CHARTA LEARNING · LESSON NOTES

Statistical characteristics of asset returns

CFA Level I · Quantitative Methods

You will learn: Describe a set of investment returns by its centre, spread, shape and position, and measure downside risk and risk per unit of return.

Running example: Examples use made-up figures. Two funds both average 8% a year, but fund 1 returned 8, 7, 9, 8, 8 and fund 2 returned 34, −18, 29, −12, 7, so one number is never enough.

1. Centre: mean, median and mode

- A measure of central tendency tells you where the middle of the data sits. The mean is the sum of all values divided by the count. It uses every value, so one huge value drags it.
- The median is the middle value once sorted (halfway between the two middle ones if the count is even). It cares only about order, so a giant value does not drag it. The mode is the value that appears most often: one, several, or none.
- When the mean sits well above the median, there is a long tail of high values: right skewed.

Mean = $(x_1 + x_2 + \ldots + x_n) \div n$. Median = the middle value, once sorted

FIVE FUNDS' RETURNS: 2%, 3%, 4%, 5%, 36%

Mean	(2 + 3 + 4 + 5 + 36) ÷ 5	**10%**
Median	3rd of 5 sorted	**4%**
Mode	no value repeats	**none**
Mean − median	the big winner pulls up	**6 points**

2. Geometric mean for compounding returns

- Returns compound, so for returns over several periods the ordinary (arithmetic) mean overstates what you earned. The geometric mean finds the single steady rate that gives the same final result.
- Turn each return R into a growth factor 1 + R (as a decimal), multiply them all, take the nth root (power 1 ÷ n), then subtract 1.

Geometric mean = $[(1 + R_1) \times (1 + R_2) \times \ldots \times (1 + R_n)]^{(1/n)} - 1$

£100 RISES 21%, THEN FALLS 19%

Year by year	£100 → £121 → £98.01	**a loss**
Growth factors	1.21 and 0.81	
Multiply	1.21 × 0.81	**0.9801**
Square root (2 years)	√0.9801	**0.99**
Geometric mean	0.99 − 1	**−1%**
Arithmetic mean, for contrast	(21 − 19) ÷ 2	**+1%**

Watch out: Using the arithmetic mean for returns over several years.

3. Spread: range, MAD, variance and standard deviation

- Spread (risk) is how far values sit from the middle. Range is the biggest minus the smallest and uses only two values. Mean absolute deviation is the average distance from the mean, ignoring signs.
- Variance squares each deviation (value minus mean) so ups and downs do not cancel. A sample divides by n − 1, because its own mean sits among its values and makes spread look smaller; a whole population divides by n. Standard deviation is the square root, in the data's units.
- Dividing a sample by n understates the risk.

> **$s^2 = \Sigma(x - \bar{x})^2 \div (n - 1)$. $s = \sqrt{s^2}$**

RETURNS 8, 12, −4, 16, 8 (MEAN 40 ÷ 5 = 8%)

Deviations	0, 4, −12, 8, 0	**sum 0**
Squared deviations, sum	0 + 16 + 144 + 64 + 0	**224**
Variance	224 ÷ (5 − 1)	**56**
Standard deviation	√56	**7.48%**
Range; MAD	16 − (−4); (0 + 4 + 12 + 8 + 0) ÷ 5	**20 points; 4.8%**
Wrongly dividing by n	√(224 ÷ 5)	**6.69%**

Watch out: Dividing by n when you have a sample.

4. Shape: skewness and kurtosis

- Skewness shows whether the shape leans to one side. Symmetric: skewness zero, mean = median = mode. Right (positive) skew: long tail of gains, mean above median, order mode, median, mean. Left (negative) skew: long tail of losses, mean below median. The tail pulls the mean.
- Kurtosis is how heavy the tails are compared with the normal distribution, whose kurtosis is 3. Excess kurtosis is kurtosis minus 3, so normal is 0. Above 0: fat tails (leptokurtic), with extreme gains and losses more common than the bell curve says. Below 0: thin tails.
- A risk limit built on a normal curve understates large losses when the tails are fat.

