Georgia Institute of Technology

Statistics for Economics

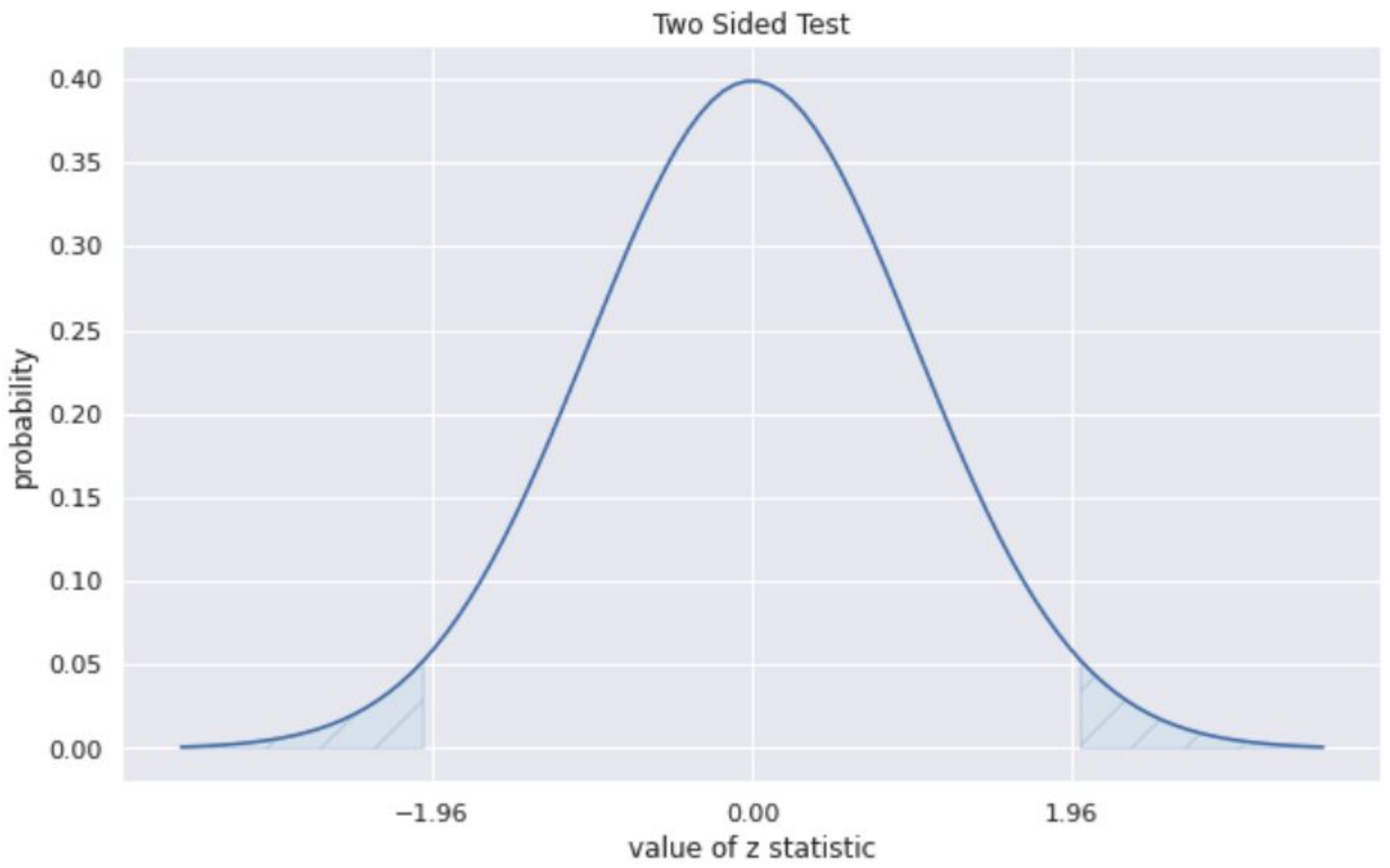
jim.crozier@gatech.edu, 2023

Today

- Review T-test
- Review confidence intervals
- Introduce binomial t-test

What are the three things that economists are looking for in applied analysis?

desired alpha level	two-sided test	one-sided test
0.100	1.644854	1.281552
0.050	1.959964	1.644854
0.010	2.575829	2.326348
0.001	3.290527	3.090232


Two Sided Test
probability
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−1.96
0.00
1.96
value of z statistic

