Regression Inference

Ani Katchova

© 2020 by Ani Katchova. All rights reserved.

Overview

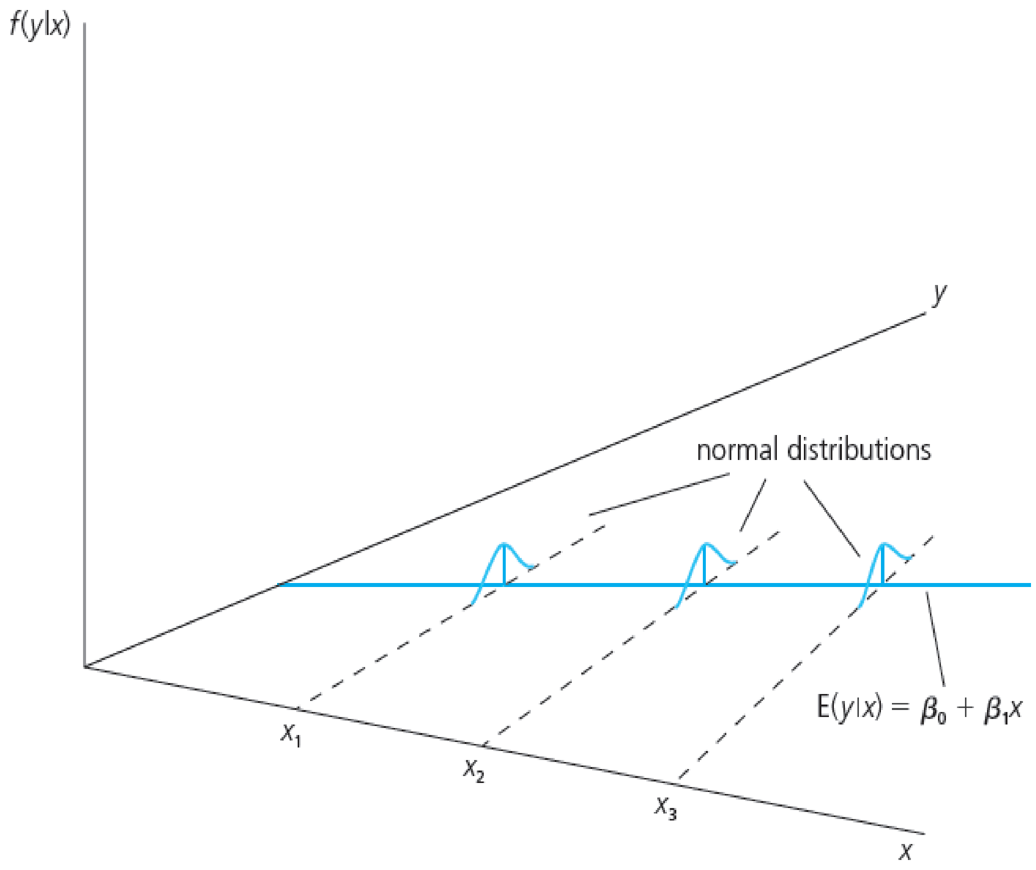
- Statistical inference in regression (hypothesis testing)
- Normality of errors assumption and test for normality
- Hypothesis tests about coefficient significance
  - T-test for coefficient significance
  - F-test for coefficient significance
    - Significance of a single coefficient
    - Joint significance of several coefficients
    - Overall significance of regression – all coefficients jointly significant
- LM chi-square test for coefficient significance

Regression inference

- Inference in regression is hypothesis testing about the population parameters
- OLS estimators have sampling distributions
- We know expected values and variances for OLS estimators
- We need to assume that errors follow the normal distribution

Assumption 6: normality of errors

- The error terms are normally distributed with mean 0 and variance σ^2
- $u_i \sim normal(0, \sigma^2)$
- From previous assumptions, $E(u_i|x_j) = 0$ and $var(u_i|x_j) = \sigma^2$
- The distribution of the error term is independent of the regressors x_j



