CHARTA LEARNING · LESSON NOTES

Linear Regression

FRM Part I

1 A story: the line through the cloud

- A cloud of dots: size of each home against its price
- One straight line through the cloud summarizes it
- Its tilt answers: how much more for one more unit?
- Regression finds the best line; OLS is the standard way

IN THE LESSON

- Picture an estate agent with a notebook full of sales.
- For each home she has written down the size and the price.
- Plot them and you get a cloud of dots, rising to the right but never in a perfect line, because location, age and the mood of the buyer also matter.
- She wants one straight line through the cloud, a line she can use to say that a home of this size should sell for about this much.
- The tilt of the line answers a practical question: how much more does one extra unit of size cost?
- Regression is the method for finding the best such line, and ordinary least squares is the standard way to do it.

2 Linear in the parameters

- Linear means linear in the parameters (a, b, c)
- X^2 is just another regressor: a quadratic model is fine
- Parameters must add up: no exponents, ratios, products
- Test the parameters, not the variables

IN THE LESSON

- Start with what the method can handle.
- Ordinary least squares, which we will call O L S, needs a model that is linear in its parameters.
- Many people hear the word linear and think of a straight line, but that is the usual mistake.
- What matters is how the unknown coefficients enter the equation.
- If they simply add up, each one multiplied by some variable, the model qualifies, even when the variables themselves are squared or logged.
- A squared variable is just another regressor with its own coefficient.
- What breaks the method is a coefficient sitting inside an exponent, or inside a ratio that cannot be removed, or multiplied by another coefficient.
- So test the parameters, not the variables.

3 Which models can OLS estimate?

What counts is how the parameters enter, not the variables

Model	OLS?
$Y = a + bX + cX^2 +$ error	Yes: parameters additive
$Y = A \times X^b \times$ error	Yes, after taking logs
$\ln Y = \ln A + b \ln X +$ error	Yes: linear in ln A, b
$Y = a + X^b +$ error	No: b sits in an exponent
$Y = a + a \times b \times X +$ error	No: nonlinear least squares

IN THE LESSON

- Here are the cases side by side.
- First, the quadratic model: the outcome equals a constant, plus a coefficient times the variable, plus another coefficient times its square, plus an error.
- The three coefficients enter additively, so O L S works.
- Second, the multiplicative model, where the outcome equals a constant times the variable raised to a power, times an error.
- As written it is not additive, but taking natural logarithms turns it into an additive equation.
- Third, that log form itself is linear in the log of the constant and the power, so it qualifies after the transformation.
- Fourth, a constant plus the variable raised to an unknown power: the power sits in an exponent and no transformation removes it.
- Fifth, parameters that multiply each other.
- Both of those need nonlinear least squares.

4 Example 1: the logarithm trick

Illustration: $Y = A \times X^b$ with $A = 2$, $b = 1.5$, $X = 4$

Level of Y	$2 \times 4^{1.5} = 2 \times 8$	**16**
Log of both sides	ln 16	**2.7726**
Right side in logs	$\ln 2 + 1.5 \times \ln 4$	**0.6931 + 2.0794**
Sum matches the left side	0.6931 + 2.0794	**2.7726**
Parameters enter additively	ln A and b are the unknowns	**OLS works**

IN THE LESSON

- Let's check the logarithm trick with made up numbers.
- Suppose the constant is two, the power is one point five, and the variable equals four.
- The outcome is two times four to the power one point five, and four to that power is eight, so the outcome is sixteen.
- Take the natural log of sixteen: two point seven seven two six.
- Now the right side in logs.
- The log of two is zero point six nine three one, and one point five times the log of four, which is one point three eight six three, gives two point zero seven nine four.
- Add them and you land on the same two point seven seven two six.
- The unknowns have become additive, so ordinary least squares can estimate them.

