CHARTA LEARNING · LESSON NOTES

Sample Moments and Estimators

FRM Part I

1 A story: tasting the soup

- Population = the whole pot; sample = one spoonful
- Your verdict from the spoonful = an estimate
- A different spoonful gives a different estimate
- Risk work: estimate the truth from a short history

IN THE LESSON

- Imagine you cook a huge pot of soup and want to know how salty it is.
- You cannot drink the whole pot.
- So you stir it and taste one spoonful.
- That spoonful is your sample, the whole pot is the population, and your verdict is an estimate.
- Another spoonful would taste a little different, so your verdict wobbles from spoonful to spoonful.
- Risk managers live in exactly this situation.
- The true behavior of a market is never fully seen.
- We only see a short history of returns, and from it we must estimate averages, spreads, tail behavior and relationships.
- This lesson is about doing that honestly and knowing how much to trust it.

2 Sample mean, variance and standard deviation

$\bar{x} = \Sigma x \div n$

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

Sample mean	Sum of data divided by n
Sum of squared deviations	Squared gaps from the sample mean
Divisor n - 1	One fewer than the data count
Sample variance	Average squared gap, adjusted

IN THE LESSON

- Here are the three basic estimates.
- The sample mean is the sum of the observations divided by how many there are.
- The sample variance takes each observation, subtracts the sample mean, squares the gap, adds up all the squares, and then divides by the number of observations minus one.
- The sample standard deviation is just the square root of that variance, which puts the spread back into the original units.
- The squares keep negative and positive gaps from canceling.
- Notice the divisor.
- It is not the number of observations.
- That small change is deliberate, and the next slide explains why it is there.

3 Why divide by n minus 1

- The sample mean is fitted to the same data
- So gaps from it are slightly too small
- Divisor n: biased low; n − 1 corrects the variance
- Even so, s itself is still slightly biased low
- Shortcut: SS = $\Sigma x^2 - n \times \bar{x}^2$

IN THE LESSON

- Why not divide by the number of observations?
- Because the sample mean is not the true mean.
- It is fitted to the very same data, so it sits right in the middle of the sample, as close to the points as any number can be.
- Gaps measured from it are therefore slightly too small, and a variance built from them with the plain count would come out too low on average.
- Dividing by one less enlarges the answer just enough to fix this.
- If you only know the sum of the data and the sum of the squares, the centered sum of squares is the sum of squares minus the count times the sample mean squared.
- A common slip is using the plain count.
- Another is forgetting that the standard deviation is still slightly biased low.

4 Example 1: five daily returns

Illustration: five daily returns (%): 4, −2, 6, 0, 2

Sample mean	(4−2+6+0+2) ÷ 5	**2%**
Sum of squared gaps	4 + 16 + 16 + 4 + 0	**40**
Sample variance	40 ÷ (5 − 1)	**10**
Sample std deviation	√10	**3.16%**

IN THE LESSON

- Let's use made up numbers.
- Five daily returns, in per cent: four, minus two, six, zero and two.
- The sum is ten, so the sample mean is ten divided by five, which is two per cent.
- Now the gaps from two: two, minus four, four, minus two and zero.
- Square them: four, sixteen, sixteen, four and zero, which add to forty.
- The sample variance divides by the count minus one, which is four, so forty over four gives ten.
- The sample standard deviation is the square root of ten, about three point one six per cent.
- So a typical daily return in this little sample sits roughly three points from its average of two per cent.

5 Example 2: from sums, and n versus n minus 1

Illustration: only totals known: n = 5, Σx = 10, Σx^2 = 60

Sample mean	10 ÷ 5	**2%**
Centered sum of squares	60 − 5 × 2²	**40**
Variance, divisor n − 1	40 ÷ 4	**10**
Variance, divisor n	40 ÷ 5	**8**
Ratio of the two	8 ÷ 10 = (n−1) ÷ n	**0.8**

IN THE LESSON

- Sometimes you are given only totals.
- Same data: the count is five, the sum of the observations is ten, and the sum of the squares is sixty.
- The mean is ten over five, two.
- The centered sum of squares is sixty minus five times two squared, which is sixty minus twenty, forty, matching what we found by hand.
- Dividing by four gives a variance of ten.
- Dividing by five instead gives eight.
- Eight is eighty per cent of ten, which is the count minus one over the count.
- So the plain count undershoots by a fifth here.
- The shortcut saves effort, but only works if you keep the sums accurate.