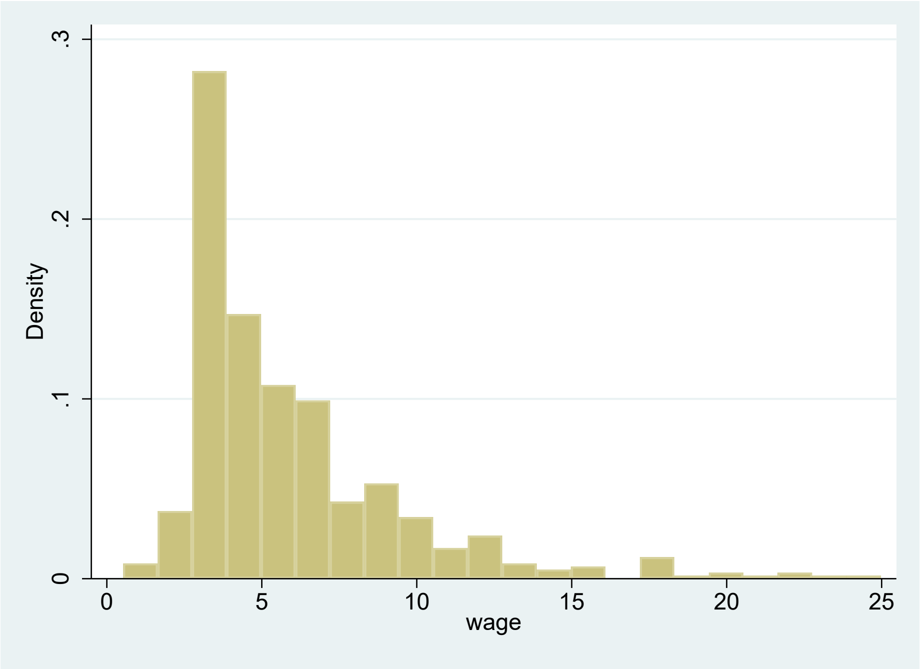
Normality assumption

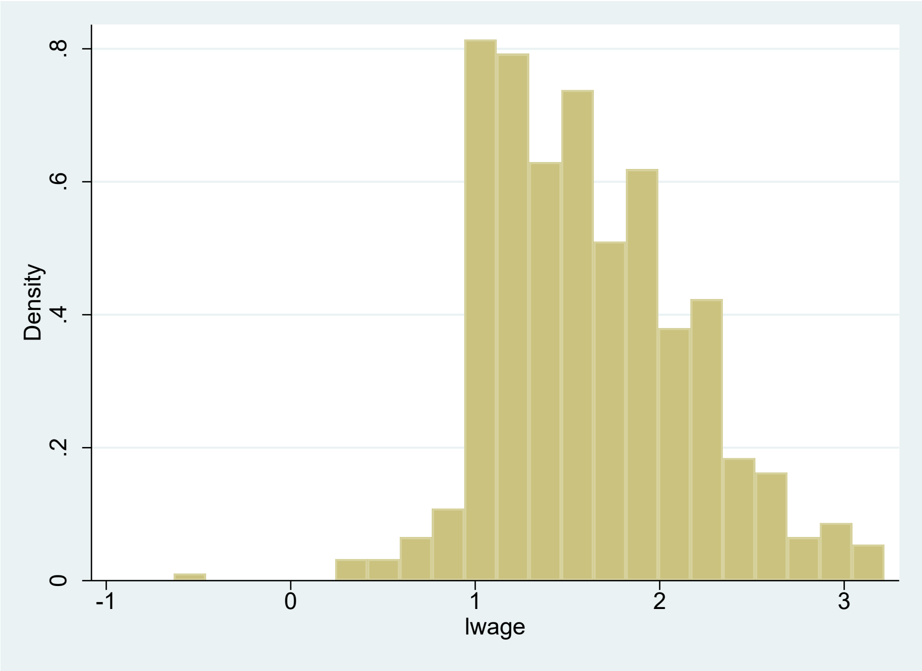
- The normality assumption is reasonable because the error terms are typically a sum of many variables approximating the normal distribution, according to the Central Limit Theorem
- The normality assumption does not hold when the dependent variable is
  - non-negative - wages, prices
  - discrete - number of purchases
  - binary (indicator variable taking on 0 or 1 values) – has insurance or not
- If the distribution of a variable is not normal, taking a log of the variable may make the distribution normal.
- The normality assumption can also be achieved with a large sample size.
- The dependent variable can be tested to see if it is normally distributed with the Shapiro-Wilk test for normality.

Shapiro-Wilk test for normality

- Test for normality: H_0: normality and H_a: non-normality
- The Shapiro-Wilk test statistic $W = \frac{\sum_{i=1}^{n} a_i y_i}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
- Where y_i is the variable, ordered from smallest to largest value and $\bar{y}$ is the sample mean.
- a_i are constants generated from covariances, variances, and means from a normally distributed sample with sample size n.
- If p-value<0.05, reject the null hypothesis of normality.

Histogram and Shapiro-Wilk test example





The histograms show the distributions for wage and log(wage). The distributions not look like the normal distribution. The Shapiro-Wilk test statistic is 0.8 for wage and 0.96 for log(wage) with p-values<0.05, so these variables are not normally distributed.

Coefficients have normal sampling distribution

- Gauss Markov assumptions are assumptions 1-5 (linearity in parameters, random sample, no perfect collinearity, zero conditional mean, and homoscedasticity)
- Classical linear model assumptions are assumptions 1-6 (Gauss Markov assumptions and the normality assumption)
- Under assumptions 1-6, the coefficients have normal sampling distribution.
- $\hat{\beta}_j \sim normal(\beta_j, var(\hat{\beta}_j))$
- The coefficient $\hat{\beta}_j$ has a normal distribution with a mean of the parameter β_j and variance $var(\hat{\beta}_j)$. Remember that $E(\hat{\beta}_j) = \beta_j$.
- Reorganizing: $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim normal(0,1)$
- The coefficient minus the parameter value divided by the standard error of the coefficient has a standard normal distribution with mean 0 and variance 1.

Hypotheses testing – setting up null and alternative hypotheses

- Hypothesis testing can be used to determine whether a statement about the value of the population parameter should or should not be rejected.
- Set up the null and alternative hypotheses

  $H_0: \beta_j = 0$ the population parameter is not significantly different from 0.

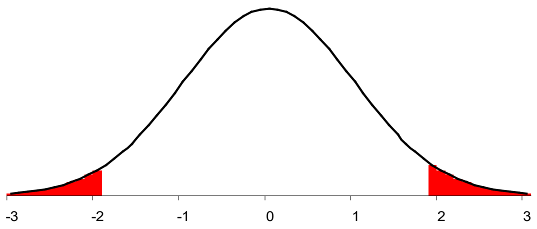
  $H_a: \beta_j \neq 0$ the population parameter is significantly different from 0.
- The null hypothesis H_0 is a tentative assumption about the population parameter.
- The alternative hypothesis H_a or H_1 is the opposite of what is stated in the null hypothesis.
- "not significant" is always in the null hypothesis and "significant" is always in the alternative hypothesis. The equal sign is always in the null hypothesis.

