- From the previous lectures, we have an idea of how OLS works (and the many assumptions required for it to work).
- How can all this help me in my research you might ask.
- Let's see…

INTERPRETATION

PROF. SEBASTIÁN VALLEJO VERA

- Let’s expand on some of what we have learned so far about the expected values and variance of OLS estimators.
- If we want to perform statistical inference, we need to know the full sampling distribution of our $\widehat{\beta_k}$ of interest.
- To do this, we require one final assumption…

ASSUMPTION MLR.6: NORMALITY

From MLR.4

- The population error μ is independent of the explanatory variables $x_1, x_2, x_3, \ldots, x_k$ and is normally distributed with zero mean and variance σ^2:

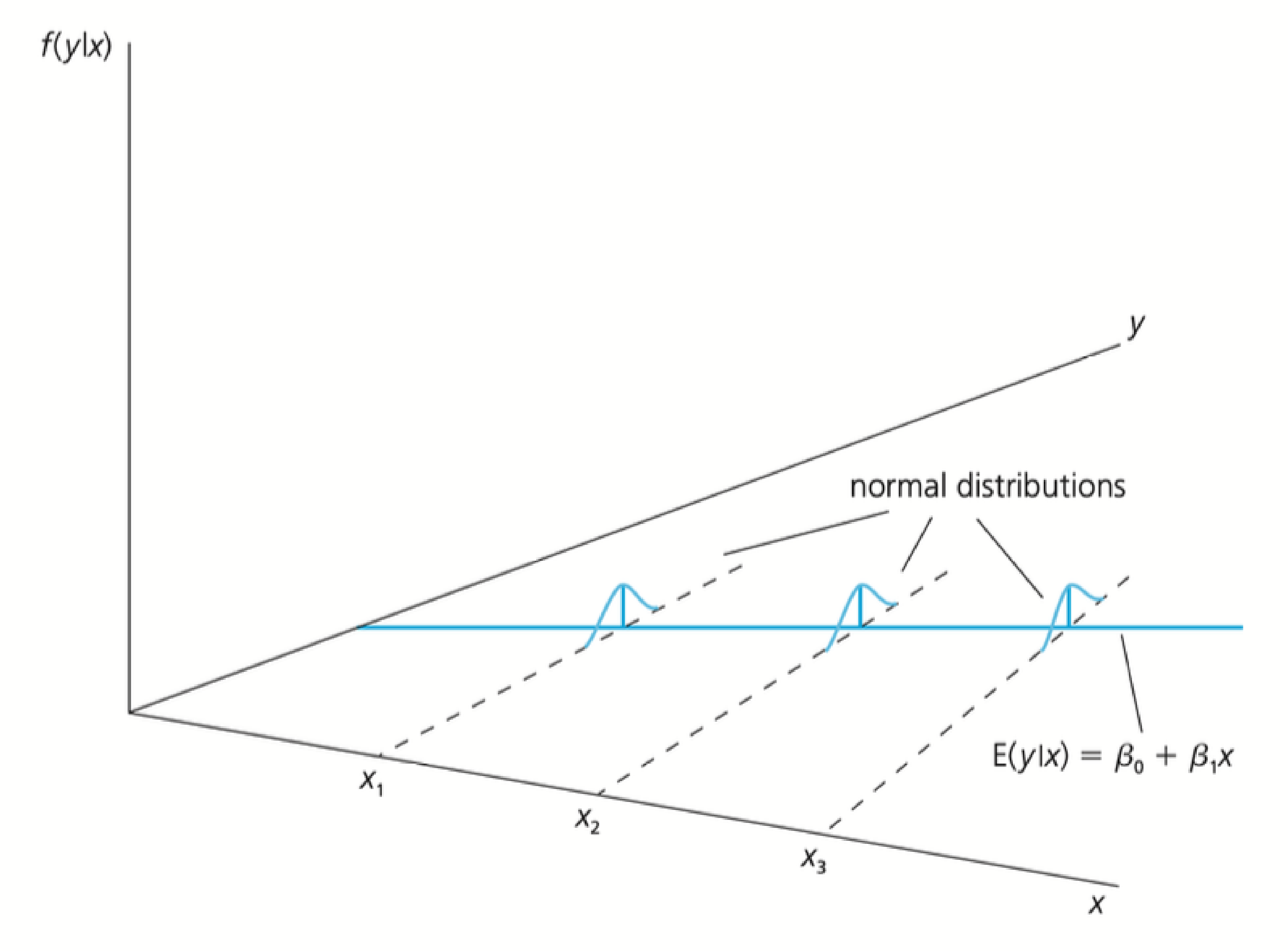
$$\mu \sim Normal(0, \sigma^2)$$

From MLR.5

- For *cross-sectional* regression applications, Assumptions MLR.1 through MLR.6 are called the **classical linear model (CLM) assumptions**. Thus, we will refer to the model under these six assumptions as the **classical linear model**.

$$CLM = Gauss - Markov + normality$$

FIGURE 4.1 The homoskedastic normal distribution with a single explanatory variable.



