

ECO 391 Economics and Business Statistics

Lecture 1: Review of Descriptive Statistics

Xiaozhou Ding

January 14, 2019

Overview

- 1 Introduction
- 2 Measures of Central Location
- 3 Measures of Dispersion
- 4 Covariance and Correlation
- 5 Summary

Introduction

Outline

- Measures of Central Location
 - ▶ Mean
 - ▶ Median
 - ▶ Mode
- Measures of Dispersion
 - ▶ Range
 - ▶ Mean Absolute Deviation (MAD)
 - ▶ Variance and Standard Deviation
 - ▶ Coefficient of Variation(CV)
- Covariance and Correlation

Introductory Case: Investment Decision

Example

As an investment counselor at a large bank, Rebecca Johnson was asked by an inexperienced investor to explain the differences between two top-performing mutual funds:

- Vanguard's Precious Metals and Mining fund (Metals)
- Fidelity's Strategic Income Fund (Income)

The investor has collected sample returns for these two funds for years 2000 through 2009. These data are presented in the next slide.

Introductory Case: Investment Decision

Example

Year	Metals	Income	Year	Metals	Income
2000	-7.34	4.07	2005	43.79	3.12
2001	18.33	6.52	2006	34.3	8.15
2002	33.35	9.38	2007	36.13	5.44
2003	59.45	18.62	2008	-56.02	-11.37
2004	8.09	9.44	2009	76.46	31.77

Rebecca would like to

- 1 Determine the typical return of the mutual funds.
- 2 Evaluate the investment risk of the mutual funds.

Measures of Central Location

Mean

Definition

The **arithmetic mean** is a primary measure of central location. It is often referred to it as the average. We usually denote the sample mean as

$$\bar{x} = \frac{\sum x_i}{n},$$

and the population mean as

$$\mu = \frac{\sum x_i}{N}.$$

Example

Now back to our example. Use the data in the introductory case to calculate and interpret the mean return of the Metals fund and the mean return of the Income fund.

$$\text{Metals fund mean return} = \frac{-7.34 + 18.33 + \dots + 76.46}{10} = \frac{246.54}{10} = 24.65\%,$$

$$\text{Income fund mean return} = \frac{4.07 + 6.52 + \dots + 31.77}{10} = \frac{85.14}{10} = 8.51\%.$$

Outliers

- The mean is sensitive to outliers.
- Consider the salaries of employees at Acetech.

Title	Salary (\$)
Administrative Assistant	40,000
Research Assistant	40,000
Computer Programmer	65,000
Senior Research Associate	90,000
Senior Sales Associate	145,000
Chief Financial Officer	150,000
President (and owner)	550,000

$$\mu = \frac{\sum x_i}{N} = \frac{40,000 + 40,000 + \dots + 550,000}{7} = \$154,286.$$

- This mean does not reflect the typical salary!

Median

- The median is another measure of central location that is not affected by outliers.
- When the data are arranged in ascending order, the median is:
 - ▶ the middle value if the number of observations is odd, or
 - ▶ the average of the two middle values if the number of observations is even.

Median

- The median is another measure of central location that is not affected by outliers.
- When the data are arranged in ascending order, the median is:
 - ▶ the middle value if the number of observations is odd, or
 - ▶ the average of the two middle values if the number of observations is even.

Example

- Consider the sorted salaries of employees at Acetech (odd number).
Median is 90,000.

Median

- The median is another measure of central location that is not affected by outliers.
- When the data are arranged in ascending order, the median is:
 - ▶ the middle value if the number of observations is odd, or
 - ▶ the average of the two middle values if the number of observations is even.

Example

- Consider the sorted salaries of employees at Acetech (odd number).
Median is 90,000.
- Consider the sorted data from the Metals funds of the introductory case study (even number).

Year	2008	2000	2004	2001	2002	2006	2007	2005	2003	2009
Metals	-56.02	-7.34	8.09	18.33	33.35	34.3	36.13	43.79	59.45	76.46

$$\text{Median} = \frac{33.35 + 34.30}{2} = 33.83.$$

Question

How to interpret the median returns for the Metals funds?

Year	2008	2000	2004	2001	2002	2006	2007	2005	2003	2009
Metals	-56.02	-7.34	8.09	18.33	33.35	34.3	36.13	43.79	59.45	76.46

- Over the period 2000-2009, the Metals funds have the median return of 33.83% , which indicates that 5 years had return less than 33.83%, and 5 years had returns greater than 33.83%.
- As a wise counselor, in order to give your customer a transparent description of these two funds, you should report both the mean and median.

Mode

Definition

The mode is another measure of central location, it is

- the most frequently occurring value in a data set,
- used to summarize qualitative data,
- a data set can have no mode, one mode (unimodal), or many modes (multimodal).

Example

Consider the salary of employees at Acetech. The mode is \$40,000 since this value appears most often.

Weighted Mean

The weighted mean is relevant when some observations contribute more than others. For example, in most classes, different assignments have different weights.

