Quantitative Methods

Statistical Measures of Asset Returns

Intro and Exam Focus

- Measures of central tendency
  - Arithmetic mean (trimmed, winsorized?)
  - Median: ½ higher and ½ lower
  - Mode: most frequent outcome
- Measures of dispersion
  - Standard deviation and variance
  - Range: highest to lowest
  - MAD: mean absolute deviation

Median

- Midpoint of a dataset: **half above and half below**

  With an odd number of observations

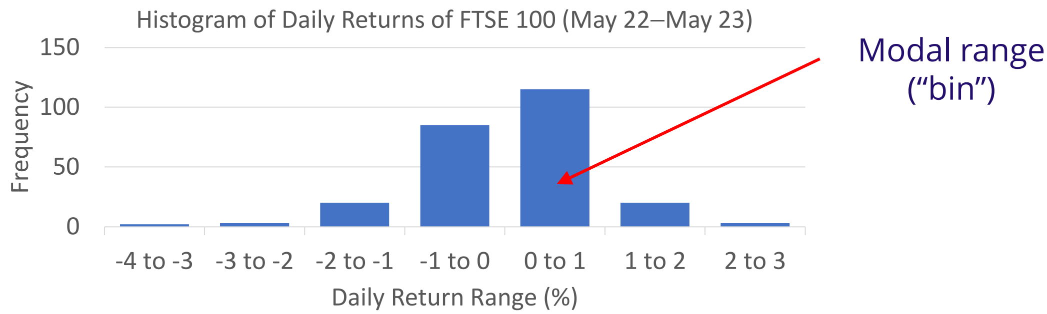
  2, 5, 7, 11, 14 Median =
- With an even number of observations, median is the average of the two middle observations

  3, 9, 10, 20 Median =
- Less affected by extreme values than the mean

-2

Mode

- Value occurring most frequently in a dataset



- Datasets can have more than one mode (bimodal, trimodal, etc.), or no mode (all values are the same)

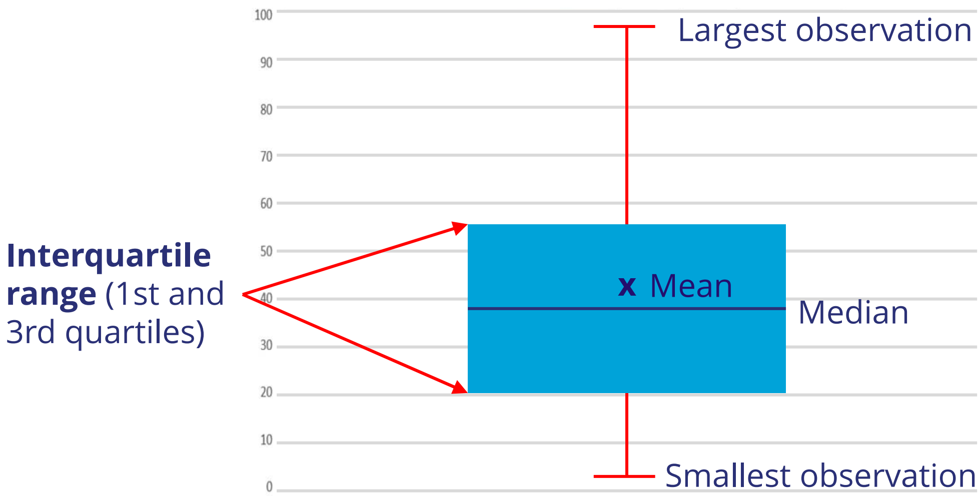
Quantiles: Example

- 75% of the data points are less than the 3rd **quartile**
- 60% of the data points are less than the 60th **percentile** (6th **decile**)
- What is the 1st decile and 1st quartile for the FTSE 100 daily return data displayed below?

		Return Range		
Bin	**Cumulative Percentage of Trading Days**	**Lower**	**Upper**	**Frequency**
1	5%	–3.83%	–1.59%	12
2	10%	–1.59%	–1.01%	12
3	15%	–1.01%	–0.76%	13
4	20%	–0.76%	–0.50%	12
5	25%	–0.50%	–0.37%	13
6	30%	–0.37%	–0.25%	12

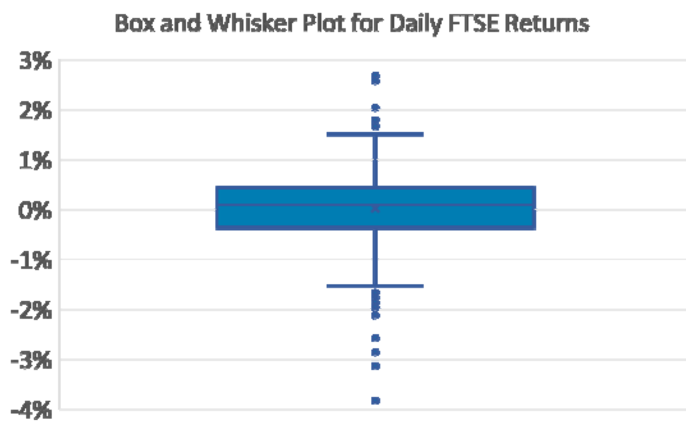
-2

Box and Whisker Plot



Box and Whisker Plot (cont.)