One-tailed vs two-tailed tests

- Two-tailed test

$H_0: \beta_j = 0$

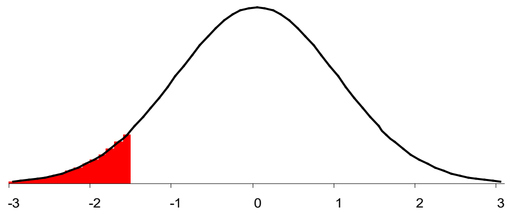
$H_a: \beta_j \neq 0$



- Lower one-tailed test

$H_0: \beta_j \geq 0$

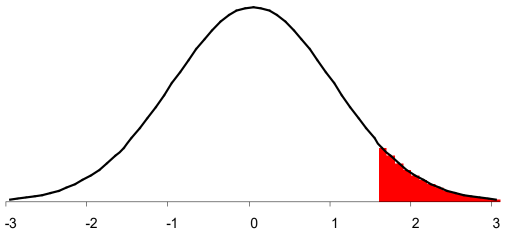
$H_a: \beta_j < 0$



- Upper one-tailed test

$H_0: \beta_j \leq 0$

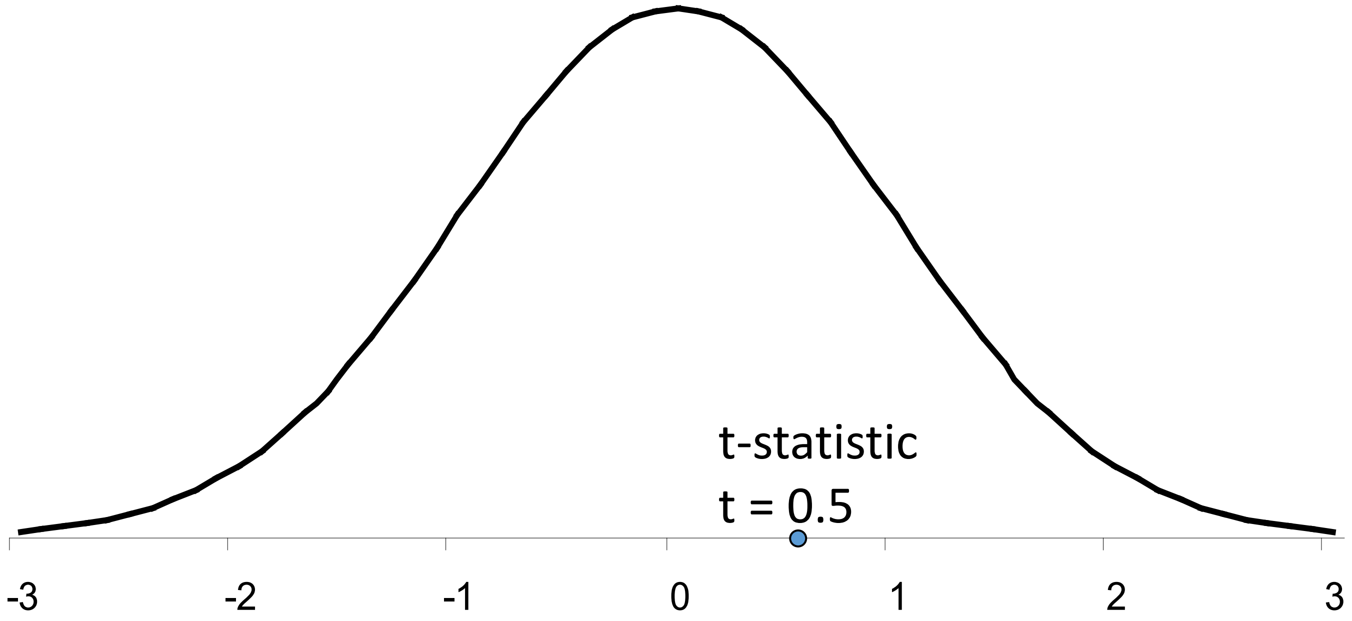
$H_a: \beta_j > 0$



Two-tailed t-tests are commonly used, and one-tailed t-tests are rarely used.

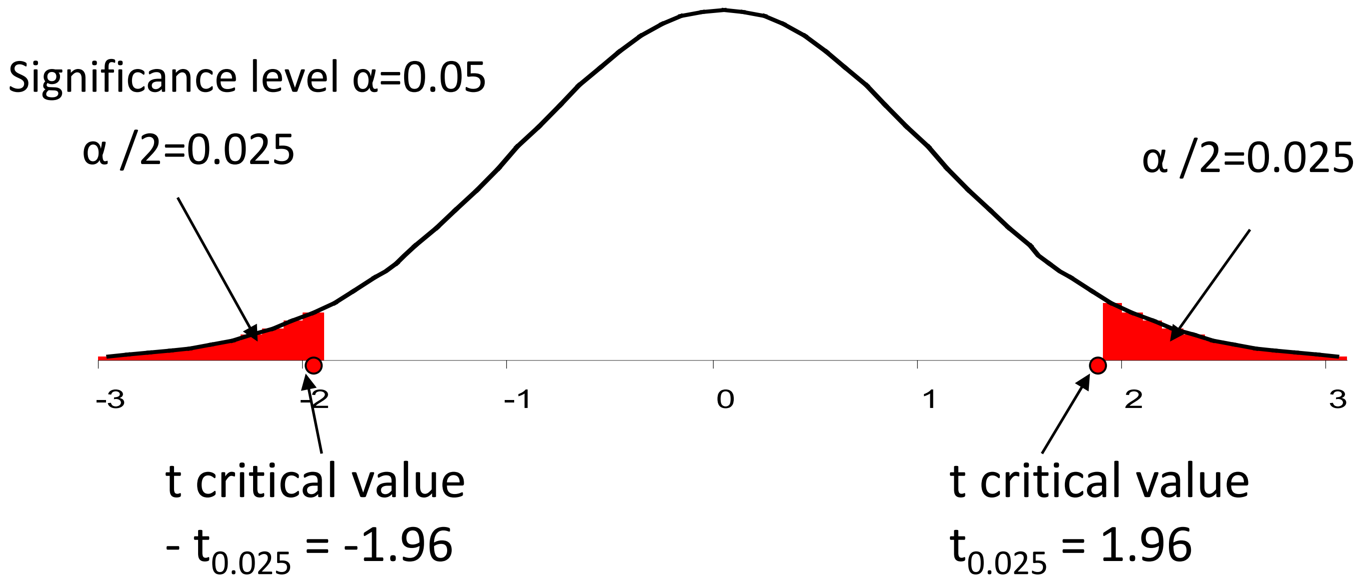
Calculate the t-statistic

- The t-statistic is the coefficient divided by the standard error of the coefficient. It follows the t-distribution with n-k-1 degrees of freedom.
- $t - stat = \frac{\widehat{\beta}_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$
- With a large sample size (n>1000), the *t*-distribution approximates the normal distribution.