5 The OLS slope and intercept

Slope = Cov(X, Y) ÷ Var(X)
Intercept = mean(Y) − Slope × mean(X)
Fitted value = Intercept + Slope × X

Slope	Change in Y per one unit of X
Covariance of X and Y	How X and Y move together
Variance of X	Spread of X, not of Y
Intercept	Fitted Y when X is zero

IN THE LESSON

- Now to the estimates.
- With one explanatory variable, O L S chooses the intercept and slope that make the sum of squared residuals as small as possible.
- A residual is the gap between an actual outcome and the fitted line, and squaring stops positive and negative gaps from cancelling.
- The answer has a neat form.
- The slope equals the covariance of the two variables divided by the variance of the explanatory variable.
- Notice, the explanatory variable's variance, not the outcome's.
- The intercept then follows from the sample means: the mean outcome minus the slope times the mean of the explanatory variable.
- A fitted value is the intercept plus the slope times the variable.

6 Example 2: computing the slope

Illustration: X = 1, 2, 3, 4, 5 and Y = 2, 4, 5, 4, 5

Means	mean X = 3, mean Y = 4	**3 and 4**
Covariance	$\Sigma(X-3)(Y-4) \div 4 = 6 \div 4$	**1.5**
Variance of X	$\Sigma(X-3)^2 \div 4 = 10 \div 4$	**2.5**
Slope	$1.5 \div 2.5$	**0.6**
Intercept	$4 - 0.6 \times 3$	**2.2**

IN THE LESSON

- Let's compute it.
- Illustration: five observations.
- The explanatory variable runs from one to five, and the outcome is two, four, five, four, five.
- The means are three and four.
- Multiply each pair of deviations and add them up: the products sum to six, and with the sample divisor of four, the covariance is one point five.
- The squared deviations of the explanatory variable sum to ten, so its variance is two point five.
- The slope is one point five over two point five, which is zero point six.
- The intercept is four minus zero point six times three, or two point two.
- So each extra unit of the variable goes with zero point six more of the outcome.

7 Example 3: fitted values and residuals

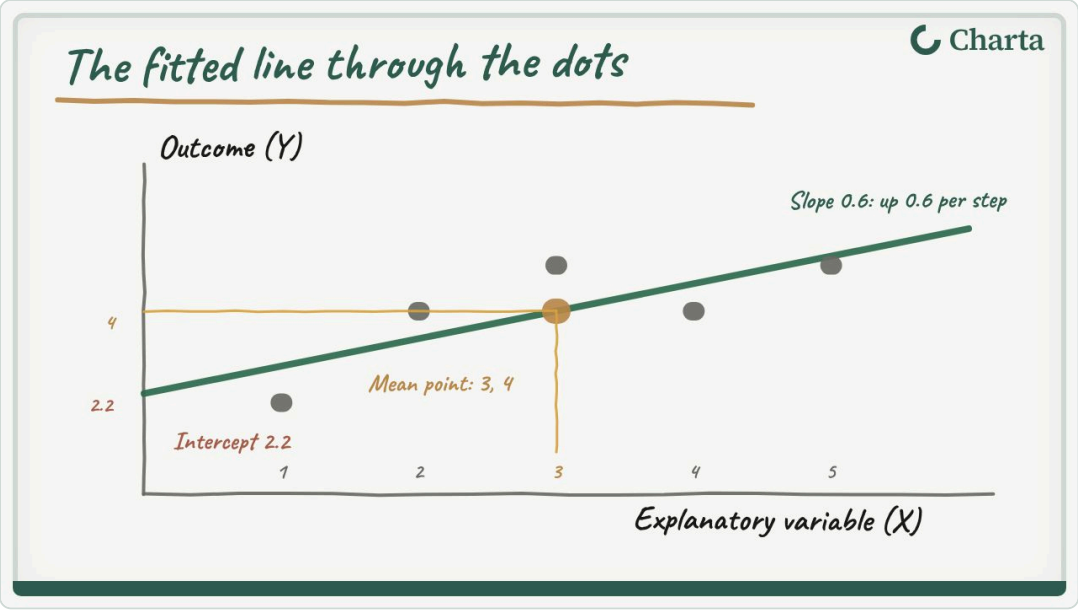
Fitted value = 2.2 + 0.6 × X

Fitted at X = 1	$2.2 + 0.6 \times 1$	**2.8**
Residual at X = 1	$2 - 2.8$	**−0.8**
Fitted at X = 3, the mean	$2.2 + 0.6 \times 3$	**4.0 = mean Y**
Sum of the five residuals	$-0.8+0.6+1.0-0.6-0.2$	**0**