Definition

Let w_1, w_2, \dots, w_n denote the weights of the sample observations x_1, x_2, \dots, x_n such that $w_1 + w_2 + \dots + w_n = 1$. The weighted mean for the sample is computed as:

$$\bar{x} = \sum w_i x_i.$$

The weighted mean for the population is computed similarly.

Example

A student scores 60 on Exam 1, 70 on Exam 2, and 80 on Exam 3. What is the student's average score for the course if Exams 1, 2, and 3 are worth 25%, 25%, and 50% of the grade, respectively?

Example

A student scores 60 on Exam 1, 70 on Exam 2, and 80 on Exam 3. What is the student's average score for the course if Exams 1, 2, and 3 are worth 25%, 25%, and 50% of the grade, respectively?

- Define $w_1 = 0.25$, $w_2 = 0.25$, and $w_3 = 0.50$.

Example

A student scores 60 on Exam 1, 70 on Exam 2, and 80 on Exam 3. What is the student's average score for the course if Exams 1, 2, and 3 are worth 25%, 25%, and 50% of the grade, respectively?

- Define $w_1 = 0.25$, $w_2 = 0.25$, and $w_3 = 0.50$.

$$\bar{x} = \sum w_i x_i = 0.25 \times 60 + 0.25 \times 70 + 0.5 \times 80 = 72.50.$$

Example

A student scores 60 on Exam 1, 70 on Exam 2, and 80 on Exam 3. What is the student's average score for the course if Exams 1, 2, and 3 are worth 25%, 25%, and 50% of the grade, respectively?

- Define $w_1 = 0.25$, $w_2 = 0.25$, and $w_3 = 0.50$.

$$\bar{x} = \sum w_i x_i = 0.25 \times 60 + 0.25 \times 70 + 0.5 \times 80 = 72.50.$$

- The unweighted mean is only 70 as it does not incorporate the higher weight given to the score on Exam 3.

Measures of Dispersion

Measures of Dispersion

- Measures of dispersion gauge the variability of a data set.
- Measures of dispersion include:
 - ▶ Range
 - ▶ Mean Absolute Deviation (MAD)
 - ▶ Variance and Standard Deviation
 - ▶ Coefficient of Variation (CV)

Range

Definition

Range is the difference between maximum value and minimum value of a data set, i.e.,

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value}.$$

- It is the simplest measure.
- It focuses on extreme values.

Range

Definition

Range is the difference between maximum value and minimum value of a data set, i.e.,

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value}.$$

- It is the simplest measure.
- It focuses on extreme values.

Example

Calculate the range using the data from the Metals and Income funds.

Range

Definition

Range is the difference between maximum value and minimum value of a data set, i.e.,

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value}.$$

- It is the simplest measure.
- It focuses on extreme values.

Example

Calculate the range using the data from the Metals and Income funds.

- Metals fund: $76.46\% - (-56.02\%) = 132.48\%$.

Range

Definition

Range is the difference between maximum value and minimum value of a data set, i.e.,

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value}.$$

- It is the simplest measure.
- It focuses on extreme values.

Example

Calculate the range using the data from the Metals and Income funds.

- Metals fund: $76.46\% - (-56.02\%) = 132.48\%$.
- Income fund: $31.77\% - (-11.37\%) = 43.14\%$.

Mean Absolute Deviation (MAD)

Definition

MAD is an average of the absolute difference of each observation from the mean.

$$\text{Sample MAD} = \frac{\sum |x_i - \bar{x}|}{n},$$

and

$$\text{Population MAD} = \frac{\sum |x_i - \mu|}{N}.$$

Example

Calculate MAD using the data from the Metals fund.

Example

Calculate MAD using the data from the Metals fund.

x_i	$x_i - \bar{x}$	$ x_i - \bar{x} $
-7.34	-7.34-24.65=-31.99	31.99
18.33	18.33-24.65=-6.32	6.32
\vdots	\vdots	\vdots
76.46	76.46-24.65=51.81	51.81
		Total=271.12

$$\text{MAD} = \frac{271.12}{10} = 27.11$$

Variance and Standard Deviation

Definition

For a given sample, variance s^2 is defined as

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}, \quad \text{and} \quad s = \sqrt{s^2}.$$

For a given population, variance σ^2 is defined as

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N}, \quad \text{and} \quad \sigma = \sqrt{\sigma^2}.$$

Example

Calculate the variance and the standard deviation using the data from the Metals fund.

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1} = \frac{12,407.44}{10 - 1} = 1378.60.$$

$$s = \sqrt{1,378.60} = 37.13.$$

Coefficient of Variation (CV)

- CV adjusts for differences in the magnitudes of the means.
- CV is unitless, allowing easy comparisons of mean-adjusted dispersion across different data sets.

$$\text{Sample CV} = \frac{s}{\bar{x}},$$

and

$$\text{Population CV} = \frac{\sigma}{\mu}.$$

Coefficient of Variation (CV)

- CV adjusts for differences in the magnitudes of the means.
- CV is unitless, allowing easy comparisons of mean-adjusted dispersion across different data sets.

$$\text{Sample CV} = \frac{s}{\bar{x}},$$

and

$$\text{Population CV} = \frac{\sigma}{\mu}.$$

Example

Calculate the coefficient of variation (CV) using the data from the Metals fund and the Income fund.