- The argument justifying the normal distribution for the errors usually runs something like this: because μ is the sum of many different unobserved factors affecting y, we can invoke the central limit theorem to conclude that μ has an approximate normal distribution.
- In any application, however, whether normality of μ can be assumed is really an empirical matter and there are some examples where MLR.6 is clearly false (e.g., whenever y takes on just a few values it cannot have anything close to a normal distribution).

- Normality of the error term translates into normal distributions of the OLS estimators:

THEOREM 4.1: NORMAL SAMPLING DISTRIBUTION

Under the CLM assumptions MLR.1 through MLR.6, conditional on the sample values of the independent variables,

$$\widehat{\beta_j} \sim Normal[\beta_j, \mathrm{Var}(\widehat{\beta_j})]$$

where

We are standardizing our estimates $\widehat{\beta_j}$

$$Var(\widehat{\beta_j}) = \frac{\sigma^2}{TSS_j\left(1 - R_j^2\right)}$$

Therefore:

$$\frac{\widehat{\beta_j} - \beta_j}{sd\left(\widehat{\beta_j}\right)} \sim Normal(0, \sigma^2)$$

Remember that this depends on σ^2, but this is unknown.

Suppose the following population model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \mu$$

and let's assume that it satisfies the CLM assumptions.

Remember that the β_j are unknown population parameters and we will never know them with certainty. Nevertheless, we can hypothesize about the value of β_j and then use statistical inference to test our hypothesis.

To construct hypothesis tests, we need the following results:

THEOREM 4.2: T DISTRIBUTION FOR THE STANDARDIZED ESTIMATORS

Under the CLM assumptions MLR.1 through MLR.6,

$$\frac{\widehat{\beta_j} - \beta_j}{se(\widehat{\beta_j})} \sim t_{n-(k+1)}$$

where $k + 1$ is the number of unknown parameters in the population model (k slope parameters and the intercept β_0) and $n - k - 1$ is the degrees of freedom (df).

Note that Theorem 4.2 is different than Theorem 4.1 because the constant σ in $sd(\widehat{\beta_j})$ has to be replaced with the random variable $\hat{\sigma}$. It now follows a t distribution with $n - k - 1$ degrees of freedom.

Theorem 4.2 is important in that it allows us to test hypotheses involving β_j. In most applications, our primary interest lies in testing the null hypothesis

$$H_0: \beta_j = 0$$

where j corresponds to any of the k independent variables.

Since β_j measures the partial effect of x_j on (the expected value of) y, after controlling for all other independent variables, H_0 means that, once $x_1, x_2, x_3, \ldots, x_{j-1}, x_{j+1}, x_k$ have been accounted for, x_j has no effect on the expected value of y.

- Let’s say that I want to know about the determinants of wages. I am particularly interested in the effect of education on wages. I estimate the following equation:

$$wage = \beta_0 + \beta_1 education + \beta_2 experience + \beta_3 tenure + u$$

- The null hypothesis $H_0\colon \beta_2 = 0$ means that, once education and tenure have been accounted for, the number of years in the workforce ($experience$) has no effect on hourly wage.

The statistic we use to test H_0 (against any alternative) is called "the" t statistic or "the" t ratio of $\widehat{\beta_j}$ and is defined as:

$$t_{\widehat{\beta}_j} = \hat{\beta}_j / se(\hat{\beta}_j)$$

Note: "the" appears in quotation marks because a more general form of the t statistic is needed for testing other hypotheses about βj (next time).

Note that:

- since $se(\hat{\beta}_j)$ is always positive, $t_{\widehat{\beta}_j}$ has the same sign as $\hat{\beta}_j$
- for a given value of $se(\hat{\beta}_j)$, a larger value of $\hat{\beta}_j$ leads to a larger value of $t_{\widehat{\beta}_j}$

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Since we are testing $H_0\colon \beta_2 = 0$, it is only natural to look at our unbiased estimator of β_j, $\hat{\beta}_j$.

The point estimate $\hat{\beta}_j$, however, will never be exactly zero, whether or not H_0 is true. The question is: **How far is $\boldsymbol{\hat{\beta}_j}$ from zero?**

A sample value of $\hat{\beta}_j$ very far from zero provides evidence against

$$H_0\colon \beta_j = 0.$$

We must recognize, however, that there is a sampling error in our estimate $\hat{\beta}_j$, so the size of $\hat{\beta}_j$ must be weighed against its sampling error.

Since the standard error of $\hat{\beta}_j$ is an estimate of the standard deviation of $\hat{\beta}_j$, $t_{\hat{\beta}_j}$ measures how many estimated standard deviations $\hat{\beta}_j$ is away from zero.

Values of $t_{\hat{\beta}_j}$ sufficiently far from zero will result in a rejection of H_0, but the precise rejection rule depends on the alternative hypothesis and the chosen significance level of the test.

Note also that determining a rule for rejecting $H_0: \beta_j = 0$ at a given significance level—that is, the probability of rejecting H_0 when it is true—requires knowing the sampling distribution of $t_{\hat{\beta}_j}$ when H_0 is true.

