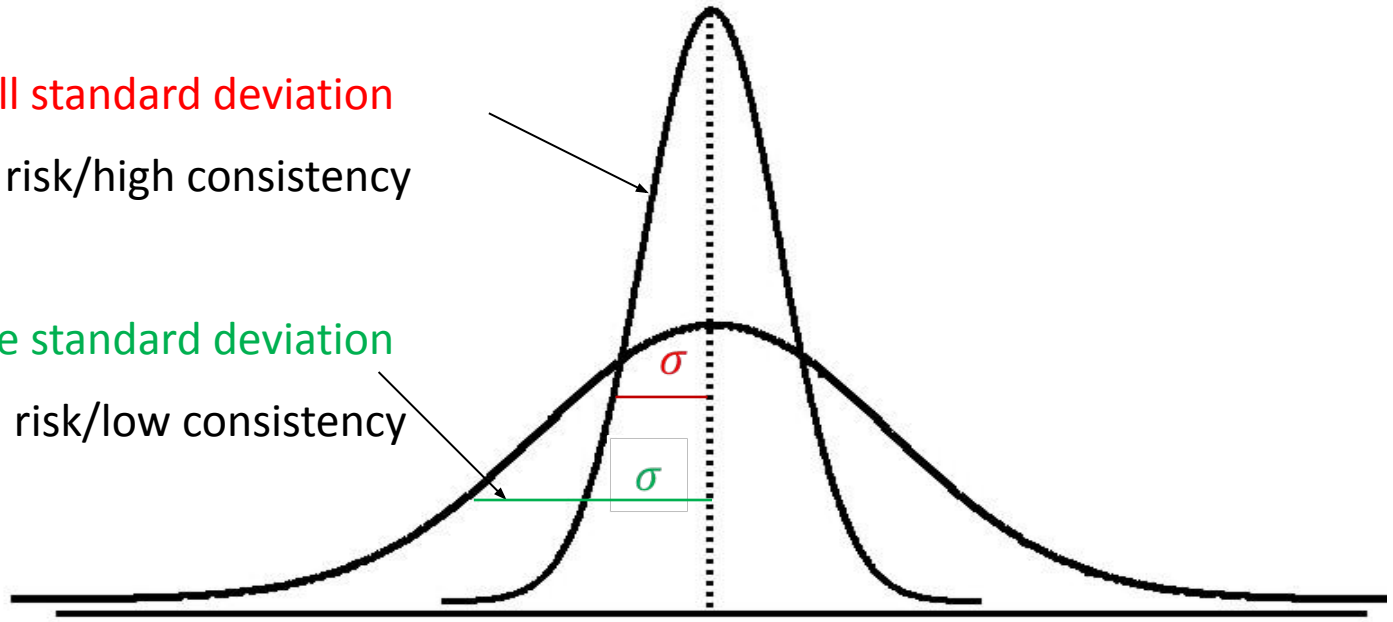
# Measuring standard deviation

Small standard deviation

Low risk/high consistency

Large standard deviation

High risk/low consistency



# asuring standard deviation

What does a standard deviation of 0 indicate?

What shape will the distribution have?

# Measuring the standard tion

Example of a data set with a standard deviation of 0:

53 53 53 53 53 53

# Advantages of Variance and Standard Deviation

**Each** single value in the data set is used in the calculation

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

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

Values far from the mean are given extra weight, such as outliers

(because deviations from the mean are squared)

COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

OKAN ÜNİVERSİTESİ
İSTANBUL

# Effect of outliers on Variance and standard deviation

A large outlier (negative or positive) will increase the variance and standard deviation

OKAN ÜNİVERSİTESİ
İSTANBUL

# Comparing the consistency of two types of Golf clubs

Golf equipment manufacturers are constantly seeking ways to improve their products. Suppose that the R&D department has developed a new golf iron (7-iron) to improve the consistency of its users.

A test golfer was asked to hit 150 shots using a 7-iron, 75 of which were hit with his current club and 75 with the newly developed 7-iron.

The distances were then measured and recorded.

COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

OKAN ÜNİVERSİTESİ
İSTANBUL

# Which iron is more consistent? The current or the newly developed? Excel output:

| *Current* | | | *Innovation* | |
|---|---|---|---|---|
| | | | | |
| Mean | 150.5466667 | | Mean | 150.1466667 |
| Standard Error | 0.668814558 | | Standard Error | 0.357011284 |
| Median | 151 | | Median | 150 |
| Mode | 150 | | Mode | 149 |
| Standard Deviation | 5.792103976 | | Standard Deviation | 3.091808416 |
| Sample Variance | 33.54846847 | | Sample Variance | 9.559279279 |
| Kurtosis | 0.126739586 | | Kurtosis | -0.885417995 |
| Skewness | -0.429888289 | | Skewness | 0.177337733 |
| Range | 28 | | Range | 12 |
| Minimum | 134 | | Minimum | 144 |
| Maximum | 162 | | Maximum | 156 |
| Sum | 11291 | | Sum | 11261 |
| Count | 75 | | Count | 75 |

COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

# terpretation of the data (golf club)

The standard deviation of the distances of the **current iron is 5.79 meters** whereas that of the newly developed **7-iron is 3.09 meters**.

Based on this sample, the  newly developed iron is more **consistent** (there is less variation in the distances shot with the innovative golf club).