Find t critical values

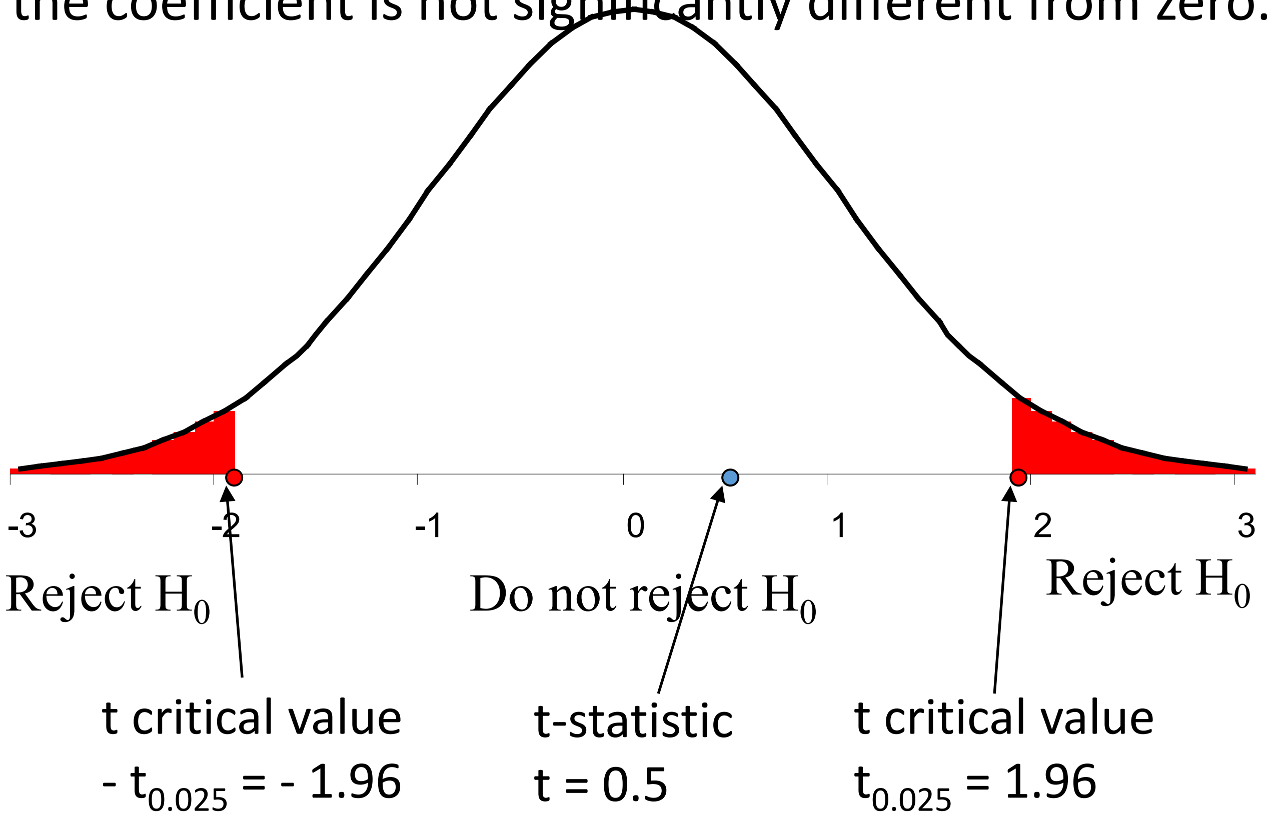
- The t-critical value t_α shows the value at which we are indifferent between rejecting the null hypothesis and not rejecting the null. Critical values are for a specific significance level (usually 0.05 or 5% but can also be 10% or 1%).



The t critical values with sample size greater than 1,000 observations are -1.96 and 1.96.
Red area is the rejection area for the null hypothesis and white area is the area of no rejection.

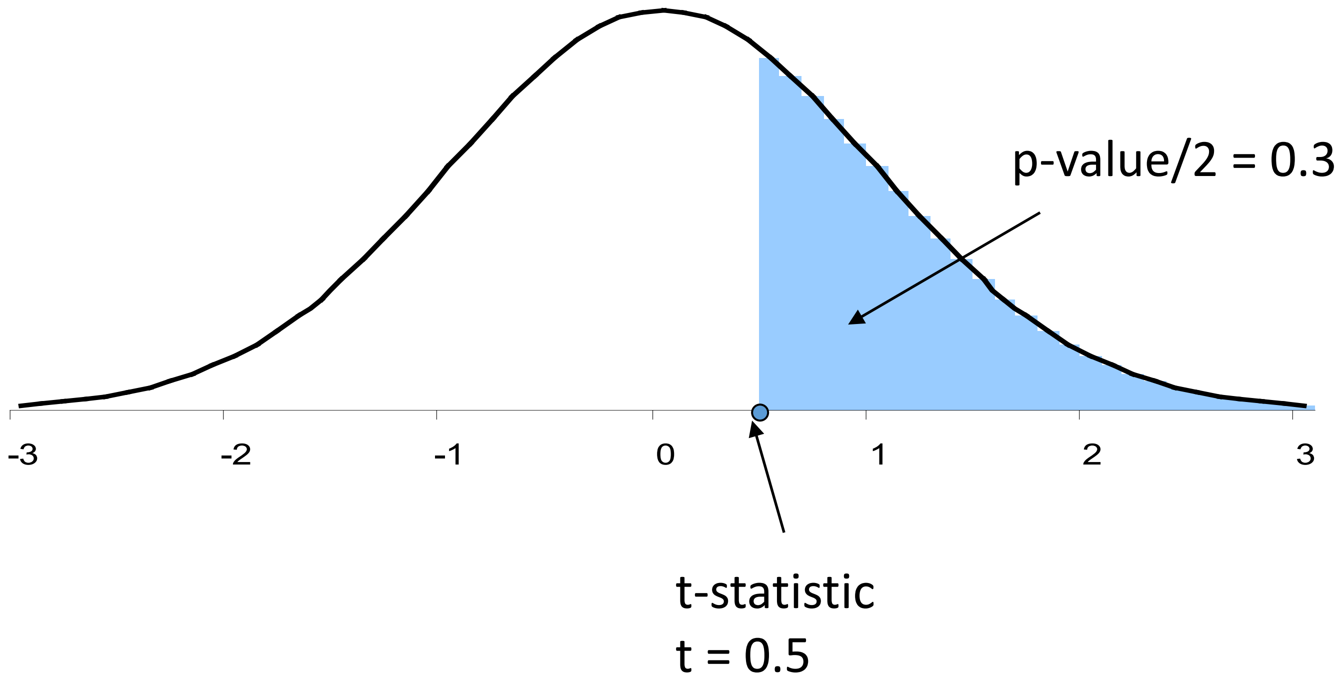
Rejection rule in two-tailed t-tests

- Compare the t-statistic with the t-critical values. Reject the null hypothesis if the t-statistic is in the rejection region, otherwise do not reject the null hypothesis.
- Here, because the t-statistic is not in the rejection region, we cannot reject the null hypothesis that the coefficient is not significantly different from zero.