From Theorem 4.2, we know this to be $t_{n-(k+1)}$. This is the key theoretical result needed for testing $H_0: \beta_j = 0$.

To determine a rule for rejecting H_0, we first need to decide on the relevant alternative hypothesis. **First**, consider a one-sided alternative of the form

$$H_1: \beta_j > 0$$

Second, we must decide on a **significance level** or the probability of rejecting H_0 when it is in fact true.

A conventional choice is the 5% significance level. Thus, we are willing to mistakenly reject H_0 when it is true 5% of the time.

Under the alternative $\beta_j > 0$, the expected value of $t_{\hat{\beta}_j}$ is positive. Thus, we are looking for a "sufficiently large" positive value of $t_{\hat{\beta}_j}$ to reject $H_0: \beta_j = 0$ in favor of $H_1: \beta_j > 0$. Negative values of $t_{\hat{\beta}_j}$ provide no evidence in favor of H_1.

The definition of "sufficiently large," with a 5% significance level, is the 95th percentile in a t distribution with $n-k-1$ degrees of freedom; denote this by c (critical value). In other words, the rejection rule is that H_0 is rejected in favor of H_1 at the 5% significance level if

$$t_{\hat{\beta}_j} > c$$

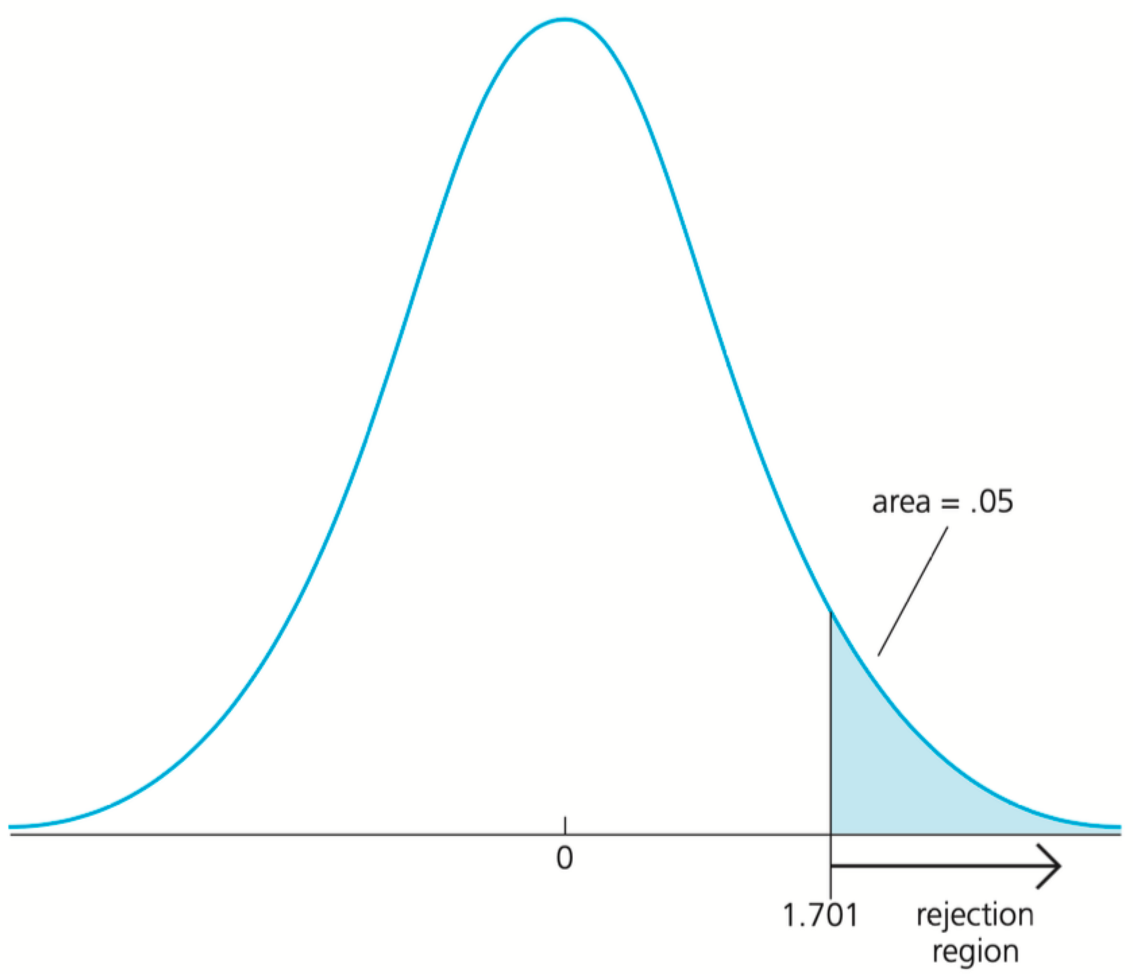
This rejection rule is an example of a one-tailed test.

To obtain c, we only need the significance level and the degrees of freedom to look up its value in a t-table.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

FIGURE 4.2 5% rejection rule for the alternative H_1: $\beta_j > 0$ with 28 *df*.



