CHARTA LEARNING · LESSON NOTES

Regression with Multiple Explanatory Variables

FRM Part I

1 A story: what does a bedroom add to a price?

- Price on bedrooms alone: a big slope
- Bigger houses have more bedrooms and cost more anyway
- Add size to the model: the bedroom slope shrinks
- Multiple regression holds other variables fixed

IN THE LESSON

- Imagine you study house prices.
- You collect data and regress price on the number of bedrooms alone.
- The slope comes out large: each extra bedroom seems to add a great deal.
- But bigger houses have more bedrooms, and bigger houses cost more anyway.
- The simple regression gives bedrooms credit for the size of the house.
- Now put both size and bedrooms into the model.
- The bedroom coefficient shrinks, because it now answers a fairer question: what does one more bedroom add among houses of the same size?
- That is the whole motivation for multiple regression.
- It lets us hold other things fixed, in the data, when we cannot hold them fixed in the world.

2 Single regression: the assumptions

- Linear model; error has zero mean given X
- Constant error variance (homoskedasticity)
- Errors not correlated across observations
- X varies: it is not one constant value
- Small samples: normal errors for exact t and F tests

IN THE LESSON

- Start with what simple regression assumes.
- The relationship is linear in the coefficients.
- The error has an average of zero, given the explanatory variable, so the model is correct on average at every value.
- The error has constant variance, which we call homoskedasticity.
- Errors from different observations are not correlated with each other.
- And the explanatory variable actually varies, because a line cannot be fitted through observations that all share one value.
- With these in place, ordinary least squares, O L S, is unbiased and gives reliable standard errors.
- In small samples we also want the errors to be normally distributed, so that the t and F tests have their exact distributions.

3 Multiple regression: the same, plus one

- All the single-regression assumptions, given every regressor
- Zero mean error given all regressors together
- New: no perfect multicollinearity
- No regressor is an exact combination of the others
- Includes the constant: no regressor may equal a constant

IN THE LESSON

- Multiple regression keeps every one of those assumptions, but now they hold conditional on all the regressors together.
- The error must have a zero mean given every explanatory variable at once.
- Then it adds exactly one new assumption: there must be no perfect multicollinearity.
- That means no explanatory variable can be an exact linear combination of the others, and that includes the constant column that stands for the intercept.
- If one regressor is simply twice another, the data cannot tell their effects apart, and the O L S formulas cannot be computed at all.
- The software will either fail or silently drop a variable.
- This extra assumption is the single formal difference to remember.

4 Perfect versus imperfect multicollinearity

Both involve correlated regressors; only one breaks OLS

	What happens
Perfect: exact combination	OLS cannot be computed
Imperfect: strong correlation	Works; estimates unbiased
Effect of imperfect case	Standard errors rise
Effect on t statistics	They fall
Fix for the perfect case	Drop or redefine a variable

IN THE LESSON

- This distinction matters a lot, so keep the two cases apart.
- With perfect multicollinearity, one regressor is an exact combination of others.
- O L S simply cannot be computed, and the model must be fixed by dropping or redefining a variable.
- With high but imperfect multicollinearity, the regressors are strongly correlated but not exactly.
- Everything still works.
- The estimates remain unbiased.
- The cost is precision: the standard errors rise, so the t statistics fall and coefficients look insignificant even when they matter.
- The model as a whole can still fit very well.
- So perfect multicollinearity breaks the method, while imperfect multicollinearity only blurs the picture and leaves the estimates unbiased.

5 Bias versus unreliable standard errors

- Bias: zero conditional mean fails, e.g. omitted variable
- Omitted variable matters and correlates with one included
- Heteroskedasticity, autocorrelation: wrong std errors
- Fix for wrong std errors: robust standard errors
- Imperfect collinearity: unbiased, larger std errors

IN THE LESSON

- Be careful about what each problem damages.
- Bias comes from a failure of the zero conditional mean assumption.
- The classic cause is an omitted relevant variable that is correlated with an included one: the included variable then absorbs the missing variable's effect, and its coefficient is biased.
- Heteroskedasticity and autocorrelation are different.
- They leave the O L S coefficients unbiased, but the usual standard errors become unreliable, so tests and intervals mislead.
- The remedy is robust standard errors rather than a new estimator.
- And imperfect multicollinearity, as we saw, does neither: no bias, just larger standard errors.
- Matching each problem to its damage is an important skill.