IN THE LESSON

- Now use the line.
- At the first observation the fitted value is two point two plus zero point six, which is two point eight.
- The actual outcome was two, so the residual is two minus two point eight, minus zero point eight.
- Next the middle observation, at the mean of the explanatory variable, which is three.
- The fitted value is two point two plus one point eight, which is four, exactly the mean of the outcome.
- That is no accident: the fitted line always passes through the point of sample means.
- Finally, compute all five residuals and add them: minus zero point eight, plus zero point six, plus one, minus zero point six and minus zero point two.
- They sum to zero, as they always do when the model has an intercept.

8 The fitted line through the dots



IN THE LESSON

- Here is the picture.
- Along the bottom is the explanatory variable, and up the side the outcome.
- The five dots are our observations.
- The line starts at two point two on the vertical axis, which is the intercept, and rises by zero point six for every step to the right.
- Notice it passes exactly through the point of means, three and four.
- Some dots sit above the line, with positive residuals, and some below, with negative ones.
- Least squares chooses this particular line because no other line gives a smaller total of squared vertical gaps.
- The line summarizes the relationship, but it does not explain every dot, and the gaps are the part it leaves out.

9 Reading the estimates in words

- Slope: average change in Y per one-unit change in X
- Units matter: Y units per X unit; rescale X, slope rescales
- Intercept: fitted Y at X = 0; weak if 0 is outside the data
- Fitted line always passes through the point of means

IN THE LESSON

- Reading the estimates correctly is half the skill.
- The slope is an average change: when the explanatory variable rises by one unit, the outcome changes by the slope, on average, in the outcome's own units.
- If the outcome were measured in dollars and the variable in basis points, the slope would be dollars per basis point, and rescaling a variable rescales the slope.
- The intercept is the fitted outcome when the variable is zero.
- If zero lies far outside the data, as it would for house size, the intercept is just a device that positions the line and means little.
- The usual slips are dividing by the wrong variance, or forgetting the units of the explanatory variable.

10 The assumptions behind OLS

- 1. Linear in the parameters
- 2. Zero conditional mean of the error (exogeneity)
- 3. Observations independent, identically distributed
- 4. Large outliers unlikely; X varies across observations
- 5. Constant error variance (normality: exact small-n t)

IN THE LESSON

- Now the small set of assumptions behind all this.
- One: the model is linear in its parameters.
- Two: the error has a zero conditional mean given the explanatory variable, which is called exogeneity.
- Whatever the value of the variable, it tells you nothing about the error's average.
- Three: the observations are independent and identically distributed draws, which is a statement about how the data were sampled.
- Four: large outliers are unlikely, meaning the variables have finite fourth moments, and the explanatory variable actually varies, because if every observation had the same value, the slope could not be computed.
- Five, for the classical results: the error variance is constant, called homoskedasticity.
- Normal errors are needed only for exact small sample t tests.

11 Exogeneity and omitted variables

- Example: pay on schooling, with ability left out
- Ability hides in the error and is tied to schooling
- Then the error's mean depends on X: the slope is biased
- iid is about sampling, not the same as exogeneity
- Heteroskedasticity does not bias the slope

IN THE LESSON

- Exogeneity deserves a closer look, because it is what protects the slope from bias.
- Imagine we regress pay on years of schooling.
- Natural ability affects pay, and it is also related to schooling.
- If ability is left out of the model, it hides inside the error, and the error is no longer unrelated to schooling.
- The slope then absorbs some of ability's effect and is biased.
- That is omitted variable bias.
- Notice the independence assumption is a different thing.
- It concerns how the sample was drawn, not whether the error is related to the variable.
- And note that a non constant error variance, called heteroskedasticity, does not bias the slope.
- It affects the standard errors instead.