The other one-sided alternative stipulates that the parameter is less than zero,

$$H_1\colon \beta_j < 0$$

The rejection rule for alternative $H_1\colon \beta_j < 0$ is just the mirror image of $H_1\colon \beta_j > 0$ and is now:

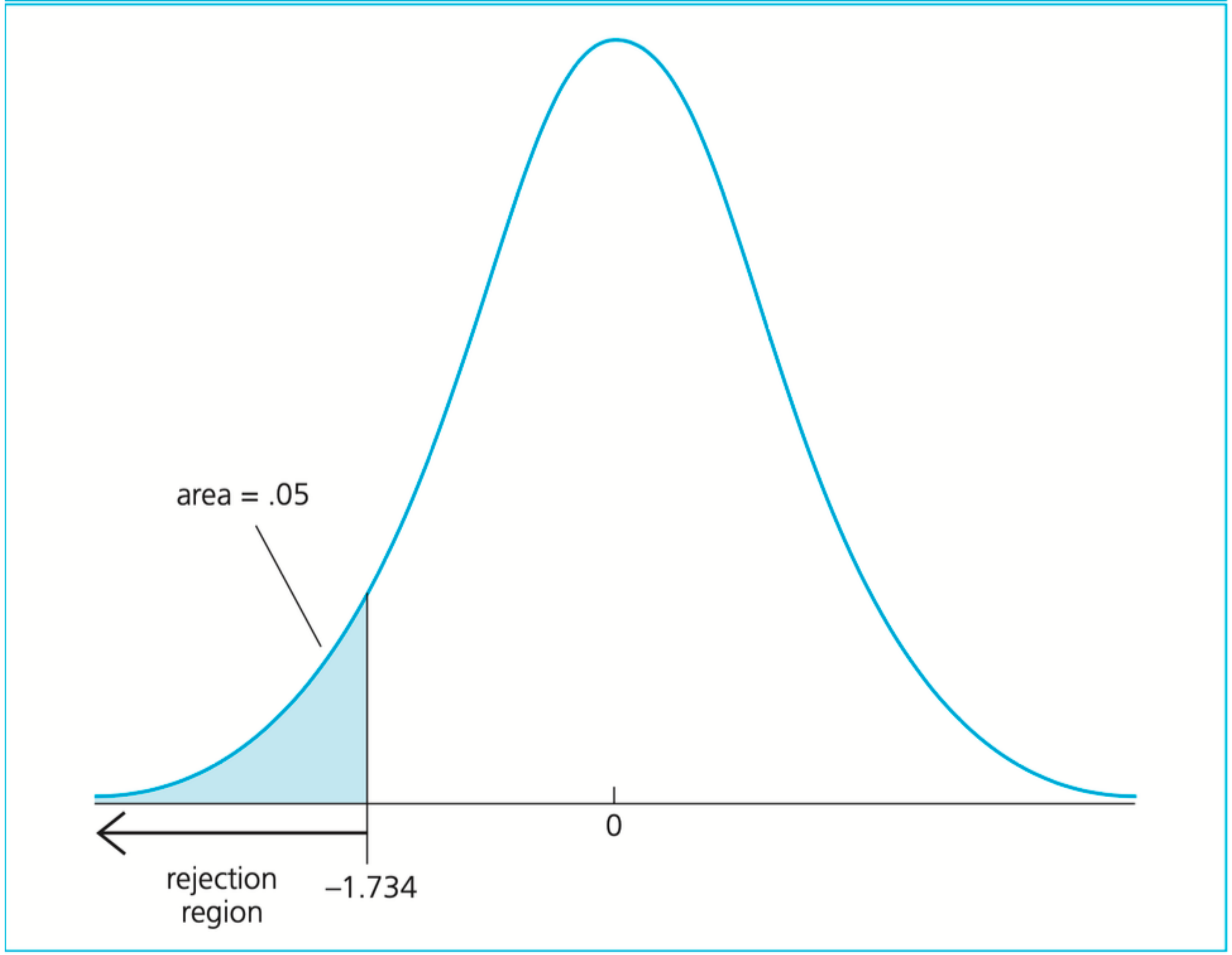
$$t_{\hat{\beta}_j} < c$$

Note that the critical value comes this time from the left tail of the t distribution.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

FIGURE 4.3 5% rejection rule for the alternative H_1: $\beta_j < 0$ with 18 *df*.