6 Population moments and sample moments

- Moment = average of a power: x, x^2, $(x - \text{mean})^2$
- Population moment: a fixed constant of the distribution
- Sample moment: same recipe computed on n data points
- A sample moment is random: it changes with the sample
- It is used to estimate the population moment

IN THE LESSON

- Now let's name the building blocks.
- A moment is an average of a power of the variable.
- For a population, the mean is the expected value of the variable, and the second raw moment is the expected value of its square.
- Central moments subtract the true mean first.
- These are fixed constants, set by the underlying distribution.
- A sample moment is the same recipe applied to data: the average of the observations, of their squares, or of deviations from the sample mean, over the number of observations.
- Because the data are random, a sample moment is itself a random variable.
- It changes from sample to sample and is used as an estimator of the population moment.
- Do not confuse the fixed target with the random statistic aiming at it.

7 Population versus sample moments

Same idea, but one is fixed and one is random

Moment	Population | Sample
Mean	$E(X)$ | $\Sigma x \div n$
Second raw	$E(X^2)$ | $\Sigma x^2 \div n$ (unbiased)
Variance	$E[(X-\mu)^2]$ | $\Sigma(x-\text{mean})^2 \div (n-1)$
Nature	Fixed | Random

IN THE LESSON

- Here is the comparison side by side.
- The first raw moment is the mean: in the population it is the expected value of the variable, in a sample it is the average of the data.
- The second raw moment is the expected square, and its sample version, the average of the squares, is unbiased for it.
- The central versions are different.
- The population variance subtracts the true mean.
- The sample version subtracts an estimated mean, and so needs the adjustment of dividing by one less than the count to be unbiased.
- The population value is one fixed number.
- The sample value is a random number, with its own spread, which we will measure shortly with the standard error.

8 Estimator or estimate?

- Estimator = a rule applied to a random sample
- Estimate = the number the rule gives for one sample
- Estimator: sampling distribution, std error (SE)
- Bias, efficiency, consistency describe the estimator
- One miss does not prove the estimator is biased

IN THE LESSON

- Two words that sound alike mean different things.
- An estimator is a rule, a formula to be applied to a random sample.
- The sample mean, the sum of the observations divided by the count, is an estimator.
- Before any data are drawn its value is uncertain, so it has a sampling distribution, with an expected value, a variance and a standard error.
- An estimate is the single number the rule produces once a particular sample has been observed, such as point zero three per cent from two hundred and fifty daily returns.
- Properties like bias, efficiency and consistency belong to the estimator, because they describe behavior across many possible samples.
- One estimate missing the target never proves the estimator is biased.

9 Sorting rules from values

If it talks about many samples, it is about the estimator

Statement	Which one?
Σx ÷ n	Estimator (a rule)
0.03% from 250 returns	Estimate (a value)
The rule is unbiased	Property of the estimator
This sample's mean is high	Estimate off in one sample
Standard error	Spread of the estimator

IN THE LESSON

- Let's practice telling them apart.
- The phrase sum of the observations divided by the count describes a rule, so it is an estimator.
- The number point zero three per cent, computed from last year's data, is an estimate.
- Saying this estimator is unbiased is a claim about the rule across many samples, so it is a property of the estimator.
- Saying our estimate of the mean is above the truth is a claim about one outcome, which is just an estimate being off in this sample.
- Finally, the standard error is a property of the estimator, because it measures how much the rule would vary from sample to sample.
- If a statement is about many hypothetical samples, it concerns the estimator.

10 Bias of an estimator

Bias = E(estimator) − θ
Divisor-n variance: bias = −σ² ÷ n

Bias	Systematic overshoot or undershoot
Estimator's average	Average over many samples
True parameter	The true value being estimated
Precision	Unbiased does not mean precise

IN THE LESSON

- Bias asks whether an estimator is aimed at the right spot.
- Imagine applying the rule to an enormous number of fresh samples and averaging all the estimates.
- Bias is that average minus the true parameter.
- Zero bias means the rule is right on average, though a single sample may still miss.
- It is a systematic tendency to overshoot or undershoot, not the error in one sample.
- The sample mean is unbiased, and so is the sample variance with the count minus one as divisor.
- The plain count version has a bias of minus the true variance divided by the sample size.
- Be careful: zero bias does not mean precise, since an unbiased estimator can still have a large variance.