12 What each assumption buys you

Break the first row and the slope is biased; break the last and only the SEs suffer

Assumption	What it protects
Zero conditional mean	The slope is unbiased
Independent, identical draws	The sampling results
No large outliers	Dependable estimates, SEs
X varies	The slope can be computed
Constant error variance	Efficiency, simple SEs

IN THE LESSON

- Here is what each assumption buys.
- Zero conditional mean protects against bias, and it is the one that matters most.
- Independent and identically distributed draws justify the sampling results we use later.
- The absence of large outliers keeps the estimates and their standard errors dependable, because one extreme point can swing a fitted line a long way.
- Variation in the explanatory variable is needed simply so the slope exists.
- Constant error variance gives the efficiency result and the simple formula for standard errors.
- Without it the slope is still unbiased, but the usual standard errors are wrong and need correcting.
- And normal errors matter only for exact small sample tests.

13 Example 4: omitted variable bias

Illustration: true model Y = 1 + 2X + 3Z + error; Z left out

True slope on X	set by the model	**2**
Effect of omitted Z on Y	coefficient on Z	**3**
Z moves with X	slope of Z on X	**0.5**
Slope if Z is left out	2 + 3 × 0.5	**3.5**
Bias	3.5 − 2	**+1.5**

IN THE LESSON

- Let's see the bias with made up numbers.
- Suppose the true model has a slope of two on the variable of interest, and a coefficient of three on a second variable that we forget to include.
- Suppose the second variable tends to rise with the first: regressing it on the first gives a slope of zero point five.
- When we leave it out, the estimated slope does not recover two.
- It picks up two plus three times zero point five, which is three point five.
- The bias is plus one point five.
- No amount of extra data fixes this, because the problem is built into the model, not into the sample size.

14 What the OLS estimators are like

- Unbiased: average over repeated samples = true slope
- Consistent: converges to the truth as n grows
- Homoskedastic errors: Gauss-Markov says OLS is BLUE
- Large n: slope is approximately normal (CLT)
- Unbiasedness needs no normality; exact t does

IN THE LESSON

- Now, what are the O L S estimators like?
- Under the assumptions they are unbiased: if you drew many samples and estimated the slope each time, the average of the estimates would equal the true slope.
- They are consistent: as the sample grows, the estimate converges to the truth.
- If the errors are also homoskedastic, the Gauss Markov theorem says that O L S is the best linear unbiased estimator, meaning it has the smallest variance among linear unbiased estimators.
- And by the central limit theorem, the sampling distribution of the slope is approximately normal in large samples, whatever the shape of the errors.
- Unbiasedness needs no normality at all.
- Exact t distributions do.

15 The standard error of the slope

SE(slope) ≈ σ_u ÷ (√n × σ_X)

Standard error	Precision of the slope estimate
Error standard deviation	Noise: bigger means worse
Sample size	More data: smaller SE
Spread of X	More spread in X: smaller SE

IN THE LESSON

- Because the slope changes from sample to sample, it has a standard deviation of its own, called its standard error.
- In large samples it is roughly the standard deviation of the error divided by the square root of the sample size times the standard deviation of the explanatory variable.
- Read the formula as three levers.
- More observations make the denominator bigger, so the standard error falls, though only with the square root of the sample size.
- A more spread out explanatory variable also helps, because it is easier to see the tilt of a line when the points stretch far apart.
- Noisier errors push the standard error up.
- Precision comes from many observations, wide variation and a tight fit.