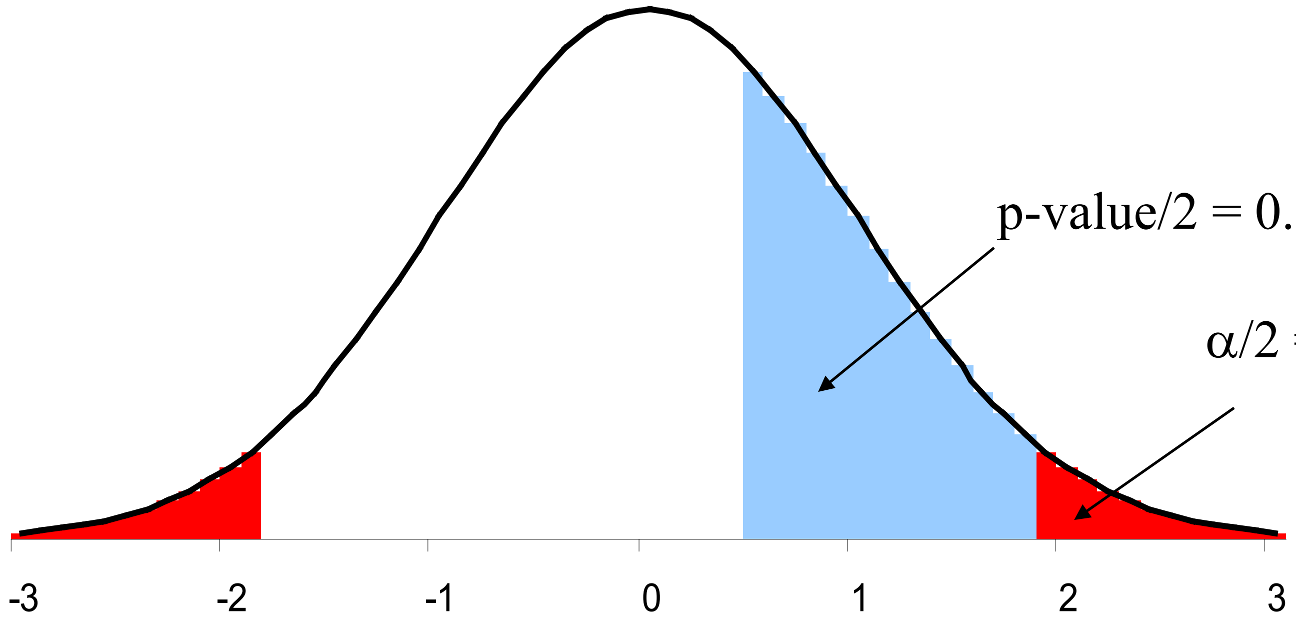
Calculate p-value

- The p-value is the smallest significance level at which the null hypothesis is still rejected.
- Use the t-statistic to calculate the p-value
  - For one-tailed tests: the p-value is the area from the t-statistic to the rejection tail
  - For two-tailed tests, the p-value is twice the area to the closer tail



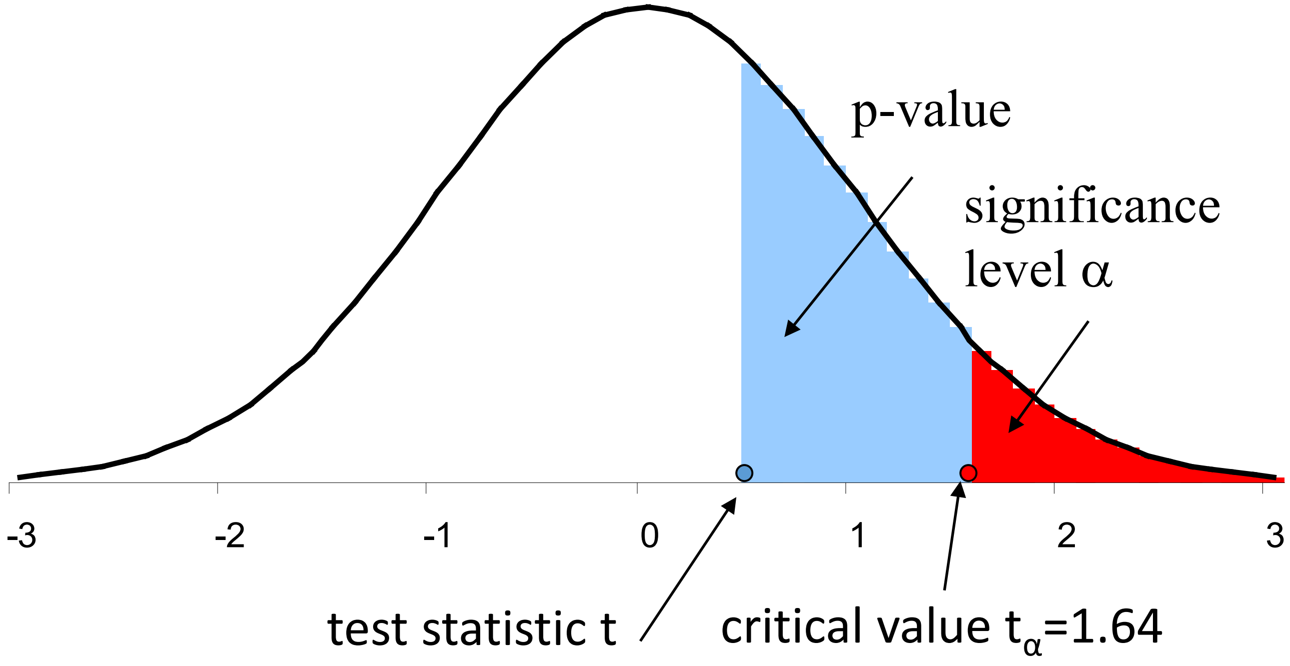
Rejection rule in two-tailed t-test

- Compare the p-value with the significance level α
  - If p-value<α then reject the null hypothesis
  - If p-value>α then do not reject the null hypothesis
- Here, because the p-value > 0.05 = significance level, we cannot reject the null hypothesis, so the coefficient is not significantly different from zero.