11 Example 3: bias of the divisor-n variance

Illustration: true variance σ² = 100, sample size n = 5

Average, divisor n − 1	unbiased: E = σ²	**100**
Average, divisor n	100 × (5 − 1) ÷ 5	**80**
Bias of divisor n	80 − 100	**−20**
Check with −σ² ÷ n	−100 ÷ 5	**−20**
Bias if n = 50	−100 ÷ 50	**−2**

IN THE LESSON

- Suppose the true variance of returns is one hundred, and each sample has five observations.
- The divisor n minus one estimator averages one hundred across many samples, so it is unbiased.
- The divisor n estimator shrinks every answer by four fifths, so it averages one hundred times four over five, which is eighty.
- Its bias is eighty minus one hundred, minus twenty.
- The formula agrees: minus the true variance over the sample size, minus one hundred over five, minus twenty.
- With a larger sample of fifty, the bias is only minus two.
- The bias fades as the sample grows, a fact that will matter when we reach consistency.
- This is an illustration with made up numbers.

12 Example 4: a biased estimator can win

MSE = variance + bias². Illustration: two estimators of one parameter

A: bias 0, variance 4	MSE = 4 + 0²	**4**
B: bias 1, variance 2	MSE = 2 + 1²	**3**
Better overall	smaller MSE wins	**B**

IN THE LESSON

- Unbiased sounds ideal, but it is not the only goal.
- The yardstick that combines accuracy and spread is the mean squared error.
- It equals the variance of the estimator plus its bias squared.
- Consider two rival estimators of the same parameter, with made up numbers.
- Estimator A is unbiased, with a variance of four, so its mean squared error is four plus zero, four.
- Estimator B has a bias of one but a smaller variance of two.
- Its mean squared error is two plus one squared, which is three.
- The slightly biased estimator is better overall, because the lower variance more than pays for the small bias.
- Unbiasedness alone never settles which estimator to use.

13 What BLUE means

- B: best = smallest variance
- L: linear = a weighted sum of the observations
- U: unbiased = the weights add up to one
- Variance = $\sigma^2 \times \Sigma(\text{weights}^2)$
- Smallest when every weight is 1 ÷ n: the sample mean

IN THE LESSON

- You will often hear that the sample mean is BLUE.
- The letters stand for Best Linear Unbiased Estimator of the population mean.
- Linear means the estimator is a weighted sum of the observations.
- Unbiased means the weights add up to one, so that its expected value equals the true mean.
- Best means it has the smallest variance among all such weighted sums.
- For uncorrelated observations sharing a common mean and variance, the variance of a weighted sum is the common variance times the sum of the squared weights.
- That sum is smallest when every weight is one over the sample size, which is exactly the sample mean.
- No assumption of normality is needed for this result.

14 Example 5: equal weights win

Illustration: 4 uncorrelated observations, common variance 16

Equal weights, 0.25 each	$\Sigma w^2 = 4 \times 0.0625$	**0.25**
Variance, equal weights	16 × 0.25	**4**
Weights 0.4, 0.3, 0.2, 0.1	Σw^2 = .16+.09+.04+.01	**0.30**
Variance, uneven weights	16 × 0.30	**4.8**

IN THE LESSON

- Let's check with four uncorrelated observations whose common variance is sixteen, made up for illustration.
- Equal weights are a quarter each.
- The squared weights are each one sixteenth, and four of them add to a quarter.
- The variance of the sample mean is sixteen times a quarter, which is four.
- Now try different weights that still add to one: point four, point three, point two and point one.
- Their squares are point one six, point zero nine, point zero four and point zero one, adding to point three.
- The variance is sixteen times point three, four point eight.
- That is higher.
- Any uneven weighting that sums to one does worse than equal weights, which is what best means here.