6 Example 1: the dummy variable trap

Illustration: a model with an intercept and one dummy per region

Number of groups	east, west, south	**3**
Three dummies added up	east + west + south	**always 1**
The intercept column	a column of ones	**same column**
Dummies to include	3 groups − 1	**2**

IN THE LESSON

- Here is a classic way to hit perfect multicollinearity by accident.
- Suppose we want to capture three regions, east, west and south, so we create one dummy variable for each.
- A dummy is one when the observation belongs to the region and zero otherwise.
- Every observation belongs to exactly one region, so the three dummies add up to one in every single row.
- But the intercept is itself a column of ones.
- The three dummies together are therefore an exact copy of the intercept column, which is perfect multicollinearity, and the regression fails.
- The fix is to include one dummy fewer than the number of groups.
- Three groups need two dummies, and the left out group becomes the base group.

7 Reading a slope: a partial effect

- Slope = expected change in outcome, others held constant
- Only the included regressors are held fixed
- Intercept = expected outcome when all regressors are zero
- Often not meaningful: zero may be outside the data
- Differs from the simple-regression slope

IN THE LESSON

- Now to reading the output.
- In a multiple regression each slope is a partial effect.
- It is the expected change in the outcome for a one unit change in that regressor, holding all the other included regressors constant.
- The words included and constant matter: the slope says nothing about variables left out of the model.
- The intercept is the expected outcome when every regressor equals zero, which is often outside the data and so not meaningful.
- And because other variables are now held fixed, a slope usually differs from the slope in a simple regression on the same variable.
- The simple slope also picks up the effect of correlated omitted variables, as the bedroom story showed.

8 Example 2: why the simple slope differs

Illustration: Return = 1 + 0.9×Market + 0.4×Value, all in percent

Partial slope on market	value held fixed	**0.9**
Fitted at market 10, value 6	1 + 0.9×10 + 0.4×6	**12.4**
Value on market slope	from a second regression	**0.5**
Simple slope on market	0.9 + 0.4 × 0.5	**1.1**

IN THE LESSON

- Let's make this concrete with made up numbers.
- A fund's excess return is modelled on the market return and a value factor return.
- The multiple regression says: intercept one, a slope of point nine on the market, and point four on the value factor.
- With a market return of ten and a value return of six, the fitted return is one plus nine plus two point four, which is twelve point four.
- Now suppose the value return itself rises by point five for each unit of market return, from a second regression.
- If we leave value out, the market slope soaks up its effect: point nine plus point four times point five equals one point one.
- The simple slope of one point one overstates the pure market effect of point nine.

9 Dummies, interactions and rescaling

- Dummy coefficient = shift versus the base group
- Interaction (X × D) lets the slope differ by group
- Dummy-group slope = base slope + interaction coefficient
- Rescale a regressor: coefficient rescales inversely
- t statistics and R^2 do not change

IN THE LESSON

- Three more reading rules.
- First, a dummy variable coefficient is a shift in the conditional mean relative to the base group, the group whose dummy equals zero.
- Second, an interaction term multiplies a regressor by a dummy.
- It lets the slope differ by group.
- For the dummy group, the effect of the regressor is the base slope plus the interaction coefficient.
- Third, rescaling.
- If you measure a regressor in thousands instead of units, its coefficient is rescaled inversely, so it becomes a thousand times larger, while its t statistic and the R squared are unchanged.
- Changing units changes the coefficient's size but never the evidence behind it.

10 Example 3: a dummy with an interaction

Illustration: Loss = 1.0 + 0.5×U + 0.8×D + 0.3×(U×D)

Base group slope on U	the plain coefficient	**0.5**
Subprime slope on U	0.5 + 0.3	**0.8**
Base group loss at U = 4	1.0 + 0.5×4	**3.0**
Subprime loss at U = 4	1.0 + 0.8 + 0.8×4	**5.0**
Gap at U = 4	0.8 + 0.3×4	**2.0**

IN THE LESSON

- Take a loan loss rate.
- Unemployment, U, is the regressor and D is a dummy equal to one for subprime loans.
- The fitted model is: one plus point five times U, plus point eight times D, plus point three times the product of U and D.
- For the base group, ordinary loans, the slope on unemployment is simply point five.
- For subprime loans it is the base slope plus the interaction, point five plus point three, which is point eight.
- Now take unemployment of four.
- The base group loses one plus point five times four, which is three.
- The subprime group loses one plus point eight plus point eight times four, which is five.
- The gap is two, the dummy plus the interaction times four.