Because the mean distances are similar it would appear that the new 7-iron is indeed superior.

COPYRIGHT © 2013 PEARSON EDUCATION, INC. PUBLISHING AS PRENTICE HALL

# Coefficient of Variation (CV)

In situations where the means are almost the same, it is appropriate to use the standard deviations to see which process is the most consistent.

In situations where the means are **different** we need to calculate the ***coefficient of variation*** to compare the consistency or riskiness.

The ***coefficient of variation*** expresses the standard deviation as a percentage of the mean.

OKAN ÜNİVERSİTESİ
İSTANBUL

# Coefficient of Variation (CV)

Measures relative variation within a dataset

Always in percentage (%) 0 – 100

A low CV translates into low variation within the same data set, a high CV into high variation

**Population coefficient of variation (CV):**

$$CV = \left( \frac{\sigma}{\mu} \right) \cdot 100\%$$

**Sample coefficient of variation (CV)**:

$$CV = \left( \frac{s}{\bar{x}} \right) \cdot 100\%$$

# Coefficient of Variation (CV)

Comparing 2 different production processes with different means:

| Process 1 | Process 2 |
|---|---|
| 18 | 18 |
| 19 | 35 |
| 15 | 12 |
| 18 | 19 |
| 17 | 16 |
| 87/ 5 = $\overline{X}$ = **17.4** | 100/5 = $\overline{X}$ = **20** |
| s = 1.51 | s = 8.80 |

# Coefficient of Variation (CV)

**Process 1:**

87/ 5 = $\bar{X}$ = 17.4

s = 1.51

**Process 2:**

100/5 = $\bar{X}$ = 20

s = 8.80

$CV = \left(\frac{s}{\bar{x}}\right) \cdot 100\%$ = $\frac{1.51}{17.4}$ x 100 % = 0.086 x 100% = **8.68%**

CV = $\frac{8.8}{20}$ = 0.44 x 100% = **44 %**

# Comparing Coefficient of Variation

**Stock A:**

- Average price last year = $ 4.00
- Standard deviation = $ 2.00

**Stock B:**

- Average price last year = $ 80.00
- Standard deviation = $ 8.00

**Note:** The standard deviation for stock A is **lower** than the standard deviation for stock B.

# Comparing Coefficient of Variation

**Stock A:**

- Average price last year = $ 4.00
- Standard deviation = $ 2.00

$$CV_A = \left(\frac{s}{\bar{x}}\right) \cdot 100\% = \frac{\$2.00}{\$4.00} \cdot 100\% = 50\%$$

**Stock B:**

- Average price last year = $ 80.00
- Standard deviation = $ 8.00

$$CV_B = \left(\frac{s}{\bar{x}}\right) \cdot 100\% = \frac{\$8.00}{\$80.00} \cdot 100\% = 10\%$$

**Note:** The standard deviation for stock A is **lower** than the standard deviation for stock B.

# Comparing Coefficient of Variation, (CV)

The standard deviation of stock A, is $2, and that of stock B, is $ 8, we would believe that stock B is more volatile or risky.

However, the average closing price for stock A is $ 4, and $ 80 for stock B.

The CV of stock A is higher, 50%, meaning that the market value of the stock fluctuates more from period to period than does that of stock B, 10%.

Therefore, a **lower CV** indicates **lower riskiness in finance** and **higher precision or consistency in a production process.**

# When to use Standard deviation and coefficient of variation, when comparing two data sets

**Use Standard deviation, SD, as a measure of risk/ consistency/reliability when comparing two or more objects:**

- Means are identical or very close

**Use Coefficient of variation, CV, as a measure of risk/ consistency/reliability when comparing two or more objects:**

- Means are different

The ***coefficient of variation, CV,*** expresses the standard deviation as a percentage of it's mean. Is measured between 0 – 100 %.

# Standard deviation and coefficient of variation – measures of variation

The standard deviation is the average distance of all the scores within a distribution around the mean.

The coefficient of variation is the standard deviation **relative (in percent)** to its' mean.

We can use the coefficient of variation to determine the relative variance **within one particular process**.

OKAN ÜNİVERSİTESİ
İSTANBUL

# Application of coefficient of variation, CV

With the following information about investment A:

| | Investment A |
|---|---|
| Average annual return | 10.4 |
| Standard deviation | 4.393 |

Can we say what risk it carries? Is this a high or low risk?

OKAN ÜNİVERSİTESİ
İSTANBUL

# Application of coefficient of variation

(continued)

With the coefficient of variation we can analyze the relative variation (in percent) around the mean:

| | Investment A |
|---|---|
| **Average annual return** | 10.4 |
| **Standard deviation** | 4.393 |
| **Coefficient of variation** | 0.422403846 |

The coefficient of variation tells us that for investment A the sample standard deviation is 42.2 % from the mean.

OKAN ÜNİVERSİTESİ
İSTANBUL

# Class quizz

What is the median?

What does the Range measure?

What does IQR measure?

How do we illustrate categorical data?

Why do we collect a sample from the population?

What are data?

What types of data do we work with in statistics?

OKAN ÜNİVERSİTESİ
İSTANBUL

# Class quizz

**Comparing the variation/ spread in two different processes: Standard deviation and Coefficient of variation:**

(1) In which situation will we use the standard deviation as the measure of variation?

(2) In which **situations** will we need to use the Coefficient of variation?