- To display outliers, "fences" can be positioned at 1.5 × the interquartile range of the data:



CFA Institute Data

Monthly portfolio returns:

Month	X_i Return (%)	Month	X_i Return (%)
January	5	July	0
February	3	August	4
March	–1	September	3
April	–4	October	0
May	4	November	6
June	2	December	5

This data will be used in the examples following this slide

Range and Mean Absolute Deviation

1. What is the data's range?

Range = – =

2. What is the data's mean absolute deviation (MAD)?

Step 1 (compute mean):

$$\text{Mean} = \frac{(5+3-1-4+4+2+0+4+3+0+6+5)}{12} =$$

-2

Mean Absolute Deviation (cont.)

2. What is the data's mean absolute deviation (MAD)?

Step 2 (compute difference between observation and mean):

5% – 2.25%

X_i Return (%)	$X_i - \overline{X}$	X_i Return (%)	$X_i - \overline{X}$
5		0	2.25
3	0.75	4	1.75
–1	–3.25	3	0.75
–4	–6.25	0	–2.25
4	1.75	6	3.75
2	–0.25	5	2.75

-2

Mean Absolute Deviation (cont.)

2. What is the data's mean absolute deviation (MAD)?

Step 3 (treat deviations as absolute values and compute mean deviation):

$$\text{MAD} = \frac{2.75 + 0.75 + 3.25 + 6.25 + 1.75 + 0.25 + 2.25 + 1.75 + 0.75 + 2.25 + 3.75 + 2.75}{12}$$

$$\text{MAD} = \qquad =$$

-1

Sample Variance and Sample Standard Deviation

- Sample variance (s^2): average squared distance from the sample mean ($\overline{X}$)

$$s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$$

- Sample standard deviation (s) = $\sqrt{s^2}$

Standard Deviation: Example

$\overline{X} = 2.25\%$

X_i Return (%)	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$	X_i Return (%)	$X_i - \overline{X}$	$\left(X_i - \overline{X}\right)^2$
5	2.75	7.5625	0	−2.25	5.0625
3	0.75	0.5625	4	1.75	3.0625
−1	−3.25	10.5625	3	0.75	0.5625
−4	−6.25	39.0625	0	−2.25	5.0625
4	1.75	3.0625	6	3.75	14.0625
2	−0.25	0.0625	5	2.75	7.5625

Σ

Much quicker on the BAII+

$S_x = \qquad =$

-2

Target Downside Deviation (or Target Semideviation)

$$s_{target} = \sqrt{\frac{\sum_{\text{all } X_i < B}^{n} (X_i - B)^2}{n - 1}}$$

Similar to sample standard deviation, but the numerator only uses those observations that are **less than a chosen target**, B

Target Downside Deviation: Example

Calculate the downside deviation when the target is 3%:

X_i Return (%)	X_i – B	$(X_i - B)^2$	X_i Return (%)	X_i – B	$(X_i - B)^2$
~~5~~			0		
~~3~~			~~4~~		
–1			3		
–4			0		
~~4~~			~~6~~		
2			~~5~~		

Σ

$s_{target} = \qquad =$

-4

Coefficient of Variation (CV): Example

- Measures dispersion per unit of mean (risk per unit of return)

$$CV = \frac{s}{\overline{X}}$$

- Which asset would be preferred based on its coefficient of variation?

