LM03 · QUANTITATIVE METHODS · CFA LEVEL I

Statistical Measures of Asset Returns

Central tendency, dispersion, skewness, kurtosis, correlation

Why Statistics for Returns?

Returns distributions are rarely normal. We describe them using **central tendency**, **dispersion**, **shape** (skew/kurtosis), and **co-movement** (covariance/correlation). Each LOS corresponds to a different tool used by analysts to quantify risk and reward.

Learning Outcome Statements

LOS a Calculate, interpret, and evaluate measures of central tendency and location to address an investment problem

LOS b Calculate, interpret, and evaluate measures of dispersion to address an investment problem

LOS c Interpret and evaluate measures of skewness and kurtosis to address an investment problem

LOS d Interpret correlation between two variables to address an investment problem

Exam Tip: Know the BAII Plus 1-V statistics keystrokes cold. Given a list of returns, you should be able to retrieve $\bar{X}$, Sx (sample SD), σx (pop SD), and n in seconds.

LOS A · MEASURES OF CENTRAL TENDENCY

Mean, Median, Mode

$$\text{Arithmetic mean } \bar{X} = \Sigma X_{i} / n \quad \cdot \quad \text{Population } \mu \; = \Sigma X_{i} / N$$

MEASURE	FORMULA / RULE	STRENGTHS / WEAKNESSES
Arithmetic mean	Sum / n	Describes a representative outcome; sensitive to outliers
Median	Middle value (odd n) or avg of two middle (even n)	Not affected by outliers; ignores most data
Mode	Most frequently occurring value	Can be unimodal, bimodal, or none; discrete data
Trimmed mean	Discard top/bottom ×% of observations, then average	Reduces outlier effect — *deletes*
Winsorized mean	Replace tails with boundary values, then average	Reduces outlier effect — *substitutes*

WORKED EXAMPLE **Mean of 3-Year Returns**

GIVEN

Returns over 3 yrs: 10%, −5%, 20%.

SOLUTION

Arithmetic mean = (10 − 5 + 20)/3 = **8.33%**

WORKED EXAMPLE **Median Calculation**

GIVEN

(A) {1,2,3,4,5}   (B) {1,2,3,4,5,6}

SOLUTION

(A) Median = **3**. (B) Median = (3+4)/2 = **3.5**.

LOS A · MEASURES OF LOCATION

Quantiles and Box-and-Whisker Plots

Quantile Family

Quartiles	25%, 50%, 75% → split into quarters
Quintiles	20%, 40%, 60%, 80% → fifths
Deciles	10%, 20%, …, 90% → tenths
Percentiles	1%, 2%, …, 99% → hundredths

$$Position L_y = (n+1)\cdot y/100 — where y = percentile, n = number of obs$$

WORKED EXAMPLE **First Quartile (25th Percentile)**

GIVEN

Returns sorted: 3%, 4%, 6%, 9%, 11%, 12%, 14% (n = 7).

SOLUTION

L_{25} = (7+1)(25)/100 = 2. ⇒ 2nd item = **4%**. 25% of observations lie at or below this.

Box-and-Whisker Plot

- Box = interquartile range (IQR = Q3 − Q1)
- Line inside box = median
- Upper fence = Q3 + 1.5×IQR; Lower fence = Q1 − 1.5×IQR
- Points outside fences = outliers

Handling Outliers

Do Nothing

Values legit → eliminates judgment bias.

Delete

Trimmed mean — remove extreme x%.

Replace

Winsorized mean — cap at percentile bounds.

Use cases: rank portfolios by quartile; long-short HF buys top return decile, shorts bottom.

LOS B · MEASURES OF DISPERSION

Range, MAD, Variance, Standard Deviation

$$Range = max - min$$

$$MAD = \Sigma |X_{i} - \bar{X}| / n$$

$$\text{Sample variance } s^{2} = \Sigma (X_{i} - \bar{X})^{2} / (n - 1) \quad \text{Pop } \sigma ^{2} = \Sigma (X_{i} - \mu)^{2} / N$$

$$SampleSDs = \quad s^{2} \quad PopSD\sigma \;= \quad \sigma^{2}$$

Sample variance uses **n − 1** degrees of freedom because one degree is consumed estimating $\bar{X}$. This correction produces an *unbiased* estimator.

WORKED EXAMPLE **MAD of 4-Year Returns**

GIVEN

4%, 2%, 6%, 8%. Mean = 5%.

SOLUTION

MAD = (|4−5| + |2−5| + |6−5| + |8−5|)/4 = (1 + 3 + 1 + 3)/4 = **2%**

WORKED EXAMPLE **Sample vs Population SD via BAII Plus**

GIVEN

Same 4%, 2%, 6%, 8%.