Coefficient of Variation (CV)

- CV adjusts for differences in the magnitudes of the means.
- CV is unitless, allowing easy comparisons of mean-adjusted dispersion across different data sets.

$$\text{Sample CV} = \frac{s}{\bar{x}},$$

and

$$\text{Population CV} = \frac{\sigma}{\mu}.$$

Example

Calculate the coefficient of variation (CV) using the data from the Metals fund and the Income fund.

- Metals fund CV: $\frac{s}{\bar{x}} = \frac{37.13\%}{24.65\%} = 1.51$.
- Income fund CV: $\frac{s}{\bar{x}} = \frac{11.07\%}{8.51\%} = 1.30$.

Synopsis of Investment Decision

- Mean and median returns for the Metals fund are 24.65% and 33.83%, respectively.
- Mean and median returns for the Income fund are 8.51% and 7.34%, respectively.
- The standard deviation for the Metals fund and the Income fund are 37.13% and 11.07%, respectively.
- The coefficient of variation for the Metals fund and the Income fund are 1.51 and 1.30, respectively.

Don't forget our initial questions. Typical return and investment risk?

Covariance and Correlation

Covariance and Correlation

- The covariance (s_{xy} or σ_{xy}) describes the direction of the linear relationship between two variables, x and y .
- The correlation coefficient (r_{xy} or ρ_{xy}) describes both the direction and strength of the relationship between x and y .

Covariance

- The sample covariance s_{xy} is computed as

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}.$$

- The population covariance σ_{xy} is computed as

$$\sigma_{xy} = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{N}.$$

Correlation

- The sample correlation r_{xy} is computed as

$$r_{xy} = \frac{s_{xy}}{s_x s_y}.$$

- The population correlation ρ_{xy} is computed as

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

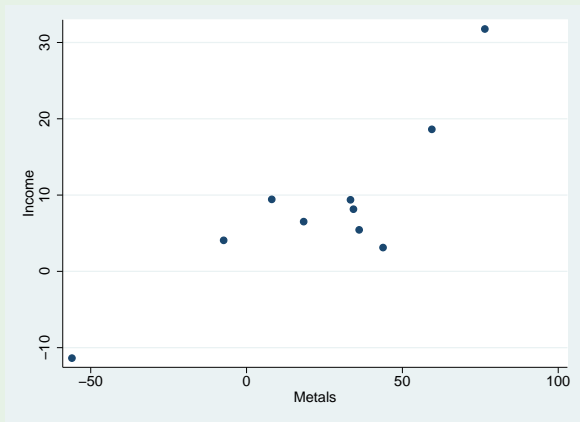
- Note, $-1 \leq r_{xy} \leq 1$ or $-1 \leq \rho_{xy} \leq 1$.

Example

Let's calculate the covariance and the correlation coefficient for the Metals (x) and Income (y) funds.

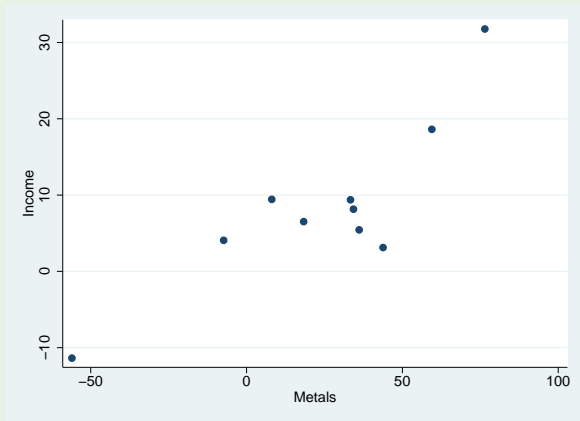
Example

Let's calculate the covariance and the correlation coefficient for the Metals (x) and Income (y) funds.



Example

Let's calculate the covariance and the correlation coefficient for the Metals (x) and Income (y) funds.



- Positive relationship

Example

Recall $\bar{x} = 24.65$, $s_x = 37.13$, $\bar{y} = 8.51$, $s_y = 11.07$.

Example

Recall $\bar{x} = 24.65$, $s_x = 37.13$, $\bar{y} = 8.51$, $s_y = 11.07$.

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{3165.55}{9} = 351.73$$

Example

Recall $\bar{x} = 24.65$, $s_x = 37.13$, $\bar{y} = 8.51$, $s_y = 11.07$.

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{3165.55}{9} = 351.73$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{351.73}{37.13 \times 11.07} = 0.86$$

Summary

Summary

- Measure of Central Location
 - ▶ Mean
 - ▶ Median
 - ▶ Mode
- Measure of Dispersion
 - ▶ Range
 - ▶ Mean Absolute Deviation (MAD)
 - ▶ Variance and Standard Deviation
 - ▶ Coefficient of Variation(CV)
- Covariance and Correlation