Asset	Mean	s
Asset A	5%	10%
Asset B	8%	12%

$CV_A =$ $=$ $CV_B =$ $=$

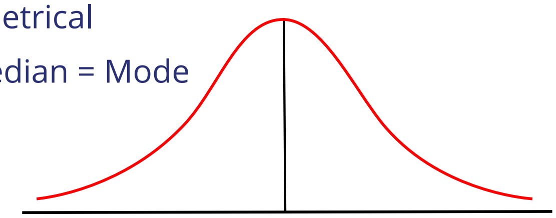
-2

Skewness

- Skew measures the degree to which a distribution **lacks** symmetry
- Based on the average *cubed* deviation from the mean
- A symmetrical distribution has skew = 0 (e.g., the normal distribution)

Symmetrical

Mean = Median = Mode



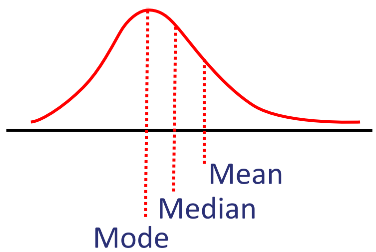
Mean

Median

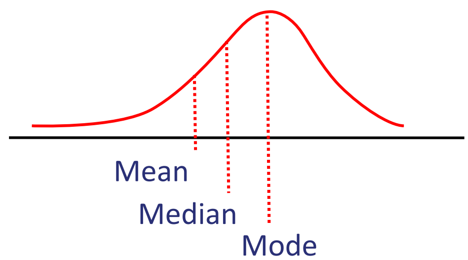
Mode

Skewness (cont.)

- **Positive** skew has outliers in the right tail

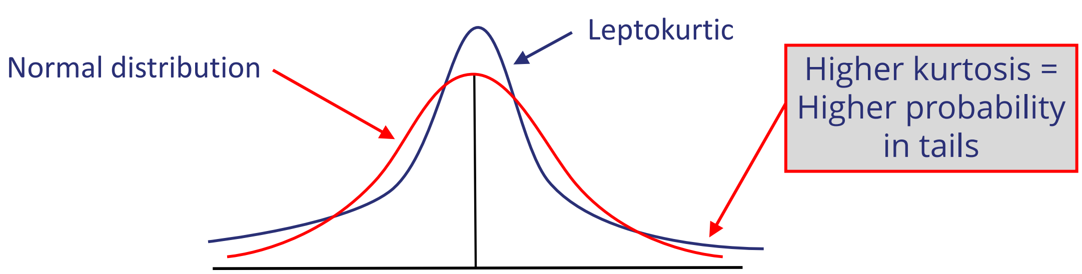


- **Negative** skew has outliers in the left tail



Kurtosis

- Measures the degree to which a distribution is more or less peaked than a normal distribution
- Based on average deviation from the mean to the power of *four*



- *Leptokurtic* (kurtosis > 3) is **more peaked** with **fatter tails** (more extreme outliers)

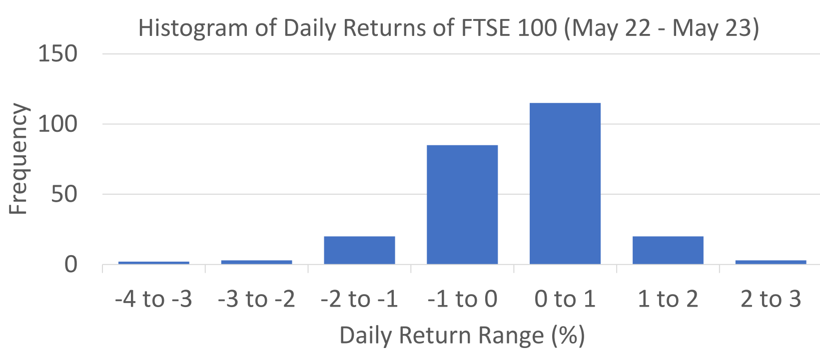
Kurtosis (cont.)

- **Kurtosis for a normal distribution = 3.0**
- **Excess** kurtosis = kurtosis – 3

Kurtosis	Excess Kurtosis	Definition	Shape vs. Normal Distribution
>3	>0	Leptokurtosis	Peaked, fat tails
3	0	Mesokurtic	Normal
<3	<0	Platykurtic	Flatter, thinner tails

Skewness/Kurtosis

- FTSE 100 daily returns revisited
  - Average = 0.01%, sample standard deviation = 0.86%
  - Skew = –0.7, kurtosis = 2.8



Sample Covariance

- A measure of how two variables move *together:*

$$s_{X,Y} = \frac{\sum_{i=1}^{n}\left[\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)\right]}{n-1}$$

- Difficult to interpret for two reasons:
  - Units of covariance are squares of the units of the underlying data
  - Does not indicate the strength of the relationship

Correlation

$s_{XY} = 0.0046$, $s_X = 0.0623$, $s_Y = 0.0991$

$$r_{XY} = \frac{S_{XY}}{S_X \times S_Y} = \qquad\qquad =$$

- Indicates strength of *linear* relationships
- Bounded between –1 and +1 → easier to interpret
- Does not indicate *nonlinear* relationships
- Correlation does not imply causality: unrelated variables may show **spurious correlation**

-1

Solutions

Median

- Midpoint of a dataset: **half above and half below**

  With an <u>odd number</u> of observations

  2, 5, 7, 11, 14   Median = 7

- With an <u>even number</u> of observations, median is the average of the two middle observations

  3, 9, 10, 20   Median = (9 + 10) / 2 = 9.5

- Less affected by extreme values than the mean

Quantiles: Example

- 75% of the data points are less than the 3rd **quartile**
- 60% of the data points are less than the 60th **percentile** (6th **decile**)
- What is the 1st decile and 1st quartile for the FTSE 100 daily return data displayed below?