15 Where the BLUE claim stops

- Compares linear estimators only
- Median is nonlinear: can beat the mean in fat tails
- Unequal variances: equal weights are no longer best
- Correlated: equal weights no longer guaranteed best
- No normality needed when the setup holds

IN THE LESSON

- The claim is narrow, so know its limits.
- It only compares linear estimators.
- It says nothing about nonlinear ones such as the median, and with heavy tailed data the median can be more precise than the mean.
- It also depends on its setup.
- If the observations have different variances, or if they are correlated with each other, equal weights are no longer the best choice, and the mean loses its title.
- So when you hear that the mean is BLUE, ask three questions.
- Are the observations uncorrelated?
- Do they share a common mean and variance?
- Are we only comparing weighted sums?
- If the answers are yes, the sample mean is the best of that class, and no normal distribution was required.

16 Consistency

- Consistent: converges in probability to the true value
- Sufficient: bias and variance both shrink to zero
- Justifies estimators biased in small samples
- Not the same as unbiased
- Says nothing about accuracy in one finite sample

IN THE LESSON

- An estimator is consistent if it converges in probability to the true parameter as the sample size grows without limit.
- In plain words, for any tolerance you choose, the chance of missing by more than that tolerance shrinks to zero as you collect more data.
- A sufficient condition is that both the bias and the variance go to zero as the sample grows.
- Consistency is useful because it justifies estimators that are biased in small samples, like the plain count variance or the sample standard deviation, provided the data set is large.
- It is not the same as unbiasedness.
- The first observation alone is an unbiased estimator of the mean, yet it is not consistent, since its variance never shrinks.
- And consistency says nothing about accuracy in any particular finite sample.

17 Consistent versus merely unbiased

Illustration: variance of one observation = 100

Sample size n	Variance: first obs. | mean
25	100 | 4
100	100 | 1
400	100 | 0.25
1,600	100 | 0.0625
Divisor-n variance	bias $-\sigma^2 \div n \to 0$: consistent

IN THE LESSON

- Here is the contrast, with a made up spread.
- Say the variance of a single observation is one hundred.
- The first observation on its own is unbiased, but its variance is one hundred whatever the sample size, so it never improves.
- The sample mean has variance one hundred over the sample size.
- At twenty five observations that is four.
- At one hundred it is one.
- At four hundred it is a quarter, and at sixteen hundred it is point zero six two five.
- Both bias and variance vanish, so the mean is consistent.
- Meanwhile the plain count variance has a bias of minus the variance over the sample size, which also vanishes, so it is biased yet consistent.

18 The LLN and the CLT

- LLN: the sample mean converges to the population mean
- CLT: the standardized mean is approximately normal
- CLT needs a finite variance; any data shape works
- Intervals and tail probabilities come from the CLT
- LLN is where it ends up; CLT is how it is spread

IN THE LESSON

- Two great results describe the sample mean, and they answer different questions.
- The Law of Large Numbers says that for independent, identically distributed observations with a finite mean, the sample mean converges to the population mean as the sample grows.
- It tells you where the estimator ends up.
- The Central Limit Theorem needs a finite variance as well.
- It says the standardized sample mean is approximately standard normal for large samples, whatever the shape of the data.
- It tells you how the estimator is spread around its target.
- So confidence intervals and tail probabilities for an average come from the Central Limit Theorem.
- A usual error is claiming the Law of Large Numbers makes averages normal.
- Another is forgetting that the result is only approximate.

19 Standard error and the z statistic

$SE = \sigma \div \sqrt{n}$
$z = (\bar{x} - \mu) \div (\sigma \div \sqrt{n})$

Standard error	Typical miss of the sample mean
Sample mean	The mean found in the sample
Population mean	The true mean (the hypothesis)
Standardized deviation	Approximately standard normal

IN THE LESSON

- Here is the working formula.
- The standard error of the mean is the standard deviation of the data divided by the square root of the sample size.
- It measures how far the sample mean typically sits from the true mean.
- The standardized mean subtracts the true mean from the sample mean and divides by the standard error.
- By the Central Limit Theorem, this is approximately a standard normal variable.
- Because of the square root, quadrupling the sample size only halves the standard error.
- To be ten times as precise you need a hundred times the data.
- In practice the true standard deviation is unknown and replaced by the sample value, which is fine when the sample is large.