It is common to test the null hypothesis $H_0\colon \beta_j = 0$ against a two-sided alternative; that is,

$$H_1\colon \beta_j \neq 0$$

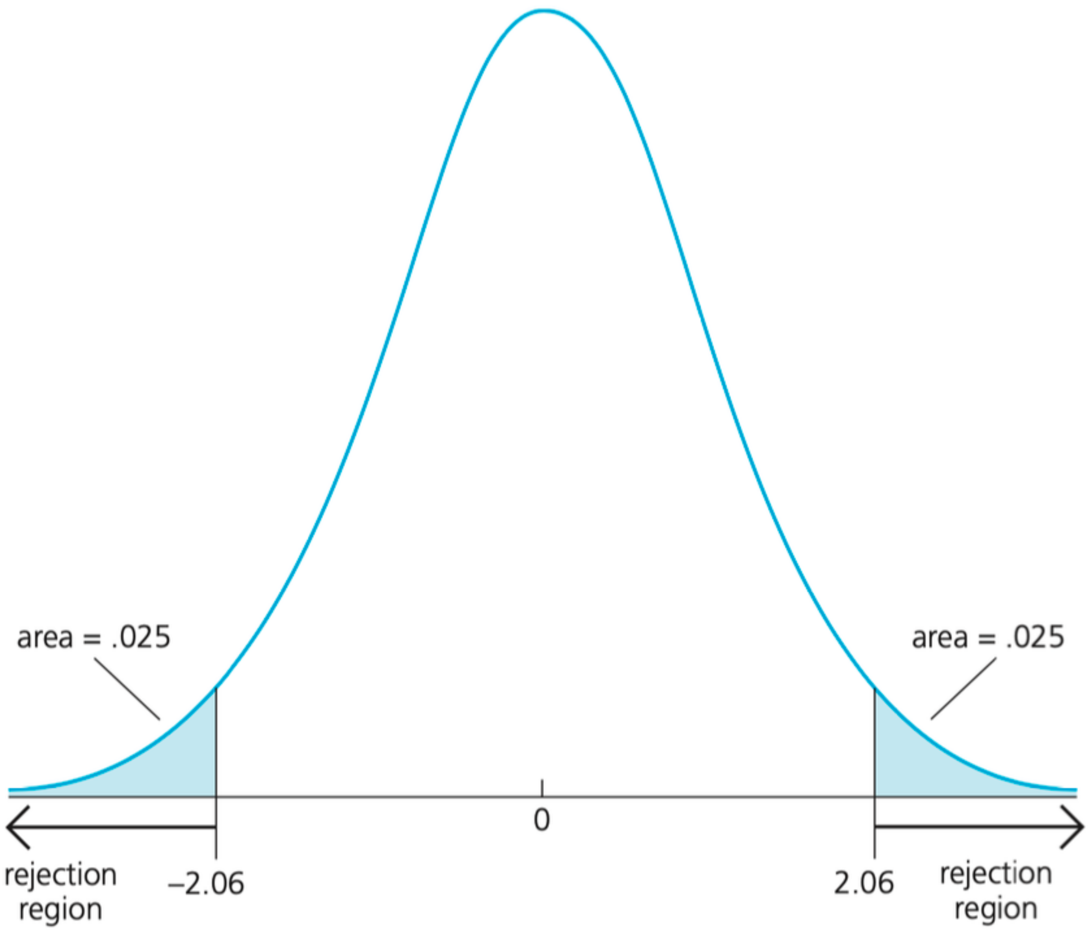
When the alternative is two-sided, we are interested in the absolute value of the t statistic. The rejection rule for $H_0\colon \beta_j = 0$ against $H_1\colon \beta_j \neq 0$ is:

$$\left| t_{\hat{\beta}_j} \right| > c$$

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

FIGURE 4.4 5% rejection rule for the alternative H_1: $\beta_j \neq 0$ with 25 *df*.



Two important things to note about the critical value, c:

- as the significance level falls, the critical value increases, so that we require a larger value of $t_{\hat{\beta}_j}$ to reject H_0
- as the degrees of freedom, df, increases, the critical value decreases, so that we require a smaller value of $t_{\hat{\beta}_j}$ to reject H_0

Note also that as the degrees of freedom in the t distribution get large (120+), the t distribution approaches the standard normal distribution (z distribution).

Although $H_0: \beta_j = 0$ is the most common hypothesis, we sometimes want to test whether β_j is equal to some other given constant. Generally, if the null is stated as

$$H_0: \beta_j = a_j$$

where a_j is our hypothesized value of β_j , then the appropriate t statistic is:

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)}$$

As before, t measures how many estimated standard deviations $\hat{\beta}_j$ is away from the hypothesized value of β_j.

Under $H_0: \beta_j = a_j$, this t statistic is distributed as $t_{n-(k+1)}$ from Theorem 4.2 and it can be used against one-sided or two-sided alternatives.

COMPUTING P-VALUES FOR T TESTS

- The classical approach to hypothesis testing consists of identifying an alternative hypothesis and then choosing a significance level to then determine a critical value. Once the critical value has been identified, the value of the t statistic is compared with the critical value, and the null is either rejected or not rejected at the given significance level.
- Even after deciding on the appropriate alternative, there is a component of arbitrariness to the classical approach, which results from having to choose a significance level ahead of time. Different researchers prefer different significance levels, depending on the particular application. There is no "correct" significance level.

- Rather than testing at different significance levels, it is more informative to answer the following question: Given the observed value of the t statistic, what is the smallest significance level at which the null hypothesis would be rejected? This level is known as the **p-value** for the test.
- Thus the p-value is the probability of observing a t statistic as extreme as we did if the null hypothesis is true. This means that small p-values are evidence against the null; large p-values provide little evidence against H_0.
- Once the p-value has been computed, a classical test can be carried out at any desired level. If α denotes the significance level of the test (in decimal form), then H_0 is rejected if $p\ value\ < \alpha$; otherwise, H_0 is not rejected at the $100\ \times\ \alpha\ \%$ level.