1st decile (10%)

1st quartile (25%)

Bin	Cumulative Percentage of Trading Days	Return Range Lower	Return Range Upper	Frequency
1	5%	–3.83%	–1.59%	12
2	10%	–1.59%	–1.01%	12
3	15%	–1.01%	–0.76%	13
4	20%	–0.76%	–0.50%	12
5	25%	–0.50%	–0.37%	13
6	30%	–0.37%	–0.25%	12

-2

Range and Mean Absolute Deviation

1. What is the data's range?

Range = 6% – (–4%) = 10%

2. What is the data's mean absolute deviation (MAD)?

Step 1 (compute mean):

$$\text{mean}=\frac{(5+3-1-4+4+2+0+4+3+0+6+5)}{12}=2.25\%$$

-2

Mean Absolute Deviation (cont.)

2. What is the data's mean absolute deviation (MAD)?

Step 2 (compute difference between observation and mean):

5% – 2.25%

X_i Return (%)	$X_i - \bar{X}$	X_i Return (%)	$X_i - \bar{X}$
5	2.75	0	–2.25
3	0.75	4	1.75
–1	–3.25	3	0.75
–4	–6.25	0	–2.25
4	1.75	6	3.75
2	–0.25	5	2.75

-2

Mean Absolute Deviation (cont.)

2. What is the data's mean absolute deviation (MAD)?

Step 3 (treat deviations as absolute values and compute mean deviation):

$$\text{MAD} = \frac{2.75 + 0.75 + 3.25 + 6.25 + 1.75 + 0.25 + 2.25 + 1.75 + 0.75 + 2.25 + 3.75 + 2.75}{12}$$

$$\text{MAD} = \frac{28.5}{12} = 2.375$$

-1

Standard Deviation: Example

$\bar{X} = 2.25\%$

X_i Return (%)	$X_i - \bar{X}$	$(x_i - \bar{x})^2$	X_i Return (%)	$X_i - \bar{X}$	$(x_i - \bar{x})^2$
5	2.75	7.5625	0	–2.25	5.0625
3	0.75	0.5625	4	1.75	3.0625
–1	–3.25	10.5625	3	0.75	0.5625
–4	–6.25	39.0625	0	–2.25	5.0625
4	1.75	3.0625	6	3.75	14.0625
2	–0.25	0.0625	5	2.75	7.5625

Σ 96.25

Much quicker on the BAII+

$$S_x = \sqrt{\frac{96.25}{12-1}} = 2.958\%$$

-2

Target Downside Deviation: Example

Calculate the downside deviation when the target is 3%:

X_i Return (%)	X_i – B	$(X_i - B)^2$	X_i Return (%)	X_i – B	$(X_i - B)^2$
~~5~~			0	–3	9
~~3~~			~~4~~		
–1	–4	16	3		
–4	–7	49	0	–3	9
~~4~~			~~6~~		
2	-1	1	~~5~~		

Σ 84

$$s_{target} = \sqrt{\frac{84}{12-1}} = 2.763\%$$

-4

Coefficient of Variation (CV): Example

- Measures dispersion per unit of mean (risk per unit of return)

$$CV = \frac{s}{\overline{X}}$$

- Which asset would be preferred based on its coefficient of variation?

Asset	Mean	s
Asset A	5%	10%
Asset B	8%	12%

$$CV_A = \frac{10}{5} = 2.0 \qquad CV_B = \frac{12}{8} = 1.5$$

-2

Correlation

$s_{XY} = 0.0046$, $s_X = 0.0623$, $s_Y = 0.0991$

$$r_{XY} = \frac{S_{XY}}{S_X \times S_Y} = \frac{0.0046}{0.0623 \times 0.0991} = 0.745$$

- Indicates strength of *linear* relationships
- Bounded between –1 and +1 → easier to interpret
- Does not indicate *nonlinear* relationships
- Correlation does not imply causality: unrelated variables may show **spurious correlation**

-1