20 Example 6: an interval for the mean

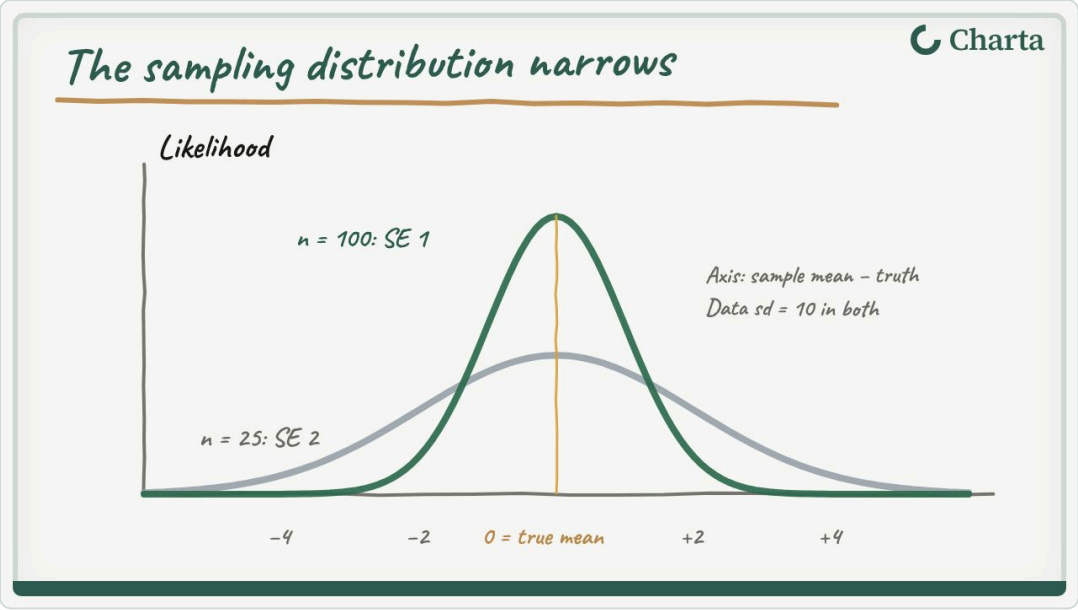
Illustration: daily $\sigma = 2\%$, $n = 100$, sample mean 0.5%

Standard error	$2\% \div \sqrt{100}$	**0.2%**
95% margin	$1.96 \times 0.2\%$	**0.392%**
95% interval	$0.5\% \pm 0.392\%$	**0.11% to 0.89%**
z to test mean = 0	$0.5 \div 0.2$	**2.5**
Standard error if n = 400	$2\% \div \sqrt{400}$	**0.1%**

IN THE LESSON

- Suppose daily returns have a standard deviation of two per cent, and we have one hundred of them with a sample mean of point five per cent.
- These are made up numbers.
- The standard error is two over the square root of one hundred, which is ten, so point two per cent.
- A ninety five per cent interval uses one point nine six standard errors, a margin of point three nine two.
- The interval runs from about point one one to point eight nine per cent.
- Is the true mean zero?
- The statistic is point five over point two, two point five, beyond one point nine six, so zero is outside.
- Quadruple the data to four hundred and the standard error halves to point one.

21 The sampling distribution narrows



IN THE LESSON

- Here is the Central Limit Theorem in a picture.
- Each curve shows how the sample mean would vary across many samples, with the true mean at the center.
- The wide, lower curve is for twenty five observations, where the standard error is two.
- The tall, narrow curve is for one hundred observations, where the standard error is one.
- Both are bell shaped, even if the raw data were not.
- Both are centered on the truth because the sample mean is unbiased, and the narrowing as the sample grows is the Law of Large Numbers at work.
- Most estimates from the larger sample land within two units of the truth, while the smaller sample often misses by a wider margin.

22 Skewness and kurtosis from data

$z = (x - \text{mean}) \div sd$
$\text{Skew} = \Sigma z^3 \div n$, $\text{Kurt} = \Sigma z^4 \div n$

Standardized deviation	Gap from the mean, in std devs
Standard deviation	Std deviation (divisor n here)
Excess kurtosis	Kurtosis minus 3
Standard error	SE: skew $\sqrt{(6 \div n)}$, kurt $\sqrt{(24 \div n)}$

IN THE LESSON

- Next, the shape of the data.
- First standardize every observation: subtract the mean and divide by the standard deviation, giving a z score.
- Sample skewness is the average of the cubed z scores.
- Cubing keeps the sign, so a long left tail makes it negative, which means large losses are more extreme than large gains, as often seen in equity returns.
- Sample kurtosis is the average of the fourth powers.
- It equals three for a normal distribution, so excess kurtosis is kurtosis minus three, and positive values mean fat tails with more extreme outcomes than normal.
- Small samples make both statistics noisy, and one outlier can dominate them.