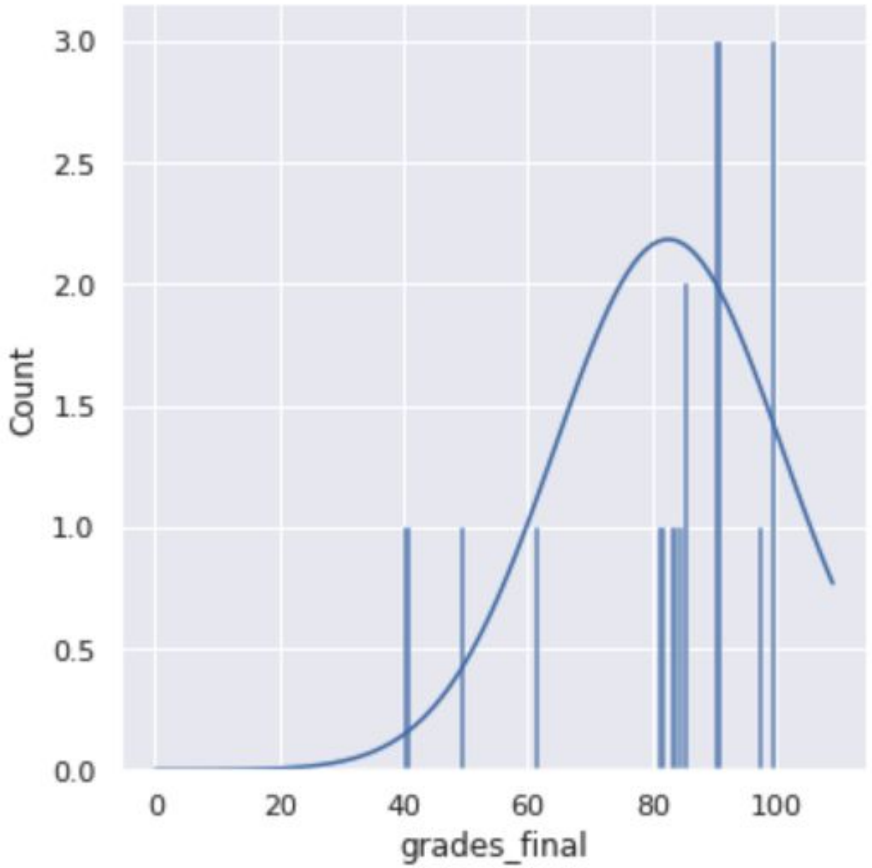
(simple) T Tests = z-test

$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{N}}$$

#1. We can make a normal with the mean and the standard deviation

$$\bar{x} = \frac{\sum x}{15} = \frac{1238}{15} = 82$$

$$\sigma = \sqrt{\frac{(\sum x - \bar{x})^2}{N-1}} = \sqrt{\frac{4672}{14}} = 18.2$$