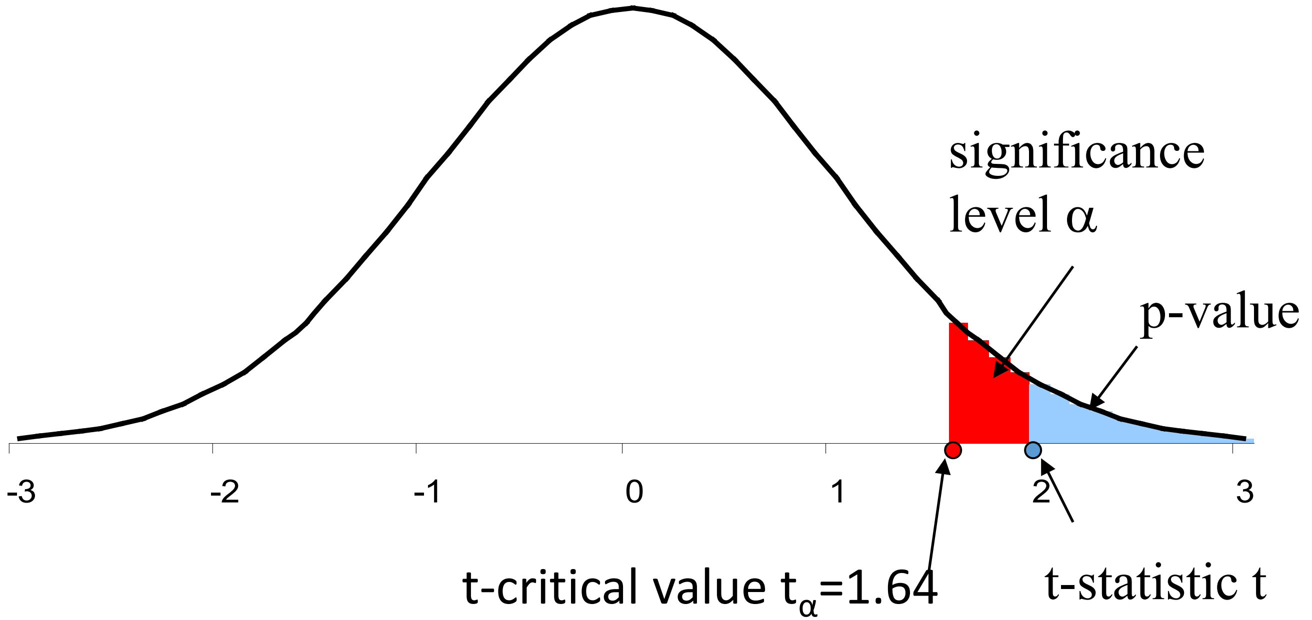
Rejection rule in one-tailed t-test

- An upper one-tailed test when the null hypothesis is not rejected:
  - The t-statistic is not in the rejection area
  - p-value > significance level



Rejection rule in one-tailed t-test

- An upper one-tailed test when the null hypothesis is rejected:
  - The t-statistic is in the rejection area
  - p-value < significance level



One-tailed tests are typically not reported with regression output. You can conduct a one-tailed t-test by comparing the t-statistic with 1.64 for an upper one-tailed test or -1.64 for a lower one-tailed test with a 5% significance level.

Significance level, p-value, critical value, and test statistic

Areas or probabilities	t
significance level α (given to you, usually 0.05)	critical value t_α -1.64 for lower one-tailed tests +1.64 for upper one-tailed tests $\pm$1.96 for two-tailed tests
p-value	test statistic t

Compare significance level α with the p-value

Compare t-critical value t_α with the t-statistic

Confidence intervals

- A 95% confidence interval indicates an interval where the population parameter will be included with 95% confidence.
- The 95% confidence interval corresponds to 5% significance level. Other common values are 90% and 99% confidence intervals.
- Confidence interval = $(\hat{\beta} - t_\alpha * se(\hat{\beta}), \hat{\beta} + t_\alpha * se(\hat{\beta}))$

where t_α is the t-critical value

- With large sample, the t-critical value is 1.96 for the 95% confidence interval and 1.64 for the 90% confidence interval.
- If the confidence interval includes 0, the null hypothesis is not rejected, and the coefficient is not significantly different from zero.
- If the confidence interval does not include 0, the null hypothesis is rejected, and the coefficient is significantly different from zero.

T-tests

- In summary, there are 3 methods to conduct t-tests:
- Method 1: Compare the t-statistic with the t-critical value, reject null hypothesis if t-statistic is in the rejection region.
- Method 2: Compare the p-value with the significance level (0.05), reject null hypothesis if p-value<significance level (0.05).
- Method 3: Compare the confidence intervals to 0, reject null hypothesis if confidence interval does not include 0.
- The same rejection decision is made when using any of the three methods to test for coefficient significance.
- Regression output typically includes asterisks (stars) to indicate coefficient significance.
- Typical significance is 5%. If a coefficient is not significant at 5%, then we can test if it is significant at the 10% significance level.

Coefficient significance interpretation

- The t-test shows “statistical” significance. If a coefficient is statistically significant, its magnitude can be discussed. If a coefficient is not significant, its magnitude is not discussed.
- The question of whether the coefficient is “economically” significant (i.e. large enough in magnitude) is a subjective judgment.
- A criterion to drop a variable from the model is if its coefficient is not significantly different from zero.
- Note that a coefficient is significant. We do not say that a variable is significant. In other words, “the coefficient on education is significant,” not “education is significant.”

T-test example in Stata

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.5556838	.0498647	11.14	0.000	.4577228	.6536447
exper	.0187449	.0120289	1.56	0.120	-.0048863	.0423762
tenure	.1387816	.0211351	6.57	0.000	.0972608	.1803024
female	-1.741397	.2664871	-6.53	0.000	-2.264921	-1.217873
married	.5592403	.2859483	1.96	0.051	-.0025156	1.120996
_cons	-1.618152	.7230508	-2.24	0.026	-3.038611	-.1976921

The t-statistic for exper = 0.019/0.012=1.56, which is less than the t-critical value of 1.96.
The p-value for the coefficient on exper is 0.12 which is greater than the 5% significance level.
The 95% confidence interval for the coefficient on exper is from 0.018-1.96*0.012=-0.005 to 0.018+1.96*0.012=0.04, which includes 0.
All 3 methods lead to the conclusion that the coefficient on experience is not significant.
This means that there is no significant relationship between experience and wage.
We do not interpret the coefficient magnitude.
To construct 90% confidence interval use 1.64 instead of 1.96.
To test if the coefficient on exper is positive (upper one tailed-test),
compare t-stat=1.56 to 1.64 critical value and conclude that the coefficient on exper is not positive.