16 Example 5: pulling the levers

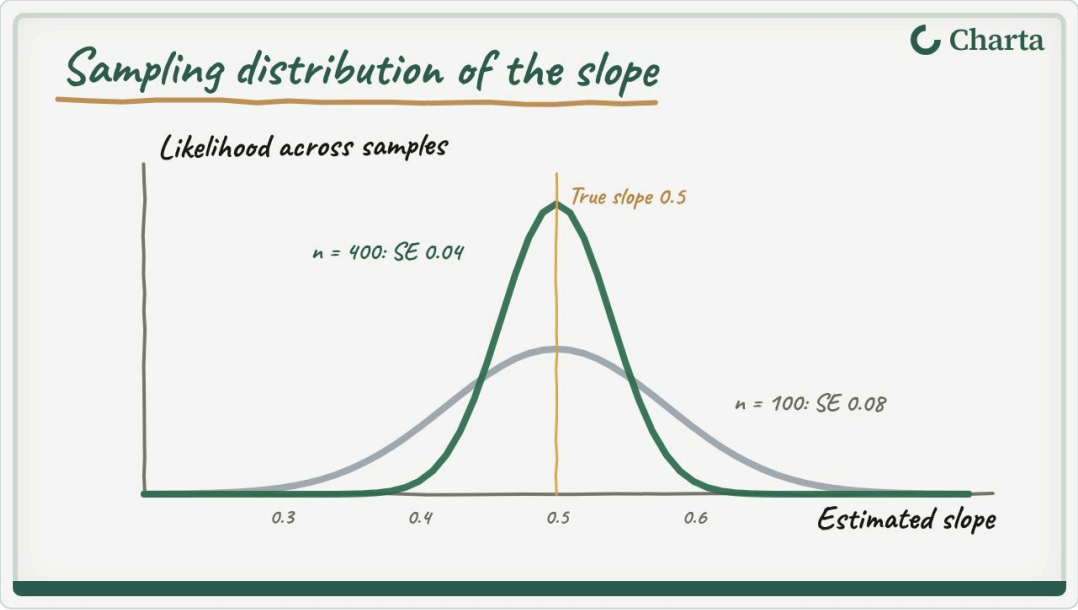
Illustration: σ_u = 4, σ_X = 5, n = 100 to start

Standard error, n = 100	4 ÷ (√100 × 5)	**0.08**
Four times the data	4 ÷ (√400 × 5)	**0.04**
X twice as spread out	4 ÷ (√100 × 10)	**0.04**
Errors twice as noisy	8 ÷ (√100 × 5)	**0.16**

IN THE LESSON

- Let's put numbers on those levers.
- Illustration: the error standard deviation is four, the standard deviation of the explanatory variable is five, and there are one hundred observations.
- The standard error is four over ten times five, which is zero point zero eight.
- Now quadruple the sample to four hundred.
- The square root is twenty, so four over twenty times five is zero point zero four.
- Doubling the data would not have been enough; we needed four times as much to halve the error.
- Alternatively, keep one hundred observations but spread the explanatory variable twice as widely, a standard deviation of ten: four over ten times ten, again zero point zero four.
- Finally, back at the first set up, double the noise to eight: eight over ten times five gives zero point one six.

17 Sampling distribution of the slope



IN THE LESSON

- Here is the sampling distribution.
- Suppose the true slope is zero point five.
- Each curve shows how the estimated slope would vary over repeated samples.
- With one hundred observations the curve is the wide one, with a standard error of zero point zero eight.
- With four hundred it is the tall, narrow one, with a standard error of zero point zero four.
- Both are centered on the truth, which is what unbiased means, and both are bell shaped, which is the central limit theorem at work.
- More data does not move the center, it squeezes the spread, so an estimate from the larger sample is likely to land closer to zero point five.

18 Testing one coefficient: t and the interval

t = (estimate − hypothesised value) ÷ SE
95% CI = estimate ± critical value × SE

t statistic	Distance from the claim, in SEs
Estimate	Your estimated coefficient
Hypothesised value	The value the null claims
Standard error	Standard error of the estimate
Critical value	1.96 large n; t if small

IN THE LESSON

- Now the main use of the standard error: testing a claim about one coefficient.
- The test statistic measures how far the estimate is from the claimed value, counted in standard errors.
- In large samples we compare it with normal critical values: one point nine six for a two sided test at five per cent, and two point five seven six at one per cent.
- In small samples with one regressor we use the t distribution with the sample size minus two degrees of freedom.
- A confidence interval is the estimate plus or minus the critical value times the standard error.
- A one sided alternative uses a single tail, so its critical value is smaller.
- Beware the default output, which tests against zero.
- If your claim is a beta of one, recompute the statistic yourself.