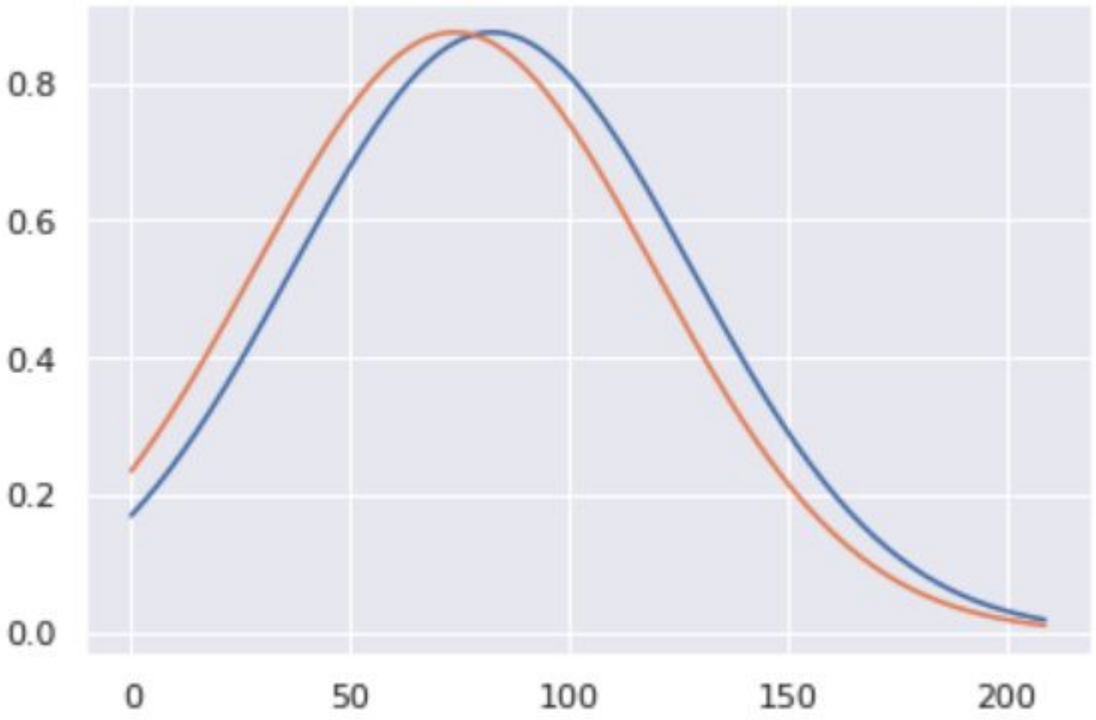


Can use the abstraction to compare to other abstractions:

Compare 2 classes:

Class1: 82 (18)

Class2: 72 (18)



T-test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}}$$

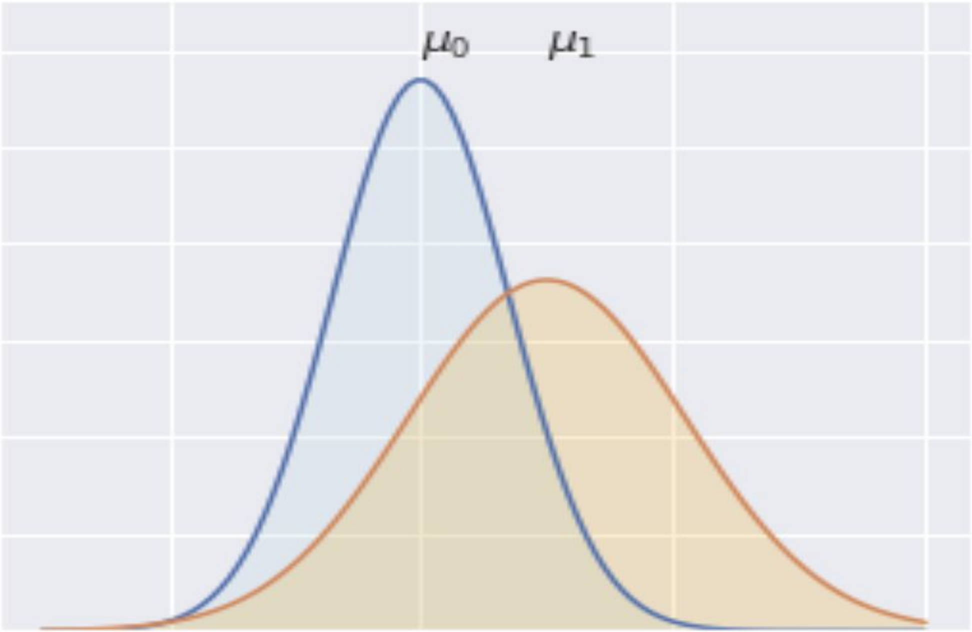
where

$$SE = \sqrt{\frac{\sigma_1^2}{N1} + \frac{\sigma_2^2}{N2}}$$

Difference between two groups with different sample sizes and variance

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

- Classroom 1
  - Avg: 82
  - Stdev: 18
  - Size: 15
- Classroom 2
  - Avg: 72
  - Stdev: 22
  - Size: 30



- Classroom 1
  - Avg: 82
  - Stdev: 18
  - Size: 15
- Classroom 2
  - Avg: 72
  - Stdev: 22
  - Size: 24

$$ttest = \frac{\mu_1 - \mu_2}{\sqrt{\frac{stdev_1^2}{n_1} + \frac{stdev_2^2}{n_2}}}$$