23 Example 7: skewness and kurtosis

Illustration: returns (%): −10, 2, 2, 2, 2, 2. Mean 0

Standard deviation	√(120 ÷ 6)	**4.47**
Skewness	(−960 ÷ 6) ÷ 4.47³	**−1.79**
Kurtosis	(10,080 ÷ 6) ÷ 4.47⁴	**4.2**
Excess kurtosis	4.2 − 3	**1.2**

IN THE LESSON

- Take six made up returns: minus ten, and then five of plus two.
- The mean is zero.
- Using the plain count as the divisor, as this illustration does, the squared gaps are one hundred plus five fours, one hundred and twenty, and divided by six that is twenty, so the standard deviation is four point four seven.
- The cubed gaps add to minus nine hundred and sixty.
- Divide by six and by the standard deviation cubed, and skewness comes out at minus one point seven nine: a long left tail.
- The fourth powers add to ten thousand and eighty.
- Divided by six and by the standard deviation to the fourth, kurtosis is four point two, an excess of one point two.

24 Reading skewness and kurtosis

- Skew < 0: heavier left tail (big losses); > 0: right
- Kurtosis = 3 is normal; excess = kurtosis − 3
- Excess > 0: fat tails, more extreme outcomes
- Normal data: SE(skew) = √(6÷n), SE(kurt) = √(24÷n)
- Small samples are noisy; one outlier can dominate

IN THE LESSON

- How do we read these numbers?
- Negative skewness signals a longer or heavier left tail, positive skewness a heavier right tail, and zero is symmetric.
- For kurtosis, compare with three.
- Above three, or positive excess kurtosis, means fat tails: rare large moves happen more often than a normal distribution allows.
- Always subtract three if a statement talks about excess.
- But also ask whether the number means anything.
- For normal data the standard error of sample skewness is about the square root of six over the sample size, and for kurtosis the square root of twenty four over the sample size.
- With one hundred observations those are about point two four and point four nine, so modest readings are just noise.

25 Example 8: sample quantiles and the median

Ten returns (%): 3, −5, 1, 8, −2, 4, 0, −9, 6, 2

Sorted data	−9,−5,−2,0,1,2,3,4,6,8	**10 values**
Median (n even)	(1 + 2) ÷ 2	**1.5%**
20% quantile	ceil(10 × 0.2) = 2nd	**−5%**
25% quantile	ceil(10 × 0.25) = 3rd	**−2%**
10% quantile (VaR style)	ceil(10 × 0.1) = 1st	**−9%**

IN THE LESSON

- Quantiles come from the ordered data, so sort first.
- Ten made up returns in per cent: three, minus five, one, eight, minus two, four, zero, minus nine, six and two.
- Sorted, they run minus nine, minus five, minus two, zero, one, two, three, four, six, eight.
- With an even count, the median is the average of the two middle values, one and two, so one point five.
- For a general quantile, one convention takes the observation at the position of the ceiling of the count times the probability.
- For the twenty per cent quantile that is the second smallest, minus five.
- For twenty five per cent it is the ceiling of two point five, the third, minus two.
- For ten per cent, the first, minus nine, as in a historical simulation of value at risk.

26 Quantiles resist outliers, means do not

Same ten returns; one value changes from 8 to 80

	Mean | Median
Original data	0.8 | 1.5
Replace 8 by 80	8.0 | 1.5
Change	+7.2 | 0

IN THE LESSON

- Why do risk people like quantiles and the median?
- Because they depend on ranks, not on the size of extreme values.
- Look at the same ten returns.
- The mean is point eight and the median is one point five.
- Now suppose the eight were a data error and read eighty.
- The mean leaps to eight, a jump of seven point two, since the mean moves with every extreme value.
- The median is still one point five, because the ordering of the middle observations did not change.
- That robustness is a strength.
- The cost is that the median ignores the size of the tails.
- The typical errors are rounding the position the wrong way and counting from the wrong end, so check which convention the question states.