11 Two groups, two slopes



IN THE LESSON

- Here are the two fitted lines.
- Along the bottom is unemployment, from zero to eight.
- Up the side is the loss rate.
- The grey line is the base group.
- It starts at one on the vertical axis and climbs at point five per unit.
- The green line is the subprime group.
- It starts higher, at one point eight, because of the dummy, and climbs more steeply, at point eight per unit, because of the interaction.
- The two lines are not parallel.
- At unemployment of four the base group sits at three and the subprime group at five, so the vertical gap is two.
- A dummy alone would shift the line up; the interaction is what changes the tilt, so the gap grows as unemployment rises.

12 Splitting the variation in the outcome

- TSS = Σ(Y − mean of Y)²: all the variation in Y
- ESS: variation captured by the fitted values
- SSR: variation left in the residuals
- With an intercept: TSS = ESS + SSR
- R^2 = ESS ÷ TSS = 1 − SSR ÷ TSS, between 0 and 1

IN THE LESSON

- Now to goodness of fit, and it starts with a decomposition.
- The total sum of squares, T S S, measures all the variation in the outcome: the sum of squared deviations of each observation from the mean.
- It splits into two parts.
- The explained sum of squares, E S S, is the variation captured by the fitted values.
- The sum of squared residuals, S S R, is the variation left over in the errors.
- When the regression includes an intercept, these add up exactly: T S S equals E S S plus S S R.
- R squared is the explained share of the total, E S S over T S S, which is also one minus S S R over T S S.
- It always lies between zero and one.

13 Example 4: R squared and the standard error

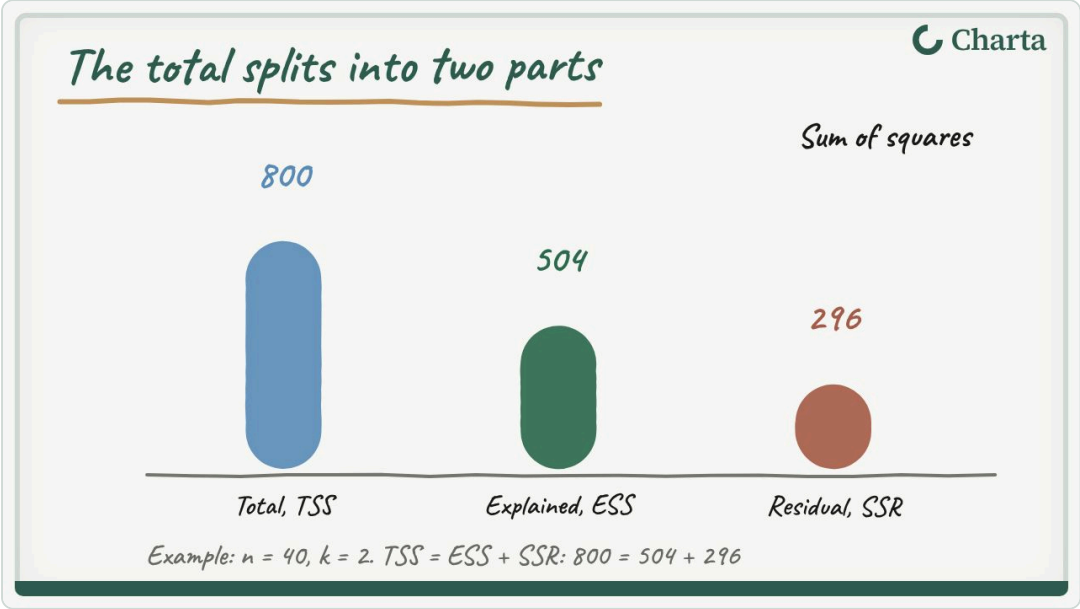
Illustration: n = 40, k = 2, TSS = 800, SSR = 296

Explained sum, ESS	800 − 296	**504**
R squared	504 ÷ 800	**0.63**
Degrees of freedom	40 − 2 − 1	**37**
Standard error of regression	√(296 ÷ 37) = √8	**2.83**

IN THE LESSON

- Let's compute it.
- A regression has forty observations and two regressors.
- The total sum of squares is eight hundred, and the sum of squared residuals is two hundred and ninety six.
- The explained sum of squares is the difference, five hundred and four.
- R squared is five hundred and four over eight hundred, which is point six three.
- So the model explains sixty three percent of the variation.
- For the standard error of the regression we need the degrees of freedom: forty minus two minus one is thirty seven.
- Divide the residuals sum by thirty seven to get eight, and take the square root: two point eight three.
- That is the typical size of a residual, in the units of the outcome.