SKEW AND FAT TAILS

Ten monthly returns: −20, 2, 3, 4, 4, 4, 5, 5, 6, 7	sum 20 ÷ 10	**mean 2%**
Median; mode	(4 + 4) ÷ 2; 4 appears three times	**4%; 4%**
Verdict	mean below median	**left skew**
1,000 monthly returns, s = 4%: bad month	3 × 4%	**12 points below**
Normal curve predicts / observed	0.13% of 1,000	**about 1 / 9 months**
Verdict	excess kurtosis > 0	**fat tails**

Watch out: Getting the skew direction backwards, and mixing up kurtosis (normal = 3) with excess kurtosis (normal = 0).

5. Position: quantiles

- A quantile cuts a sorted set into groups with the same number of values. Quartiles give 4 groups, deciles 10, percentiles 100. The median is the 2nd quartile, 5th decile and 50th percentile.
- The interquartile range, Q3 minus Q1, is the spread of the middle half and ignores the extremes.
- Sort the data, then find the position L: n is the count, y the percentile wanted. If L is a whole number, take that value. If it has a fraction, go to the value just below and add that fraction of the gap to the next value.

L = (n + 1) × y ÷ 100. Interquartile range = Q3 − Q1

SORTED: −8, −3, 1, 4, 6, 7, 9, 11, 14, 18, 25 (N = 11)		
Q1 (L = 12 × 0.25)	L = 3	**1%**
Median (L = 12 × 0.50)	L = 6	**7%**
Q3 (L = 12 × 0.75)	L = 9	**14%**
Interquartile range	14 − 1	**13 points**
90th percentile	L = 10.8: 18 + 0.8 × 7	**23.6%**
Add a 12th year of +120%: IQR	Q1 = 1 + 0.25 × 3 = 1.75; Q3 = 14 + 0.75 × 4 = 17	**15.25 points**
Standard deviation		**9.38 to 33.65**

Watch out: Quartiles resist outliers; standard deviation does not.

6. Downside risk and risk per unit of return

- Standard deviation treats upside surprises like downside ones. Semi-deviation uses only returns below the mean: square the shortfalls, add them, divide by n − 1 where n counts all observations, then take the square root. Measuring shortfall against a target such as zero is target semi-deviation.
- The coefficient of variation (CV) is risk per unit of return. It is a pure number with no units, and a lower CV is better. If the mean is close to zero, the CV becomes huge and unreliable.

Semi-deviation = √[Σ(shortfall)² ÷ (n − 1)]. CV = standard deviation ÷ mean

FUND A (−12, 4, 6, 10, 12) AND FUND B (20, 4, 2, −2, −4): MEAN 4%, S = 9.49%		
A: shortfall squared	−12 is 16 below: 16^2	**256**
A: semi-deviation	√(256 ÷ 4)	**8.00%**
B: shortfalls squared	$2^2 + 6^2 + 8^2$	**104**
B: semi-deviation	√(104 ÷ 4)	**5.10%**
CV of A and B	9.49 ÷ 4	**2.37**
Fund X (mean 10, s 15); fund Y (mean 6, s 6)	15 ÷ 10; 6 ÷ 6	**1.50; 1.00 (Y)**

Watch out: In semi-deviation, dividing by only the number of bad years.

Key takeaways

- One number is never enough: describe returns by centre, spread, shape and position.
- Use the geometric mean for returns over several periods; the arithmetic mean overstates.
- Sample variance divides by n − 1; standard deviation is its square root.
- Mean above median means right skew; positive excess kurtosis means fat tails (normal = 0).
- Percentile position is L = (n + 1) × y ÷ 100, and the interquartile range resists outliers.
- Semi-deviation looks only below the mean; CV = standard deviation ÷ mean, and lower is better.

Quick check

1. Mean return 6%, median 2%. What is the skew?
   Answer: Right skewed: mean above median, long tail of high values.
2. Mean 5%, standard deviation 10%. What is the CV?
   Answer: 10 ÷ 5 = 2.0 risk per unit of return.
3. What does positive excess kurtosis mean?
   Answer: Fat tails: extreme outcomes more likely than the normal curve says.