19 Example 6: is the beta equal to one?

Illustration: beta estimate 1.15, SE 0.10, large sample. H0: β = 1

Distance from the claim	1.15 − 1	**0.15**
t statistic	0.15 ÷ 0.10	**1.5**
Critical value, two-sided 5%	large sample	**1.96**
Decision	1.5 is inside ±1.96	**Do not reject**
95% interval	1.15 ± 1.96 × 0.10	**0.954 to 1.346**

IN THE LESSON

- Illustration: a regression of a fund's returns on the market gives a beta estimate of one point one five, with a standard error of zero point one zero, in a large sample.
- We want to test whether the true beta equals one.
- The distance from the claim is one point one five minus one, which is zero point one five.
- Divided by the standard error, the statistic is one point five.
- The critical value for a two sided five per cent test is one point nine six.
- Because one point five lies inside plus or minus one point nine six, we do not reject.
- The confidence interval agrees: one point one five plus or minus one point nine six times zero point one zero runs from zero point nine five four to one point three four six, and one sits comfortably inside.

20 Example 7: a small sample

Illustration: n = 27, one regressor, estimate 0.8, SE 0.35, H0: β = 0

Degrees of freedom	27 − 2	**25**
t statistic	(0.8 − 0) ÷ 0.35	**2.29**
Critical value, two-sided 5%	t with 25 degrees	**2.060**
Decision	2.29 > 2.060	**Reject**
95% interval	0.8 ± 2.060 × 0.35	**0.08 to 1.52**

IN THE LESSON

- Now a small sample.
- Illustration: twenty seven observations, one regressor, an estimate of zero point eight and a standard error of zero point three five, testing against zero.
- Degrees of freedom are twenty seven minus two, which is twenty five.
- The statistic is zero point eight over zero point three five, which is two point two nine.
- The critical value from the t distribution with twenty five degrees of freedom is two point zero six zero, a little larger than one point nine six, because small samples carry extra uncertainty.
- Two point two nine exceeds two point zero six zero, so we reject.
- The interval is zero point eight plus or minus two point zero six zero times zero point three five, from about zero point zero eight to one point five two.
- Zero is outside it, consistent with the rejection.

21 Critical values to know

Larger values mean a harder test to pass

Test	Critical value
Large n, two-sided 5%	1.96
Large n, two-sided 1%	2.576
Large n, one-sided 5%	1.645
df 25, two-sided 5%	2.060
df 25, one-sided 5%	1.708

IN THE LESSON

- Keep these critical values at your fingertips.
- For a large sample and a two sided test at five per cent, use one point nine six.
- At one per cent, two point five seven six.
- A one sided test at five per cent puts all of the five per cent in one tail, so the critical value drops to one point six four five.
- In a small sample the t distribution has fatter tails, so the values are larger.
- With twenty five degrees of freedom, the two sided five per cent value is two point zero six zero, and the one sided five per cent value is one point seven zero eight.
- As the degrees of freedom grow, the t values shrink toward the normal ones.

22 The stages of a hypothesis test

Five stages, always in this order

Stage	What you do
1 Hypotheses	Wanted claim goes in H1
2 Level	Choose before looking
3 Statistic	(estimate − claim) ÷ SE
4 Benchmark	Critical value, or p-value
5 Decide and say it	Compare, then use words

IN THE LESSON

- A hypothesis test on a coefficient follows a fixed sequence, and following it keeps you honest.
- First, state the null and the alternative, and put the claim you want evidence for in the alternative.
- Second, choose the significance level, the probability of rejecting a true null, before you look at the result.
- Third, compute the test statistic from the estimate and its standard error.
- Fourth, find the benchmark: the critical value for your level, tail and degrees of freedom, or alternatively the p-value.
- Fifth, compare and decide.
- Reject if the statistic falls beyond the critical value.
- Then finish by stating the conclusion in plain words, because a reader needs a sentence and not just a number.