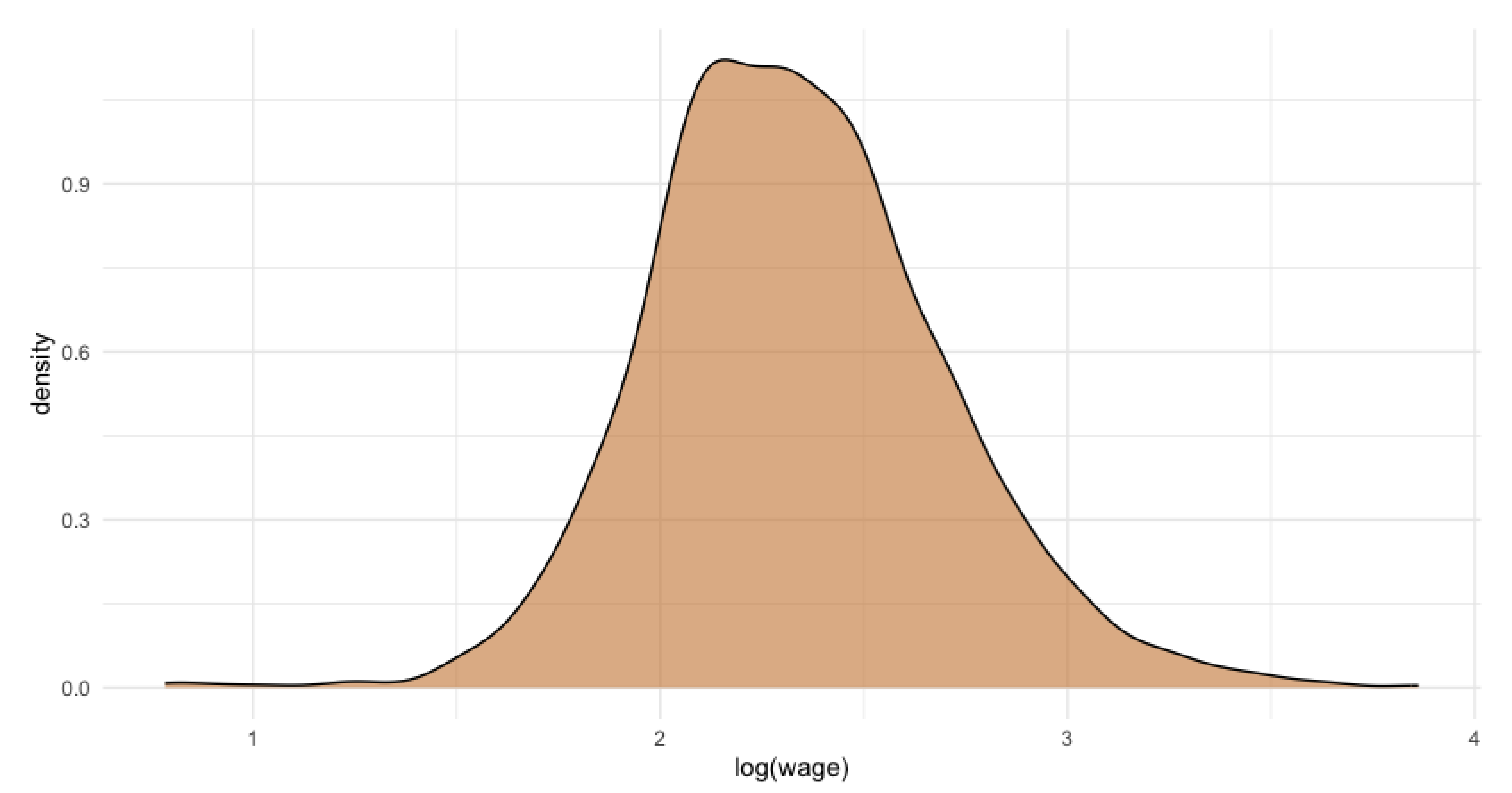
EXAMPLE

- Let's say that I want to know that I have some theory about the determinants of wages. I am particularly interested in the effect of experience and education on wages. I estimate the following equation:

$$wage = \beta_0 + \beta_1 experience + \beta_2 education + u$$

- Where wage is gross hourly wage rate in euro, education is education level from 1 [low] to 5 [high], and experience is years of experience. The data covers a sample of wage-earners from Belgium in 1994.

- Before running any model, I want to look at the distribution of my variables.