T-test example in Stata

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.5556838	.0498647	11.14	0.000	.4577228	.6536447
exper	.0187449	.0120289	1.56	0.120	-.0048863	.0423762
tenure	.1387816	.0211351	6.57	0.000	.0972608	.1803024
female	-1.741397	.2664871	-6.53	0.000	-2.264921	-1.217873
married	.5592403	.2859483	1.96	0.051	-.0025156	1.120996
_cons	-1.618152	.7230508	-2.24	0.026	-3.038611	-.1976921

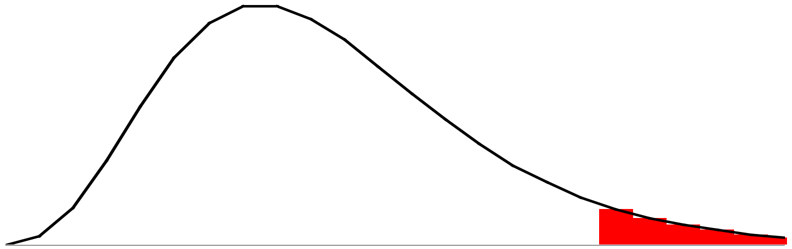
The t-statistic for educ = 0.56/0.05=11.14, which is greater than the t-critical value of 1.96.
The p-value for the coefficient on educ is 0.000 which is less than the 5% significance level.
The 95% confidence interval for the coefficient on educ is from 0.556-1.96*0.050=0.46 to 0.556+1.96*0.050=0.65, which does not include 0.
All 3 methods lead to the conclusion that the coefficient on educ is significant.
This means that there is a significant relationship between education and wage.
Then interpret the coefficient: one additional year of education is associated with $0.55 increase in wages.
To construct 90% confidence interval use 1.64 instead of 1.96.
To test if the coefficient on educ is positive (upper one tailed-test),
compare t-stat=11.14 to 1.64 critical value and conclude that the coefficient on educ is positive.

F-test for significance of a single coefficient

- The F test is used to see if a single coefficient is significant. The F-test uses the F distribution. The conclusion is the same as using a t-test for coefficient significance.
- Unrestricted regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$
- Estimate the unrestricted model and get the residual sum of squares, SSR_{ur}
- Test H_0: β_2=0 H_a: $\beta_2 \neq$0, number of restrictions $q = 1$
- Restricted regression model with β_2 set to zero: $y = \alpha_0 + \alpha_1 x_1 + \alpha_3 x_3 + \alpha_4 x_4 + e$
- Estimate the restricted model and save the residual sum of squares, SSR_r
- Compute $F - stat = \frac{(SSR_r - SSR_{ur})\ /q}{SSR_{ur}/(n-k-1)}$
- SSR_r is residual sum of squares from the restricted model, SSR_{ur} is residual sum of squares from the unrestricted model, q=numerator df = number of restrictions and n-k-1= denominator df
- Compare F-stat to the F critical value with an F(q,n-k-1) distribution
    - If F-stat$>$F-crit or p-value<0.05, the null hypothesis H_0 is rejected (coefficient is significant)
    - If not, we fail to reject the null hypothesis H_0 (coefficient is not significant)

F-test for significance of a single coefficient example

- Unrestricted regression model:
- $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 female + u$
- Test H$_0$: β_2=0   H$_a$: β_2 ≠0, number of restrictions $q = 1$
- Restricted regression model:
- $wage = \alpha_0 + \alpha_1 educ + \alpha_3 tenure + \alpha_4 female + e$
- $F - stat = \frac{\frac{(SSR_r - SSR_{ur})}{q}}{\frac{SSR_{ur}}{n-k-1}} = \frac{(4599-4557)/1}{4557/(526-4-1)}$=4.82
- F critical value (1,521) = 3.86 < F-stat
- p-value=0.029 < 0.05



- Coefficient on experience is significant. We can interpret the magnitude.
- Note that the F-test and t-test give the same p-value of 0.029, so the conclusion is the same.

F-test for joint significance of several coefficients

- The F test is used to see if several coefficients are jointly significant.
- Unrestricted regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$
- Estimate the unrestricted model and get the residual sum of squares, SSR_{ur}
- H_0: β_2=0 and β_3=0 H_a: $\beta_2 \neq$0 or $\beta_3 \neq$0, number of restrictions $q = 2$
- Restricted regression model with β_2 and β_3 set to zero: $y = \alpha_0 + \alpha_1 x_1 + \alpha_4 x_4 + e$
- Estimate the unrestricted model and save the residual sum of squares, SSR_r
- Compute $F - stat = \frac{(SSR_r - SSR_{ur})\ /q}{SSR_{ur}/(n-k-1)}$
- SSR_r is residual sum of squares from the restricted model, SSR_{ur} is residual sum of squares from the unrestricted model, q=numerator df = number of restrictions and n-k-1= denominator df
- Compare F-stat to the F critical value with an F(q,n-k-1) distribution
  - If F-stat$>$F-crit or p-value<0.05, the null hypothesis H_0 is rejected (coefficients are jointly significant)
  - If not, we fail to reject the null hypothesis H_0 (coefficients are not jointly significant)

F-test for joint significance of several coefficients example

- Unrestricted regression model:
- $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 female + u$
- H_0: β_2=0 and β_3=0 H_a: $\beta_2 \neq$0 or $\beta_3 \neq$0, number of restrictions $q = 2$
- Restricted regression model:
- $wage = \alpha_0 + \alpha_1 educ + \alpha_4 female + e$
- $F - stat = \frac{\frac{(SSR_r - SSR_{ur})}{q}}{\frac{SSR_{ur}}{n-k-1}} = \frac{(5307-4557)/2}{4557/(526-4-1)}$=42.86
- F critical value (2,521) = 3.01 < F-stat
- p-value < 0.05
- Coefficients on experience and tenure are jointly significant. Their magnitudes can be interpreted.

F-test for overall significance

- The F test for overall significance is used to test if all coefficients are jointly significant.
- Unrestricted regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$
- Estimate the unrestricted model and get the residual sum of squares, SSR_{ur}
- H_0: β_1=0 and β_2=0 and β_3=0 and β_4=0 H_a:$\beta_1 \neq 0$ or $\beta_2 \neq 0$ or $\beta_3 \neq$0 or $\beta_4 \neq$0, $q = k = 4$
- Restricted regression model with β_1, β_2, β_3, and β_4 set to zero: $y = \alpha_0 + e \quad (\alpha_0 = \bar{y})$
- Estimate the unrestricted model and save the residual sum of squares, SSR_r
- Compute $F - stat = \frac{(SSR_r - SSR_{ur})\,/q}{SSR_{ur}/(n-k-1)}$
- Note that $SSR_r = SST_{ur}$, so $F - stat = \frac{\frac{(SST_{ur} - SSR_{ur})}{q}}{\frac{SSR_{ur}}{n-k-1}} = \frac{\frac{(SSE_{ur})}{k}}{\frac{SSR_{ur}}{n-k-1}}$
- Compare F-stat to the F critical value with an F(q,n-k-1) distribution
  - If F-stat$>$F-crit or p-value<0.05, the null hypothesis H_0 is rejected (all coefficients are jointly significant)
  - If not, we fail to reject the null hypothesis H_0 (all coefficients are not jointly significant)