23 Example 8: all the stages at work

Illustration: estimate 1.20, SE 0.12, large sample

1 Hypotheses	H0: β = 1, H1: β > 1	**One-sided**
2 Level	chosen before the data	**5%**
3 Statistic	(1.20 − 1) ÷ 0.12	**1.667**
4 Critical value	one-sided, large sample	**1.645**
5 Decision	1.667 > 1.645	**Reject H0**

IN THE LESSON

- Let's run all the stages.
- Illustration: we suspect a fund's beta exceeds one, so that goes in the alternative.
- The null says beta equals one, the alternative says beta is greater than one.
- We choose a level of five per cent before looking.
- The estimate is one point two zero with a standard error of zero point one two, so the statistic is zero point two over zero point one two, which is one point six six seven.
- A one sided five per cent test in a large sample has a critical value of one point six four five.
- One point six six seven is beyond it, so we reject the null.
- In words: there is evidence, at the five per cent level, that the fund's beta is above one.

24 Reading the verdict with care

- Fail to reject is not proof that the null is true
- A 5% level: about one in twenty true nulls is rejected
- Same data, two-sided: 1.667 < 1.96, so no rejection
- Choose the tail beforehand, from the question asked

IN THE LESSON

- Three points of care.
- First, failing to reject does not prove the null true; it only says the data did not give enough evidence against it.
- Second, a five per cent level means that about one in twenty true nulls will be rejected by chance alone.
- Third, the choice of tail matters.
- In our example the statistic of one point six six seven beat the one sided value of one point six four five, but a two sided test would need one point nine six, and would not reject.
- That is why the tail must be chosen beforehand, from the question being asked, and not picked afterwards to get the answer you like.

25 t, p-value and interval: one test

- One test of H0: β = b0, three ways to read it
- p-value below α ⇔ |t| above the critical value
- ⇔ b0 lies outside the (1 − α) confidence interval
- p-value: chance of a result this far out if H0 were true
- It is not the probability that the null is true

IN THE LESSON

- Now, three views of one test.
- Take a null that the coefficient equals some value, against a two sided alternative.
- The t statistic measures the distance in standard errors.
- The p-value is the probability, if the null were true, of getting an estimate at least as far from the claimed value as the one observed.
- And the confidence interval is the set of claimed values that the data would not reject.
- These always agree.
- Reject at level alpha when the p-value is below alpha.
- That is the same as the absolute t statistic exceeding the critical value.
- And that is the same as the claimed value lying outside the confidence interval.
- Do not read the p-value as the probability that the null is true.

26 Example 9: moving between the three views

Illustration: large sample, H0: β = 0, estimate 0.50, t = 2.0

t statistic	given	**2.0**
Two-sided p-value	2 × (1 − Φ(2.0))	**0.046**
Reject at 5%?	0.046 < 0.05, 2.0 > 1.96	**Yes**
Standard error	0.50 ÷ 2.0	**0.25**
95% interval	0.50 ± 1.96 × 0.25	**0.01 to 0.99**

IN THE LESSON

- Let's move between the three views.
- Illustration: a large sample, testing a coefficient against zero, with an estimate of zero point five zero and a t statistic of two.
- The p-value is two times one minus the normal cumulative probability at two, about zero point zero four six.
- That is below five per cent, and the statistic of two exceeds one point nine six, so we reject.
- Now recover the standard error: the estimate over the statistic, zero point five over two, is zero point two five.
- The interval is zero point five plus or minus one point nine six times zero point two five, from zero point zero one to zero point nine nine.
- Zero is outside it.
- All three views give the same verdict.
- Beware of using a one tailed p-value for a two sided test: it would be half as big.

27 R squared and the correlation

R^2 = ESS ÷ TSS = 1 − RSS ÷ TSS with TSS = ESS + RSS
r = ± √R^2, sign taken from the slope
Slope = r × σ_Y ÷ σ_X

R squared	Share of variation explained
Explained sum of squares	Explained variation
Residual sum of squares	Residual variation
Correlation	Size is √R^2, sign is the slope's

IN THE LESSON

- Finally, the link between R squared and correlation.
- R squared is the share of the variation in the outcome that the regression explains: the explained sum of squares divided by the total sum of squares.
- Total variation splits into an explained part and a residual part, so the sums of squares add up.
- With a single explanatory variable, R squared is exactly the square of the sample correlation between the two variables.
- So the size of the correlation is the square root of R squared.
- The sign comes from the slope, because the slope equals the correlation times the outcome's standard deviation over the explanatory variable's standard deviation, and standard deviations are positive.
- The same R squared results whichever variable you regress on the other.