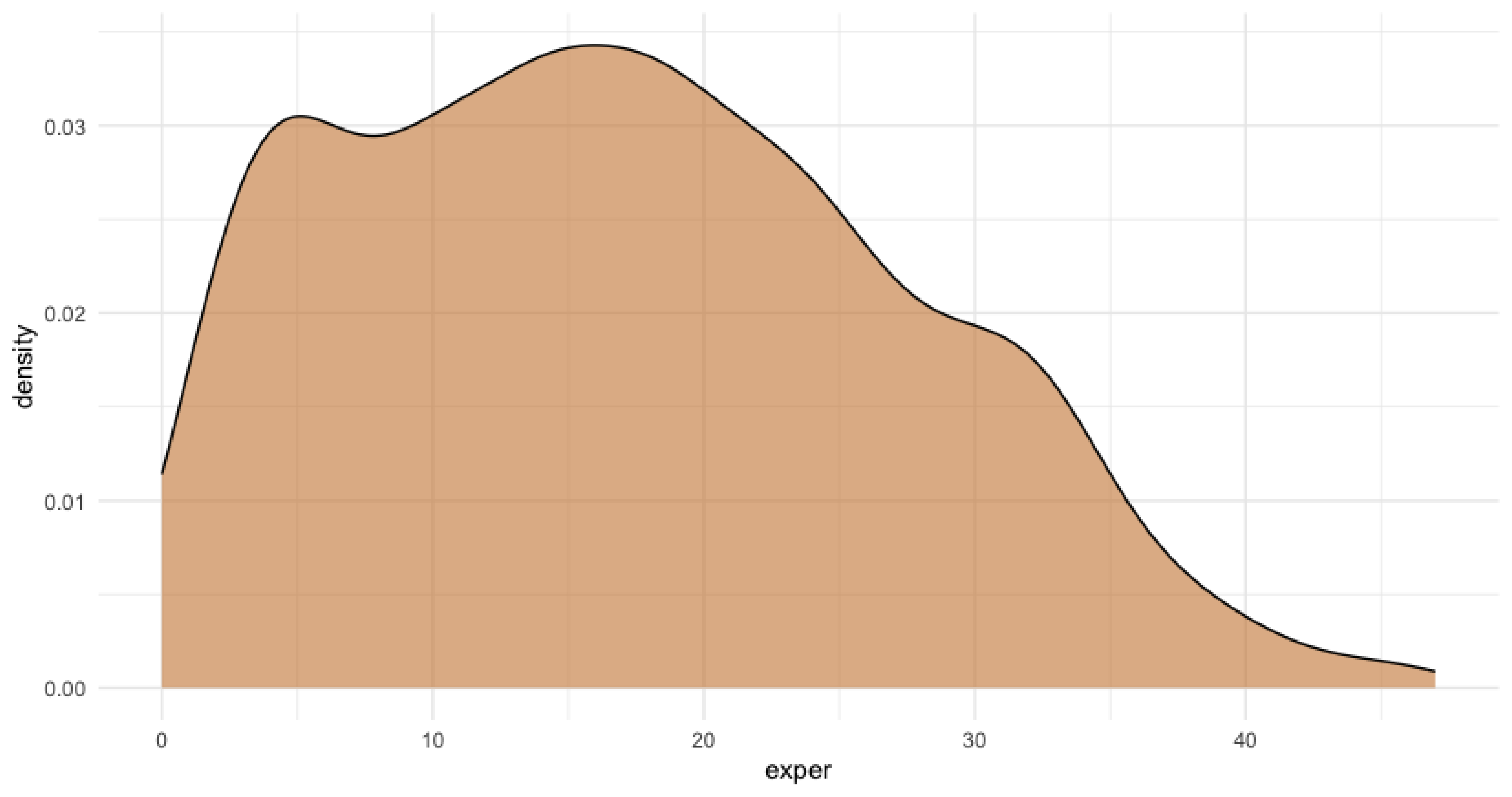
density
0.9
0.6
0.3
0.0
1
2
3
4
log(wage)

- Let’s say that I want to know that I have some theory about the determinants of wages. I am particularly interested in the effect of experience and education on wages. I estimate the following equation:

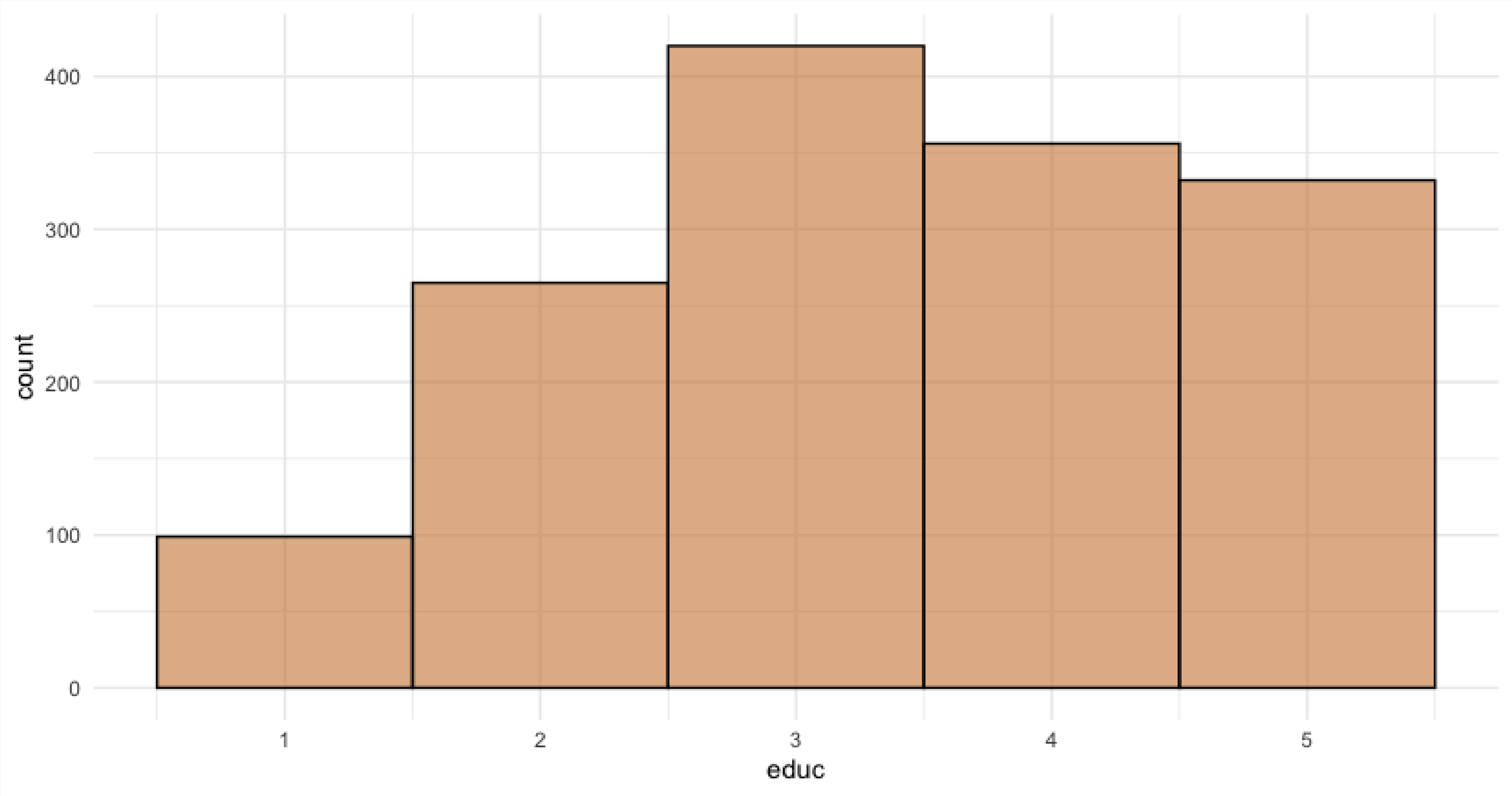
$$\mathbf{log}(\boldsymbol{wage}) = \beta_0 + \beta_1 experience + \beta_2 education + u$$

- Where wage is gross hourly wage rate in euro, education is education level from 1 [low] to 5 [high], and experience is years of experience. The data covers a sample of wage-earners from Belgium in 1994.

- Before running any model, I want to look at the distribution of my variables.



density
0.03
0.02
0.01
0.00
0
10
20
30
40
exper


count
400
300
200
100
0
1
2
3
4
5
educ

$$\log(\boldsymbol{wage}) = \beta_0 + \beta_1 experience + \beta_2 education + u$$

	(1)
(Intercept)	1.513***
	(0.030)
exper	0.016***
	(0.001)
educ	0.159***
	(0.007)
Num.Obs.	1472
R2	0.351

* p < 0.1, ** p < 0.05, *** p < 0.01

1. How would you interpret the Intercept coefficient? Does it make any sense to interpret the Intercept?

The intercept coefficient would suggest that, all else equal, a worker in Belgium in 1994 with no experience (just entering the work force) and with less than low education (?) will receive, on average, an hourly wage of ln(1.51) euros, or $e^{1.51} = 4.53$ euros.

2. How would you interpret the *exper* coefficient?

Our model would suggest that, all else equal, an additional year of experience of a worker in Belgium in 1994 will increase their wage, on average, by 1.6% and this relation is statistically significant at the 95% level.

3. How would you interpret the *educ* coefficient?

Our model would suggest that, all else equal, wages in Belgium in 1994 increase, on average, by 16% for every additional an additional level of education (p<0.01).

$$\log(\boldsymbol{wage}) = \beta_0 + \beta_1 experience + \beta_2 education + u$$

	(1)
(Intercept)	1.513***
	(0.030)
exper	0.016***
	(0.001)
educ	0.159***
	(0.007)
Num.Obs.	1472
R2	0.351

* p < 0.1, ** p < 0.05, *** p < 0.01

What about substantive significance?

Our model would suggest that, all else equal, moving one standard deviation in the distribution of experience (10 years) of workers in Belgium in 1994 will increase their wage, on average, by 16.3%. This is similar to the gains from an additional level of education.

What other information do have?

We know that our model explains 35.1% of the variation.

Can we argue that *increases* in experience is causing *increases* in wages? Can we argue that *increases* in education is causing *increases* in wages?