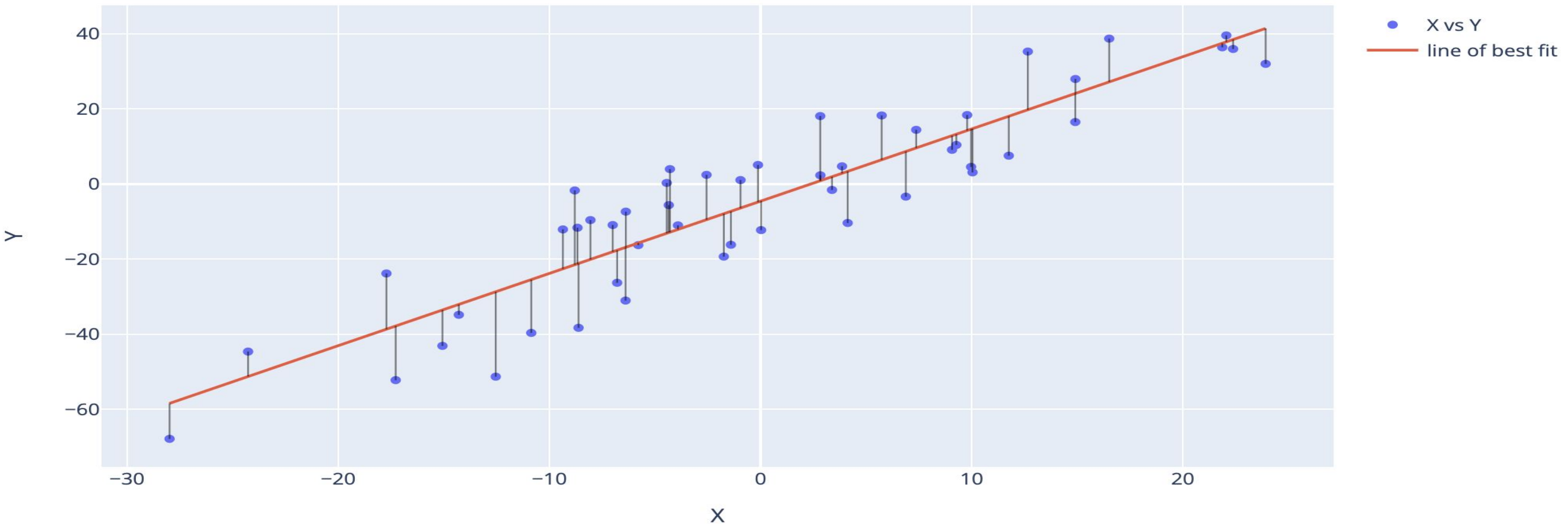
$$t - stat = \frac{82 - 72}{\sqrt{\frac{18^2}{15} + \frac{22^2}{24}}} = \frac{10}{6.46} = 1.54$$

Linear Regression

$$y_i = a + b * x_i + u_i$$

$$\hat{y}_i = \hat{a} + \hat{b} * x_i$$

$$\hat{u}_i = y_i - \hat{y}_i$$



$$\hat{y}_i = \hat{a} + \hat{b} * x_i$$

$$\hat{b} = \frac{\text{cov}(x, y)}{\text{var}(x)} = \frac{\sum(x_i - \bar{x}) \sum(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$\hat{a} = \bar{y} - \hat{b} * \bar{x}$$