27 Example 9: the CLT for two sample means

Illustration: $\sigma X^2 = 4$, $\sigma Y^2 = 9$, covariance 3 ($\rho = 0.5$), $n = 100$

Variance of each mean	4 ÷ 100 and 9 ÷ 100	**0.04, 0.09**
Covariance of the means	3 ÷ 100	**0.03**
Variance of difference	(4 + 9 − 2×3) ÷ 100	**0.07**
Variance of sum	(4 + 9 + 2×3) ÷ 100	**0.19**
Wrong: ignore covariance	(4 + 9) ÷ 100	**0.13**

IN THE LESSON

- Often we estimate the means of two variables from the same pairs of observations.
- Then the Central Limit Theorem applies jointly.
- For a large sample the pair of means is approximately bivariate normal, centered on the true means, with the population covariance matrix divided by the sample size.
- Take made up numbers: variances of four and nine, a covariance of three, which is a correlation of point five, and one hundred pairs.
- The two means have variances point zero four and point zero nine, and covariance point zero three.
- The difference has variance four plus nine minus six, over one hundred, point zero seven.
- The sum has four plus nine plus six, point one nine.
- Ignoring the covariance would give point one three, which is wrong.

28 Sample covariance and correlation

$\text{Cov} = \Sigma(x - \bar{x})(y - \bar{y}) \div (n - 1)$
$\text{Corr} = \text{Cov} \div (s_x \times s_y)$

Covariance	Do they move together? Unit-bound
Sample standard deviations	Each variable's own spread
Correlation	Unit-free, between −1 and 1
Divisor n - 1	Same divisor in all three
Linear dependence	Linear only: 0 is not independent

IN THE LESSON

- Now two variables observed together.
- The sample covariance multiplies, for each pair, the gap of the first variable from its mean by the gap of the second from its mean, adds these cross products, and divides by the count minus one.
- It is positive when the variables tend to be above or below their means together, and its size depends on the units.
- The sample correlation divides the covariance by the product of the two sample standard deviations.
- It is unit free and lies between minus one and one.
- From totals, the centered cross sum is the sum of the products minus the product of the two sums over the count.
- If you use the plain count in the covariance, use it in the variances too, and it cancels.

29 Example 10: covariance and correlation

Illustration: pairs (x, y): (2,1) (4,5) (6,3) (8,7)

Means of x and y	20 ÷ 4 and 16 ÷ 4	**5 and 4**
Sum of cross-products	9 − 1 − 1 + 9	**16**
Sample covariance	16 ÷ (4 − 1)	**5.33**
Std dev of each	√(20 ÷ 3)	**2.58**
Correlation	5.33 ÷ (2.58 × 2.58)	**0.80**

IN THE LESSON

- Four made up pairs: two and one, four and five, six and three, eight and seven.
- The mean of the first variable is five and of the second is four.
- The cross products of the gaps are nine, minus one, minus one and nine, which add to sixteen.
- The shortcut agrees: the sum of the products is ninety six, minus twenty times sixteen over four, eighty, gives sixteen.
- The sample covariance is sixteen over three, five point three three.
- Each variable has squared gaps adding to twenty, a variance of twenty over three and a standard deviation of two point five eight.
- The correlation is five point three three divided by two point five eight squared, which is point eight.
- Strong, positive, and linear.

30 Coskewness and cokurtosis

- $S(X,X,Y) = E[(X-\mu_x)^2(Y-\mu_y)] \div (\sigma_x^2\,\sigma_y)$
- $K(X,X,Y,Y) = E[(X-\mu_x)^2(Y-\mu_y)^2] \div (\sigma_x^2\,\sigma_y^2)$
- S(X,X,X) = skewness of X; K(X,X,X,X) = kurtosis of X
- Y = −X: an odd number of Y factors flips the sign
- n variables, order k: C(n + k − 1, k) moments

IN THE LESSON

- Skewness and kurtosis describe one variable.
- Coskewness and cokurtosis extend them to several, which matters for the tail risk of a portfolio.
- Coskewness with the first variable twice and the second once is the average of the squared gap of the first times the gap of the second, divided by the variance of the first times the standard deviation of the second.
- Cokurtosis with each variable twice uses squared gaps of both, over both variances.
- When every argument is the same variable they collapse to the single variable measures: three copies is its skewness, four copies its kurtosis.
- If the second variable is the negative of the first, an odd number of its factors flips the sign.
- For several variables the count of distinct moments of order k is n plus k minus one, choose k.