SOLUTION

Keystrokes: [2nd][DATA], [2nd][CLR WRK], 4[ENTER], 2[ENTER], 6[ENTER], 8[ENTER], [2nd][STAT], ensure 1-V mode. Retrieve: N=4, $\bar{X}$=5, **Sx=2.58** (sample SD), **σx=2.24** (pop SD).

Exam Tip: Variance is additive; standard deviation is NOT. That's why portfolio risk calculations work in variance space (then take the square root at the end).

LOS B · DOWNSIDE RISK AND RELATIVE DISPERSION

Target Semideviation and Coefficient of Variation

Target Semideviation (Downside Deviation)

$$S_{target} = \quad [\Sigma_{X_i \le B}(X_i - B)^2/(n-1)]$$

Counts only observations that fall at or below target B. As B ↑, S_{target} ↑ because more observations breach the threshold. Denominator uses **full sample n**, not just the downside count.

Coefficient of Variation (CV)

$$CV = s / \bar{X}$$

Measures *relative* dispersion — risk per unit of mean return. Lower CV = tighter distribution per unit of return (better risk-adjusted).

WORKED EXAMPLE

CV Comparison Across Portfolios

GIVEN

A: mean 10%, SD 9%. B: mean 6%, SD 4%. C: mean 9%, SD 2%.

SOLUTION

CV_A = 9/10 = **0.90**

CV_B = 4/6 = **0.66**

CV_C = 2/9 = **0.22**

⇒ Portfolio C offers the least risk per unit of return — most attractive.

CHECK YOUR UNDERSTANDING · QUIZ 1

Quiz 1

Select the best answer. Feedback appears after each click.

Q1. Returns are 10%, −5%, 20%. The arithmetic mean return is closest to:

Q2. Quartile 1 of the sorted data {3%, 4%, 6%, 9%, 11%, 12%, 14%} is at position:

Q3. For returns 4%, 2%, 6%, 8%, the mean absolute deviation (MAD) equals:

Q4. Portfolios: A (10%, SD 9%), B (6%, SD 4%), C (9%, SD 2%). Which has the lowest CV?

Q5. A winsorized mean differs from a trimmed mean in that it:

LOS C · SKEWNESS

Shape of the Distribution — Skewness

Normal Distribution

Symmetrical; completely described by 2 parameters: μ, σ^2. Notation: $X \sim N(\mu, \sigma^2)$. For symmetry, **mean = median = mode**.

```
Skew ≈ (1/n) \; \cdot \; \Sigma (X_{i} - X̄)^{3} / s^{3} \quad (for n > 100)
```

SHAPE	ORDERING	INTERPRETATION
Skew = 0 (symmetric)	mean = median = mode	Normal, symmetric

Positive skew (right)	mean > median > mode	Long right tail; frequent small losses, few large gains
Negative skew (left)	mean < median < mode	Long left tail; frequent small gains, few large losses — most financial returns

Investor intuition: Equity returns historically exhibit negative skew — rare but severe drawdowns. Risk management models that assume normality will underestimate tail losses.

LOS C · KURTOSIS

Kurtosis — Tail Weight

```
K_{\text{E}} = (1/n) \; \cdot \; \Sigma (X_{i} - X̄)^{4} / s^{4} - 3
\quad \; (excess kurtosis)
```

Kurtosis captures combined weight of tails relative to the body.

TYPE	KURTOSIS	EXCESS KURTOSIS	DESCRIPTION
Leptokurtic	$K > 3$	$K_E > 0$	Fat tails, peaked center — over-weight head & tails, under-weight shoulders
Mesokurtic	$K = 3$	$K_E = 0$	Normal distribution
Platykurtic	$K < 3$	$K_E < 0$	Thin tails, flat center — under-weight head & tails, over-weight shoulders

Significance rule: Excess kurtosis with absolute value > 1 is considered significant. Equity returns typically show $K_E > 0$ (leptokurtic): a normal-based model will under-state the probability of extreme outcomes.

Exam Tip: Securities returns typically exhibit BOTH negative skew AND excess kurtosis. Risk managers focus on the tails, not the mean and SD alone.

LOS D · COVARIANCE AND CORRELATION

Co-Movement of Two Variables

Scatter Plot

Visualizes joint variation in two numerical variables. Identifies outliers, linear / non-linear relationships, or absence of relationship. A *scatter plot matrix* assesses pairwise association among many variables.

$$\text{Cov}(X, Y) = S_{XY} = \Sigma (X_{i} - \bar{X})(Y_{i} - \bar{Y}) / (n - 1)$$

$$Correlation r_{XY} = S_{XY}/(S_X \cdot S_Y)$$

Covariance = joint variability, expressed in *units squared*. Correlation standardizes it to [−1, +1] and has no units.