14 The total splits into two parts



IN THE LESSON

- Here is the same example as bars.
- The heights are sums of squares, and every bar starts from zero.
- The tall bar is the total, eight hundred.
- It is fixed by the outcome variable alone, so it does not change when we add regressors to the same sample.
- The green bar is the explained part, five hundred and four, and the red bar is the unexplained part, two hundred and ninety six.
- Together they rebuild the total.
- Now picture adding another regressor.
- The total stays at eight hundred, the residual bar can only shrink or stay put, and the explained bar can only grow or stay put.
- That is exactly why R squared never falls when a variable is added.

15 Example 5: converting an analysis of variance table

Illustration: n = 61, k = 4. Regression: df 4, MS 84. Residual: df 56, MS 4

Explained sum, ESS	84 × 4	**336**
Residual sum, SSR	4 × 56	**224**
Total sum, TSS	336 + 224	**560**
R squared	336 ÷ 560	**0.60**
Standard error of regression	√4	**2.0**

IN THE LESSON

- A common trap comes from the analysis of variance table.
- That table often reports mean squares, which are sums of squares divided by their degrees of freedom, not the sums themselves.
- Say the regression row has four degrees of freedom and a mean square of eighty four, and the residual row has fifty six degrees of freedom and a mean square of four.
- First convert back.
- Explained sum: eighty four times four, three hundred and thirty six.
- Residual sum: four times fifty six, two hundred and twenty four.
- The total is their sum, five hundred and sixty.
- Now R squared is three hundred thirty six over five hundred sixty, point six.
- And the standard error of the regression is the square root of the residual mean square, which is two.

16 What R squared means and what it hides

- R^2 = share of the variation in Y that is explained
- Simple regression: R^2 = (corr of X and Y)^2
- OLS with intercept: R^2 = (corr of Y and fitted Y)^2
- Adding a regressor never lowers R^2
- Does not show unbiased coefficients or a correct model

IN THE LESSON

- So what does R squared tell us?
- It is the fraction of the variation in the outcome that the regression explains.
- In a simple regression it equals the squared correlation between the explanatory variable and the outcome.
- In any O L S regression with an intercept it equals the squared correlation between the outcome and its fitted values.
- But R squared has limits.
- Adding a regressor never lowers it, even if the new variable is pure noise, so R squared cannot be used to choose between models with different numbers of variables.
- And neither R squared nor its adjusted version shows that the coefficients are unbiased or that the model is correct.
- A high value can sit beside a badly specified model.

17 Adjusted R squared

Adj $R^2 = 1 - (1 - R^2) \times (n - 1) \div (n - k - 1)$

R squared	Share of variation explained
Observations, n	Sample size
Regressors, k	Number of explanatory variables
Adjusted R squared	Penalized for extra variables

IN THE LESSON

- Adjusted R squared fixes the one-way street of R squared by charging a price for each added variable.
- Start with the unexplained share, one minus R squared.
- Multiply it by the number of observations minus one, divided by the number of observations minus the number of regressors minus one.
- That ratio is bigger than one, so it scales the unexplained share up.
- Subtract the result from one.
- Adjusted R squared is below R squared whenever there is at least one regressor and R squared is below one, and it can even be negative.
- It rises when an added variable has a t statistic above one in absolute value, and falls otherwise.
- Watch the denominator: it is n minus k minus one, not n minus k.

18 Example 6: computing adjusted R squared

Illustration: n = 61, k = 4, R squared = 0.60

Ratio of degrees	(61 − 1) ÷ (61 − 4 − 1)	**60 ÷ 56**
Unexplained share	1 − 0.60	**0.40**
Adjusted R squared	1 − 0.40 × 60 ÷ 56	**0.571**
Wrong: n − k denominator	1 − 0.40 × 60 ÷ 57	**0.579**

IN THE LESSON

- Let's compute it with the numbers from before: sixty one observations, four regressors and an R squared of point six.
- The two degrees of freedom terms are sixty one minus one, which is sixty, and sixty one minus four minus one, which is fifty six.
- Their ratio is about one point zero seven.
- The unexplained share is one minus point six, point four.
- Multiply: point four times one point zero seven gives about point four three.
- Adjusted R squared is one minus that, point five seven one.
- Compare that with R squared of point six: the penalty shaved about three points.
- As a warning, if you wrongly used sixty one minus four in the denominator, you would get point five seven nine.