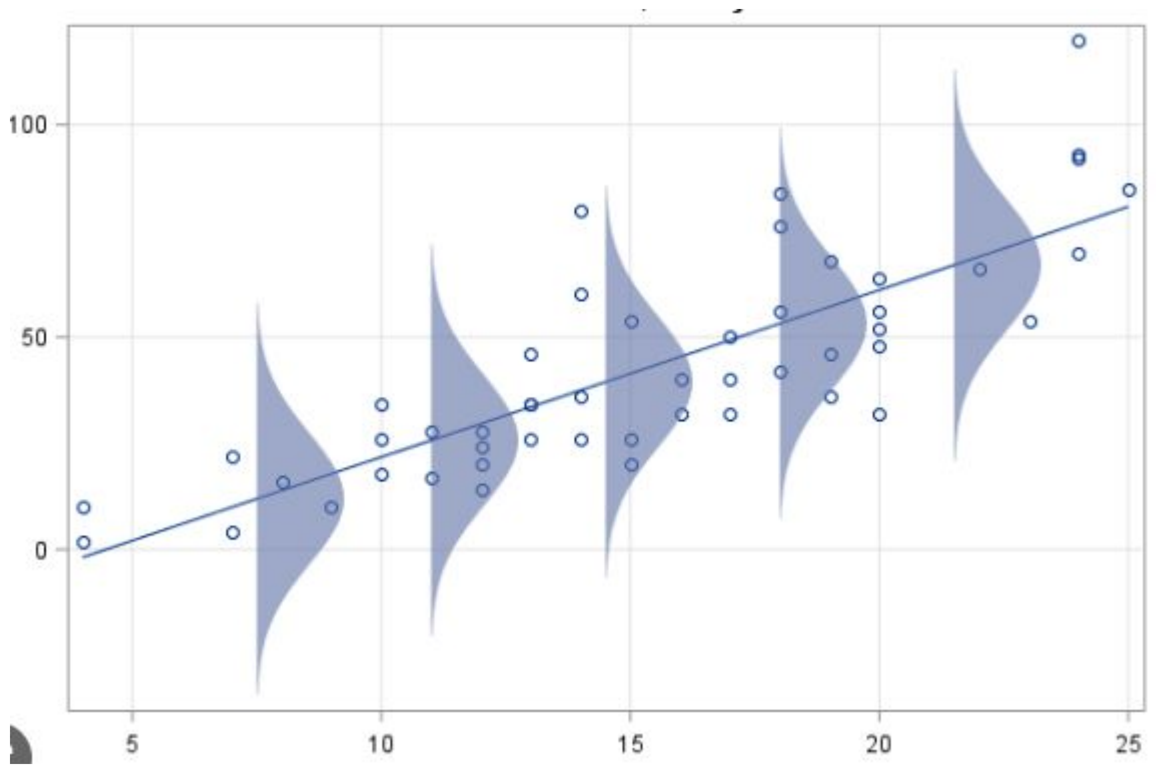
Walk through example

								yhat	y-yhat
x	y	demean_x	demean_x_sq	demean_y	demean_y_sq	demean_x*demean_y		b0 + b1*x	error
3	6	-2.4	5.76	1.6	2.56	-3.84		4.90	1.10
5	9	-0.4	0.16	4.6	21.16	-1.84		4.48	4.52
2	2	-3.4	11.56	-2.4	5.76	8.16		5.11	-3.11
8	1	2.6	6.76	-3.4	11.56	-8.84		3.85	-2.85
9	4	3.6	12.96	-0.4	0.16	-1.44		3.65	0.35
			37.2		41.2	-7.8	sum		
						-1.95	sum/(n-1)		

mean_y	4.4
mean_x	5.4

numerator	-1.95	-1.95	**-0.199**	**correlation**
denom	sqrt(37.2/4 * 41.2/4)	9.79		

slope	-0.210
intercept	5.53


100
50
0
5
10
15
20
25

Confidence Interval

H_0: b_1= 0 (we don't believe there is a slope (relationship))

H_A: b_1 not equal 0 (we believe there is a slope (relationship))

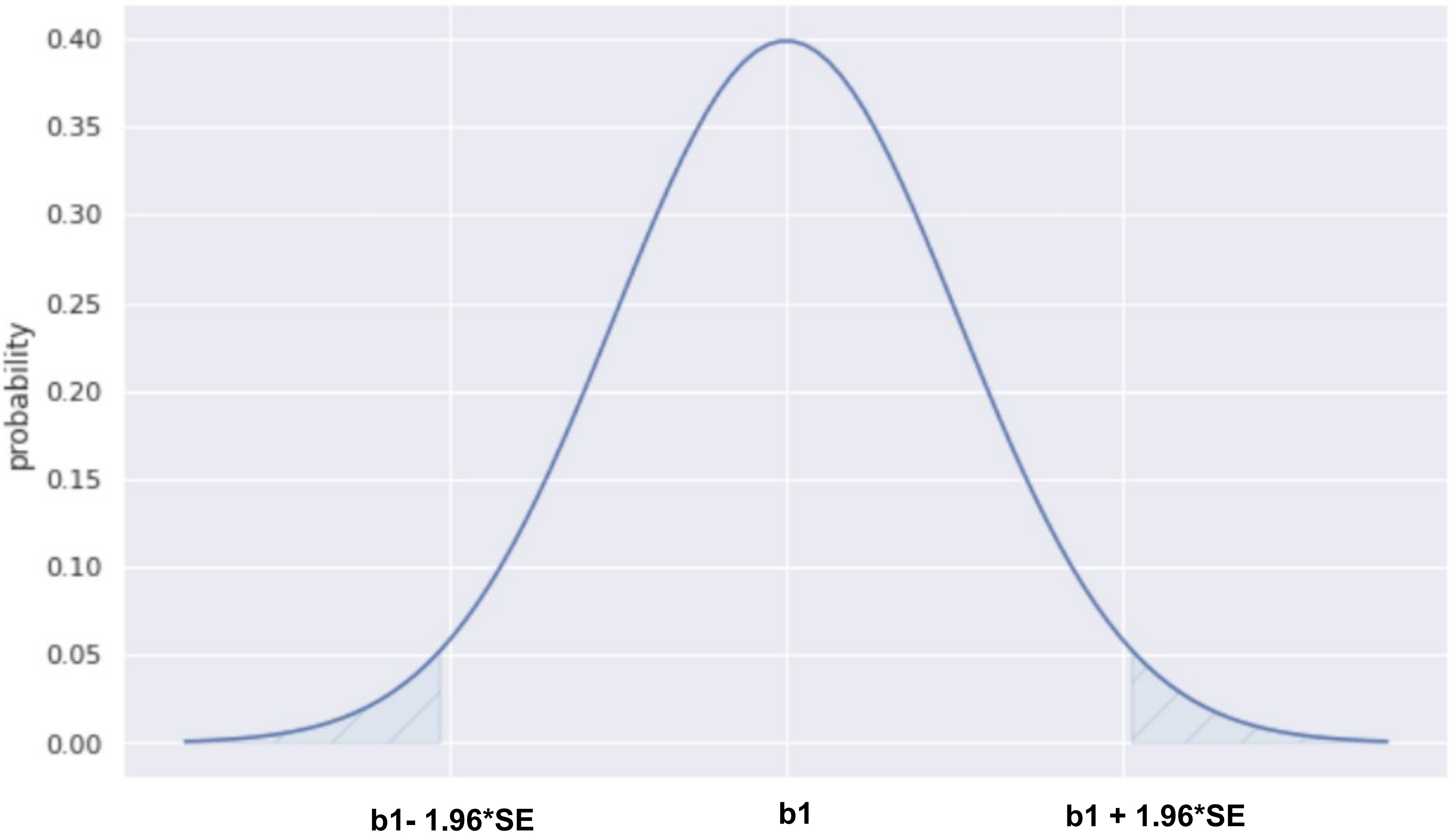
$$t = \frac{b_1 - 0}{SE_{b_1}} = \frac{b_1}{SE_{b_1}}$$