28 Example 10: correlation from R squared

Our five observations: TSS 6, ESS 3.6, RSS 2.4, slope +0.6

Sums of squares add up	3.6 + 2.4	**6 = TSS**
R squared	3.6 ÷ 6	**0.60**
Size of the correlation	√0.60	**0.775**
Sign from the slope	slope is +0.6	**+0.775**
R^2 0.36, slope negative	−√0.36	**−0.60**

IN THE LESSON

- Back to our five observations.
- The total sum of squares was six, the explained part three point six and the residual part two point four.
- They add up: three point six plus two point four is six.
- R squared is three point six over six, which is zero point six zero.
- The size of the correlation is the square root of zero point six, about zero point seven seven five.
- The slope was positive, so the correlation is plus zero point seven seven five.
- Now a quick contrast.
- If R squared is zero point three six and the slope is negative, the correlation is minus the square root of zero point three six, minus zero point six zero.
- It is not zero point three six.
- This shortcut works only with one regressor.

29 Four mistakes to avoid

- Linear = straight line? No: test the parameters
- Cov ÷ Var(Y)? No: Cov ÷ Var(X), in Y per unit of X
- Heteroskedasticity biases slope? No: it hits SEs only
- Default null 0 fits any claim? No: recompute t; p ≠ P(null)

IN THE LESSON

- Four mistakes to avoid.
- One: reading linear as a straight line.
- Square and log the variables as you like; what must be linear is the way the parameters enter.
- Two: dividing the covariance by the variance of the outcome instead of the explanatory variable, or ignoring the units when reading the slope.
- Three: believing that heteroskedasticity biases the slope.
- It does not; it upsets the standard errors.
- Only a failure of zero conditional mean, such as an omitted variable, creates bias.
- Four: using the default test against zero when your claim is another value, or reading the p-value as the probability that the null is true.
- Recompute the statistic yourself, and remember the p-value describes the data, given the null.

30 Recap: the main ideas

- Linear in the parameters, not necessarily the variables
- Slope = Cov ÷ Var(X); line passes through the means
- Zero conditional mean keeps the slope unbiased
- SE = σ_u ÷ (√n × σ_X); t, p and interval tell one story
- One regressor: r = ± √R², sign from the slope

IN THE LESSON

- Let's pull it together.
- O L S can estimate any model that is linear in its parameters, even if the variables are squared or logged.
- With one explanatory variable the slope is the covariance over the variance of that variable, and the line passes through the sample means.
- Zero conditional mean is what keeps the slope unbiased, and with homoskedastic errors O L S is the best linear unbiased estimator.
- The standard error falls with more data and more spread, and the t statistic, the p-value and the confidence interval all tell one story.
- Test a claim in five stages and say the conclusion in words.
- And with one regressor, the correlation is the square root of R squared, with the sign of the slope.

31 Try three quick questions

1. Which can OLS estimate: a + bX + cX², a + X^b, ln Y on ln X?
   Answer: 1st and 3rd. In a + X^b the parameter b sits in an exponent.
2. Estimate 0.9, SE 0.2, H0: β = 0.5, large sample. Reject at 5%?
   Answer: t = 0.4 ÷ 0.2 = 2.0 > 1.96, so reject.
3. R² = 0.25 and the slope is negative. Correlation?
   Answer: −√0.25 = −0.50. Take the sign from the slope.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- Which of these can ordinary least squares estimate: a quadratic in the variable, a constant plus the variable raised to an unknown power, or the log of the outcome on the log of the variable?
- The first and the third, because their parameters enter additively.
- The second has a power that no transformation removes.
- Question two.
- An estimate is zero point nine with a standard error of zero point two, the claim is zero point five, and the sample is large.
- Do we reject at five per cent?
- The statistic is zero point four over zero point two, which is two, beyond one point nine six, so yes.
- Question three.
- R squared is zero point two five and the slope is negative.
- What is the correlation?
- Minus zero point five.