19 Adding variables: R squared versus adjusted

Same sample: n = 61, TSS = 560. Each row adds one regressor

k = 1	R^2 0.411, adjusted	**0.401**
k = 2	R^2 0.518, adjusted	**0.501**
k = 3	R^2 0.580, adjusted	**0.558**
k = 4: adjusted peaks	R^2 0.600, adjusted	**0.571**
k = 5: adjusted drops	R^2 0.602, adjusted	**0.566**

IN THE LESSON

- Now watch the two measures as we add regressors one at a time to the same sample of sixty one observations, with a total sum of squares of five hundred and sixty.
- With one regressor, R squared is point four one and adjusted is point four zero.
- With two, point five two and point five zero.
- With three, point five eight and point five six.
- With four, point six zero and point five seven.
- With a fifth regressor, the residual sum barely moves, so R squared edges up to point six zero two, but adjusted R squared falls to point five six six.
- R squared climbs with every variable.
- Adjusted R squared peaks at four and then drops, which is the signal that the fifth variable added too little.

20 The F test for joint hypotheses

F = [(SSR_r − SSR_u) ÷ q] ÷ [SSR_u ÷ (n − k − 1)]

Restricted SSR	Residual sum, claim imposed
Unrestricted SSR	Residual sum of the full model
Restrictions, q	Number of restrictions
Degrees of freedom	n − k − 1, from the full model

IN THE LESSON

- Sometimes the claim involves several coefficients at once, for example that two of them are both zero.
- For that we use the F test.
- Estimate the model twice: the unrestricted model, and the restricted model with the claim imposed.
- The restricted model always fits a little worse, so its residual sum is larger.
- The F statistic is the increase in the residual sum, divided by the number of restrictions, over the unrestricted residual sum divided by its degrees of freedom, n minus k minus one.
- Reject when F exceeds the critical value for q and n minus k minus one degrees of freedom.
- A big increase in the residual sum means the restrictions hurt the fit, which is evidence against them.

21 Example 7: testing two coefficients together

Illustration: n = 61, k = 4, SSR unrestricted 224, restricted 260

Increase in SSR	260 − 224	**36**
Numerator	36 ÷ 2	**18**
Denominator	224 ÷ 56	**4**
F statistic	18 ÷ 4	**4.5**
Versus critical 3.16	(2, 56) df, 5%	**Reject**

IN THE LESSON

- Use the model with sixty one observations and four regressors, with an unrestricted residual sum of two hundred and twenty four.
- We claim two coefficients are both zero, so q is two.
- Impose that by dropping the two variables, and the residual sum rises to two hundred and sixty.
- The increase is thirty six.
- Divide by two restrictions to get eighteen.
- For the denominator, divide two hundred twenty four by fifty six degrees of freedom to get four.
- F is eighteen over four, four point five.
- A table gives a five percent critical value of about three point one six for two and fifty six degrees of freedom.
- Four point five is larger, so we reject: the two coefficients are jointly significant.

22 Example 8: the overall regression F test

Illustration: n = 61, k = 4, R squared = 0.60; restrictions q = 4

Explained, per regressor	0.60 ÷ 4	**0.15**
Unexplained, per df	0.40 ÷ 56	**0.00714**
F statistic	0.15 ÷ 0.00714	**21.0**
Same from the ANOVA table	84 ÷ 4	**21**

IN THE LESSON

- There is a special joint test that asks whether the regression explains anything at all, that is, whether all slope coefficients are zero.
- Here q equals the number of regressors, four.
- When the dependent variable is the same in both models, we can write F with R squared.
- The restricted model has only an intercept, so its R squared is zero.
- Take our R squared of point six.
- The numerator is point six over four, which is point one five.
- The denominator is one minus point six, point four, over fifty six, about point zero zero seven one four.
- The ratio is twenty one.
- That matches the analysis of variance table, mean square eighty four over four, also twenty one.
- Far above any critical value, so we reject.