31 Example 11: coskewness and moment counts

Illustration: pairs (x, y): (−2,−3) (0,0) (1,1) (1,2). Means are 0

Variances of x and y	6 ÷ 4 and 14 ÷ 4	**1.5, 3.5**
Average of $x^2 \times y$	(−12+0+1+2) ÷ 4	**−2.25**
Denominator	1.5 × √3.5	**2.81**
Coskewness S(X,X,Y)	−2.25 ÷ 2.81	**−0.80**
Moments, 3 variables	C(5,3) and C(6,4)	**10 and 15**

IN THE LESSON

- Take four equally likely pairs, made up for illustration: minus two and minus three, zero and zero, one and one, one and two.
- Both means are zero.
- The first variable has variance six over four, one point five.
- The second has fourteen over four, three point five.
- The product of the squared first value and the second value averages to minus twelve plus zero plus one plus two, over four, minus two point two five.
- The denominator is one point five times the square root of three point five, two point eight one.
- So coskewness is minus point eight: the first variable's big moves tend to coincide with downward moves of the second.
- Moment counts: with three variables there are ten of order three and fifteen of order four.

32 Four mistakes to avoid

- Divided by n: use n − 1 for a sample variance
- One estimate off: bias is the average miss over samples
- LLN makes it normal: no, that is the CLT
- Excess kurtosis = kurtosis − 3; skew < 0 is left tail

IN THE LESSON

- Mistake one: dividing the squared gaps by the count when you want the variance of a sample.
- Use the count minus one, and remember the standard deviation is still slightly biased low.
- Mistake two: judging an estimator by one estimate.
- A single miss is not bias, because bias is the average miss over many samples, and a zero bias does not guarantee precision.
- Mistake three: saying the Law of Large Numbers makes averages normal.
- It says the average converges, and the Central Limit Theorem gives the approximate normal shape.
- Mistake four: forgetting to subtract three from kurtosis to get the excess, or misreading the sign of skewness.
- Negative skewness means the heavier tail is on the left.

33 Recap: the main ideas

- s^2 uses n − 1; estimator = rule, estimate = its value
- Bias = average miss; MSE = variance + bias²
- Mean is BLUE; consistent = converges with data
- LLN: where it ends up. CLT: SE = σ ÷ √n, normal shape
- Skew, kurtosis, quantiles, Cov, Corr, co-moments

IN THE LESSON

- Let's pull it together.
- Sample moments are random statistics used to estimate fixed population moments.
- The sample variance divides by one less than the count.
- An estimator is a rule and an estimate is its value for one sample.
- Bias is the average miss, mean squared error adds variance and squared bias, and the sample mean is the best linear unbiased estimator under uncorrelated, equal variance data.
- Consistency means converging with more data.
- The Law of Large Numbers says where the mean ends up and the Central Limit Theorem gives its approximate normal spread, with a standard error that shrinks with the square root of the sample size.
- Then skewness, kurtosis, quantiles, covariance, correlation and the higher co moments.

34 Try three quick questions

1. Sum of squared deviations 12, n = 4. s^2 and s?
   Answer: 12 ÷ (4 − 1) = 4, so s = 2.
2. σ = 12 and n = 36. SE? n to halve it?
   Answer: SE = 12 ÷ 6 = 2. Use n = 144 for SE = 1.
3. n = 100, skew −0.6, SE = √(6÷n). Beyond 1.96 SE?
   Answer: SE = 0.245; z ≈ −2.45, beyond 1.96: yes.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- Four observations have a sum of squared deviations from their mean of twelve.
- What are the sample variance and standard deviation?
- Twelve over three is four, and the square root is two.
- Question two.
- The standard deviation of the data is twelve and there are thirty six observations.
- What is the standard error, and how many observations halve it?
- Twelve over six is two, and a quadrupled sample of one hundred and forty four halves it to one.
- Question three.
- With one hundred observations, a sample skewness of minus point six, and a standard error of the square root of six over the sample size.
- Is it beyond one point nine six standard errors from zero, the five per cent level?
- The standard error is point two four five, so minus point six is about minus two point four five standard errors, beyond one point nine six.
- Yes.