Correlation Properties

1. $-1 \le r \le +1$
2. $r = 0$ → no linear relationship ⇒ maximum diversification benefit
3. $r = +1$ → perfect positive correlation ⇒ perfect replication (no diversification)
4. $r = -1$ → perfect negative correlation ⇒ perfect hedge

WORKED EXAMPLE

Scatter Plot Statistics

GIVEN

$n = 11$, $\bar{X} = 9$, $\bar{Y} = 7.5$, $S_X = 3.32$, $S_Y = 2.03$.

SOLUTION

For all four scatter-plot data configurations in the classical *Anscombe quartet*-style example, r_{XY} = **0.82**, yet visual inspection shows different shapes ⇒ always inspect the scatter plot!

Limitations of Correlation

- Linear association only — misses non-linear structure
- Unreliable when outliers are present

- Correlation does NOT imply causation — may be spurious (chance, or due to a third variable)

CHECK YOUR UNDERSTANDING · QUIZ 2

Quiz 2

Select the best answer. Feedback appears after each click.

Q1. For a positively skewed distribution, the MOST accurate ordering is:

Q2. A distribution with excess kurtosis K_E = +2 is BEST described as:

Q3. Correlation of +1.0 between two assets implies:

Q4. Covariance between two variables is always expressed in:

Q5. Which of the following is a limitation of Pearson correlation?

LOS D · CORRELATION IN PORTFOLIO CONTEXT

How Correlation Drives Portfolio Risk

$$\mathrm{Var}(R_p) = w_A^2\sigma_A^2 + w_B^2\sigma_B^2 + 2w_Aw_B\sigma_A\sigma_B\rho_{AB}$$

Covariance enters the cross term. Portfolio SD < weighted SD of components *whenever* $\rho < 1$ — this is the quantitative core of diversification.

WORKED EXAMPLE **Two-Asset Portfolio Variance**

GIVEN

$w_A = w_B = 0.50$, $\sigma_A = \sigma_B = 20\%$, $\rho_{AB} = 0.4$.

SOLUTION

$Var(R_p) = 0.25(0.04) + 0.25(0.04) + 2(0.5)(0.5)(0.2)(0.2)(0.4)$

$= 0.01 + 0.01 + 0.008 = \mathbf{0.028}$

$\sigma_p = \sqrt{0.028} =$ **16.73%**, versus 20% weighted average ⇒ 327 bps of diversification.

Exam Tip: On the exam, expect a question that switches ρ from 0.4 to 0.8 (variance rises) or to −0.3 (variance drops). Always plug into the two-asset variance formula and remember the cross term doubles.

CHECK YOUR UNDERSTANDING · QUIZ 3

Quiz 3

Select the best answer. Feedback appears after each click.

Q1. For a sample of 10 returns with sum of squared deviations = 0.090, the sample variance is:

Q2. Target semideviation (downside deviation) vs standard deviation will be LARGER when:

Q3. Two assets with ρ = 0.40, equal weights 50/50, and equal σ = 20%. Portfolio SD is closest to:

Q4. A data set has mean 8% and standard deviation 12%. The coefficient of variation equals:

Q5. Which statement BEST describes a spurious correlation?

MODULE COMPLETE

LM03 · Statistical Measures of Asset Returns Summary

Final Score

0 / 0

Cheat Sheet

$$Mean = \Sigma X_i/n; median = middle; mode = most frequent$$

$$L_y = (n+1) \cdot y/100$$

$$s^2 = \Sigma(X_i - \bar{X})^2/(n-1); s = \quad s^2$$

```
CV = s / X̄ \quad \cdot \quad S_{target} = \sqrt{}[\Sigma _{X_{i}
\le B}(X_{i}-B)^{2} / (n-1)]
```

$$\text{Cov}(X,Y) = \Sigma(X_i - \bar{X})(Y_i - \bar{Y})/(n-1)$$

$$r_{XY} = S_{XY}/(S_X S_Y), -1 \le r \le +1$$

Last-Minute Exam Checklist

Skew: right-skew mean>median>mode; left-skew mean<median<mode

Kurtosis: K=3 normal; K>3 leptokurtic (fat tails); K<3 platykurtic

Trimmed vs winsorized: delete vs replace outliers

CV: lower = better risk per return (e.g., 2/9 = 0.22)

Correlation: +1 replication, −1 hedge, 0 max diversification

Portfolio var: ρ < 1 ⇒ diversification benefit

Step 1 of 13