23 Why not run separate t-tests?

- Separate t-tests ignore how the estimates move together
- Several tests also inflate the false-rejection rate
- Symptom of multicollinearity: every t is small
- ...yet the joint F test rejects strongly
- Variables matter jointly; the data cannot split the credit

IN THE LESSON

- A tempting shortcut is to test each coefficient with its own t test and combine the answers.
- Do not.
- The individual tests ignore how the estimates move together, and running several tests also raises the chance of a false rejection.
- A joint claim needs a joint test.
- There is a striking symptom to know.
- When regressors are highly correlated, every individual t statistic can be small, say around one point two, so none looks significant, while the F test on all of them together strongly rejects.
- The variables jointly matter, but the data cannot say which one deserves the credit.
- That pattern is a typical sign of multicollinearity, and it is the reason the F test exists.

24 Example 9: a confidence interval for a slope

Illustration: estimate 0.9, standard error 0.3, n = 61, k = 4, df = 56

t statistic against zero	0.9 ÷ 0.3	**3.0**
Critical t, 95%, 56 df	about 2.00	**2.00**
Margin	2.00 × 0.3	**0.60**
95% interval	0.9 ± 0.6	**0.3 to 1.5**
Is a slope of 1.0 rejected?	0.3 < 1.0 < 1.5	**No**

IN THE LESSON

- Last, a confidence interval.
- Take a slope estimate of point nine with a standard error of point three, from the model with sixty one observations and four regressors, so fifty six degrees of freedom.
- The t statistic against zero is point nine over point three, which is three.
- The interval is the estimate plus or minus the critical t value times the standard error.
- For ninety five percent confidence and fifty six degrees of freedom, the critical t is about two.
- Two times point three is a margin of point six.
- The interval runs from point three to one point five.
- Now ask whether the true slope could be one.
- One sits inside the interval, so we do not reject that claim.

25 Four mistakes to avoid

- Adjusted R^2: denominator is $n - k - 1$, not $n - k$
- More variables always lift R^2; compare adjusted R^2
- A joint claim needs one F test, not separate t tests
- Imperfect collinearity: wider errors; perfect: no OLS

IN THE LESSON

- Mistake one: using the wrong denominator in adjusted R squared.
- It is the number of observations minus the number of regressors minus one, not minus the number of regressors alone.
- Mistake two: choosing between models of different sizes on R squared.
- It rises with every added variable, so compare adjusted R squared, and remember that neither shows the model is correct.
- Mistake three: testing a joint claim with separate t tests.
- Estimate both models and use one F test.
- Mistake four: mixing up the two kinds of multicollinearity.
- Perfect multicollinearity breaks O L S, while the imperfect kind leaves estimates unbiased and only widens standard errors.

26 Recap: the main ideas

- Same assumptions plus no perfect multicollinearity
- Slope = partial effect; dummy shifts, interaction tilts
- TSS = ESS + SSR; R^2 = ESS ÷ TSS; compare adjusted R^2
- Joint claims: F test; interval = estimate ± t × std error
- Bias: omitted variables. Imperfect collinearity: wider errors

IN THE LESSON

- Let's pull it together.
- Multiple regression keeps the single regression assumptions, now conditional on every regressor, and adds no perfect multicollinearity.
- Imperfect multicollinearity leaves estimates unbiased and widens standard errors.
- Bias comes from omitted variables correlated with included ones.
- Each slope is a partial effect, with the other included variables held constant, and dummies and interactions shift the mean and the slope for a group.
- Total variation splits into explained and residual parts.
- R squared is the explained share and never falls; adjusted R squared penalizes extra variables.
- Joint hypotheses need an F test, and a confidence interval is the estimate plus or minus the critical t times the standard error.

27 Try three quick questions

1. n = 50, k = 4, R^2 = 0.5. Adjusted R^2?
   Answer: 1 − 0.5 × 49 ÷ 45 = 0.456.
2. SSR_u 100, SSR_r 130, q = 3, n = 64, k = 5. F?
   Answer: (30 ÷ 3) ÷ (100 ÷ 58) = 5.8.
3. One regressor is exactly twice another. OLS?
   Answer: Cannot be computed: perfect multicollinearity.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- A model has fifty observations, four regressors and an R squared of point five.
- What is the adjusted R squared?
- One minus point five times forty nine over forty five gives about point four five six.
- Question two.
- The unrestricted residual sum is one hundred, the restricted one is one hundred and thirty, there are three restrictions, sixty four observations and five regressors.
- What is F?
- Thirty over three is ten, divided by one hundred over fifty eight, about five point eight.
- Question three.
- If one regressor is exactly twice another, can O L S be computed?
- No: that is perfect multicollinearity.