F-test for overall significance

- Unrestricted regression model:
- $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 female + u$
- H_0: β_1=0 and β_2=0 and β_3=0 and β_4=0 H_a:$\beta_1 \neq 0$ or $\beta_2 \neq 0$ or $\beta_3 \neq$0 or $\beta_4 \neq$0, $q = k = 4$
- Restricted regression model:
- $wage = \alpha_0 + e$ (α_0= average value of wage)
- $F - stat = \frac{\frac{(SSR_r - SSR_{ur})}{q}}{\frac{SSR_{ur}}{n-k-1}} = \frac{(7160-4557)/4}{4557/(526-4-1)}$=74.60 (see in regression output)
- F critical value (4,521) = 2.39 < F-stat
- p-value < 0.05
- All coefficients for the regression are jointly significant. All regressors jointly explain wage.
- Almost always the F-test will show overall significance, otherwise the regression model is poorly specified.

ANOVA: Analysis of Variance

Source of variation	SS (Sum of Squares)	df (Degrees of freedom)	MS (Mean squares)	F
Model (Explained)	SSE	k	MSE=SSE/k	F-stat=MSE/MSR
Residual	SSR	n-k-1	MSR=SSR/(n-k-1)	
Total	SST	n-1	MST=SST/(n-1)	

ANOVA Stata example

Source	SS	df	MS
Model	2603.1066	4	650.77665
Residual	4557.30771	521	8.74723169
Total	7160.41431	525	13.6388844

Number of obs	=	526
F(4, 521)	=	74.40
Prob > F	=	0.0000
R-squared	=	0.3635
Adj R-squared	=	0.3587
Root MSE	=	2.9576

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.5715048	.0493373	11.58	0.000	.4745803	.6684293
exper	.0253959	.0115694	2.20	0.029	.0026674	.0481243
tenure	.1410051	.0211617	6.66	0.000	.0994323	.1825778
female	-1.810852	.2648252	-6.84	0.000	-2.331109	-1.290596
_cons	-1.567939	.7245511	-2.16	0.031	-2.99134	-.144538

F-tests for coefficient significance

	Unrestricted model	Model with 1 restriction	Model with 2 restrictions	Model with constant (overall significance)
VARIABLES	wage	wage	wage	wage
educ	0.572***	0.538***	0.506***	
	(0.0493)	(0.0471)	(0.0504)	
exper	0.0254**			
	(0.0116)			
tenure	0.141***	0.164***		
	(0.0212)	(0.0183)		
female	-1.811***	-1.788***	-2.273***	
	(0.265)	(0.266)	(0.279)	
Constant	-1.568**	-0.845	0.623	5.896***
	(0.725)	(0.648)	(0.673)	(0.161)
SSR	4557	4599	5307	7160
F-stat for restrictions		4.82	42.86	74.40
F-critical value		3.86	3.01	2.39
p-value for restrictions		0.029	0	0

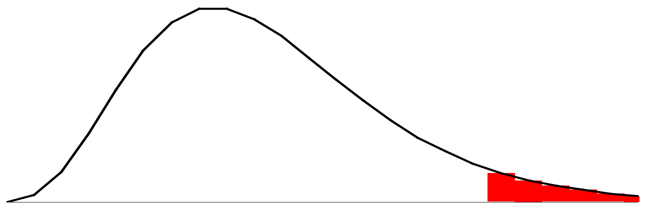
With more restrictions, SSR_r increases (variation that cannot be explained), and F-stat increases, so more evidence of joint significance of coefficients. F-test with 1 restriction has same p-value of t-test.

Lagrange Multiplier test

- Lagrange Multiplier (LM) test is used to see if coefficients are jointly significant. The LM test uses the χ^2 distribution. It is less commonly used than the F-test.
- Regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u$
- Test H_0: β_2=0 and β_3=0 H_a: $\beta_2 \neq$0 or $\beta_3 \neq$0
- Regress y on the restricted set of independent variables and save the residuals, $\hat{e}$

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_4 x_4 + e$$

- Regress $\hat{e}$ on all independent variables and obtain the R-squared, R_e^2
- $\hat{e} = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + v$
- Compute $LM - stat = nR_e^2$
- Compare LM-stat to the chi-square critical value with a χ_q^2 distribution
  - If $LM - stat >$ chi-square critical value or p-value<0.05, the null hypothesis H_0 is rejected (coefficients are jointly significant)
  - If not, we fail to reject the null hypothesis H_0 (coefficients are not jointly significant)



Lagrange Multiplier test example

- Regression model: $wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 female + u$
- Test H_0: β_2=0 and β_3=0 H_a: $\beta_2 \neq$0 or $\beta_3 \neq$0 with q=2 for two restrictions
- Regress $wage$ on the restricted set of independent variables and save the residuals, $\hat{e}$

$$wage = \alpha_0 + \alpha_1 educ + \alpha_4 female + e$$

- Regress $\hat{e}$ on all independent variables and obtain the R-squared, R_e^2
- $\hat{e} = \gamma_0 + \gamma_1 educ + \gamma_2 tenure + \gamma_3 exper + \gamma_4 female + v$
- Compute $LM - stat = nR_e^2$

LM test example

	Restricted Model	Residuals on all regressors
VARIABLES	wage	ehat
educ	0.506***	0.0651
	(0.0504)	(0.0493)
exper		0.0254**
		(0.0116)
tenure		0.141***
		(0.0212)
female	-2.273***	0.463*
	(0.279)	(0.265)
Constant	0.623	-2.191***
	(0.673)	(0.725)
Observations	526	526
R-squared	0.259	0.141

$LM - stat = nR_e^2 = 526 * 0.141 = 74.32$, chi-square critical value with 2 degrees of freedom is 5.99.
p-value<0.05. The null hypothesis is rejected so the coefficients on exper and tenure are jointly significant.

Review questions

- Describe assumption 6. What does assumption 6 imply about the estimators?
- Describe the test for normality.
- Describe the steps of hypothesis testing.
- How many methods can we use with a t-test to reach conclusion about coefficient significance?
- Describe the three F-tests for coefficient significance.
- Describe the LM test for coefficient significance.