Confidence interval is b1 +/- CV* SE, where CV is critical value. This is easier to show than anything else (see notebook).

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

$$(\bar{x}_1 - \bar{x}_2) = SE * t$$

$$(\bar{x}_1 - \bar{x}_2) \pm SE * t = 0$$


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
probability
b1- 1.96*SE
b1
b1 + 1.96*SE

```
==============================================================================
Dep. Variable:                      y   R-squared:                       0.269
Model:                            OLS   Adj. R-squared:                  0.254
Method:                 Least Squares   F-statistic:                     17.68
Date:                Tue, 04 Apr 2023   Prob (F-statistic):           0.000113
Time:                        12:59:08   Log-Likelihood:                -294.80
No. Observations:                  50   AIC:                             593.6
Df Residuals:                      48   BIC:                             597.4
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     15.4229     12.700      1.214      0.231     -10.113      40.959
x             -5.7164      1.360     -4.204      0.000      -8.450      -2.983
==============================================================================
Omnibus:                        1.825   Durbin-Watson:                   1.558
Prob(Omnibus):                  0.401   Jarque-Bera (JB):                1.428
Skew:                          -0.414   Prob(JB):                        0.490
Kurtosis:                       2.975   Cond. No.                         9.34
==============================================================================
```

	coef	std err	t	$P>\lvert t\rvert$	[0.025	0.975]
Intercept	15.4229	12.700	1.214	0.231	-10.113	40.959
x	$\frac{Cov(x,y)}{Var(x)}$	1.360	$\frac{b1}{SE}$	0.000	b1-1.96*b1	b1+1.96*b1

Binary Standard Error

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

When an estimate is of a binomial (1,0) there are some simplifications in the variance.

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

See here if you are interested in the derivation https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$$

A newspaper says a randomized national survey found that 40% of the population has covid.
What are the confident intervals with different n?

$$n = 10, \qquad \sqrt{\frac{0.4(1-0.4)}{10}} = 0.15, CI = [10\%, 70\%]$$

$$n = 100, \qquad \sqrt{\frac{0.4(1-0.4)}{100}} = 0.05, CI = [30\%, 50\%]$$

$$n = 1000, \qquad \sqrt{\frac{0.4(1-0.4)}{1,000}} = 0.015, CI = [37\%, 43\%]$$

$$n = 10,000, \qquad \sqrt{\frac{0.4(1-0.4)}{10,000}} = 0.015, CI = [39\%, 41\%]$$