ECON 113 DI A/B

Dizhi Wang

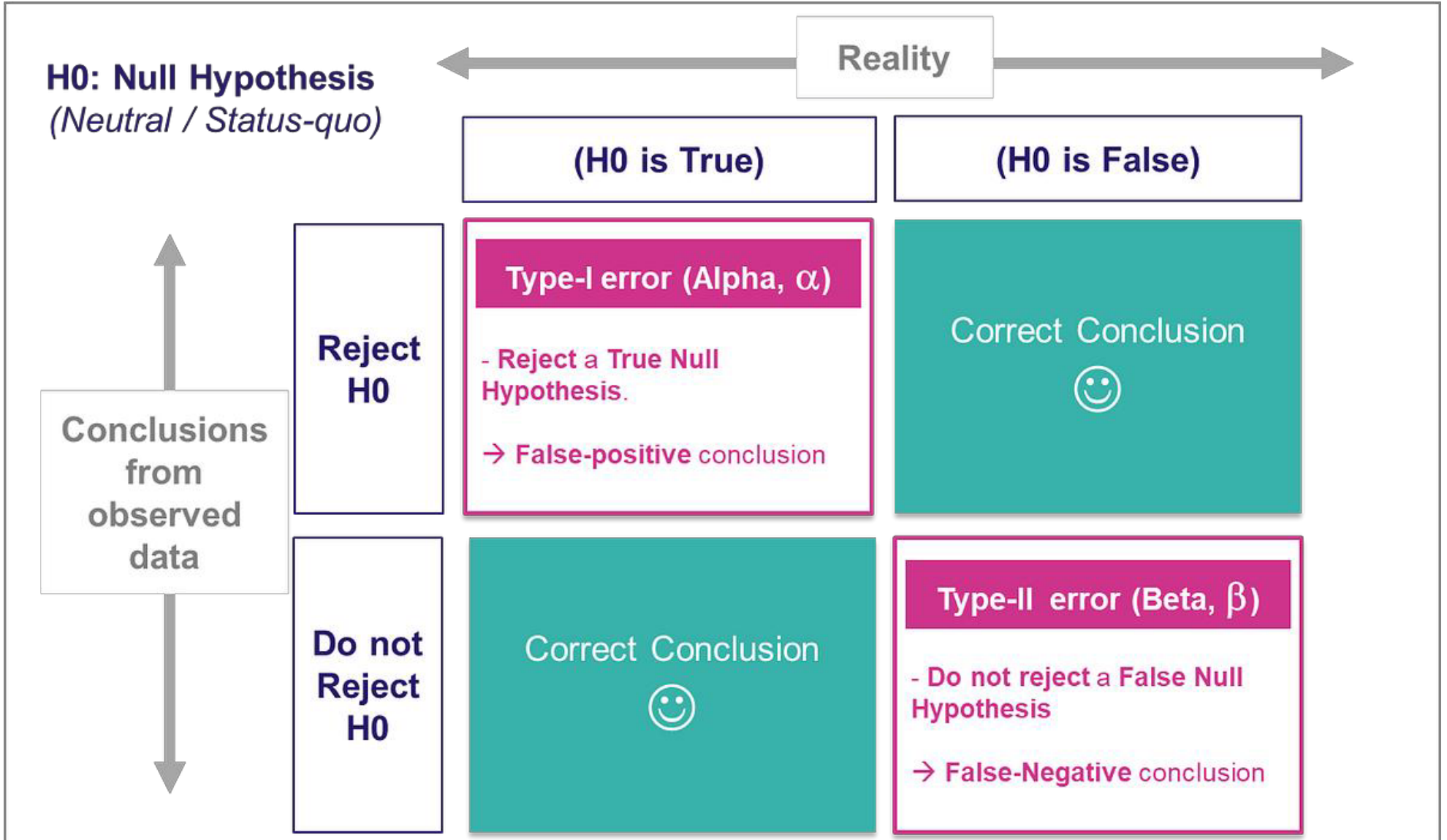
OH : 8:30 am – 10:30 am E2 403B

Agenda

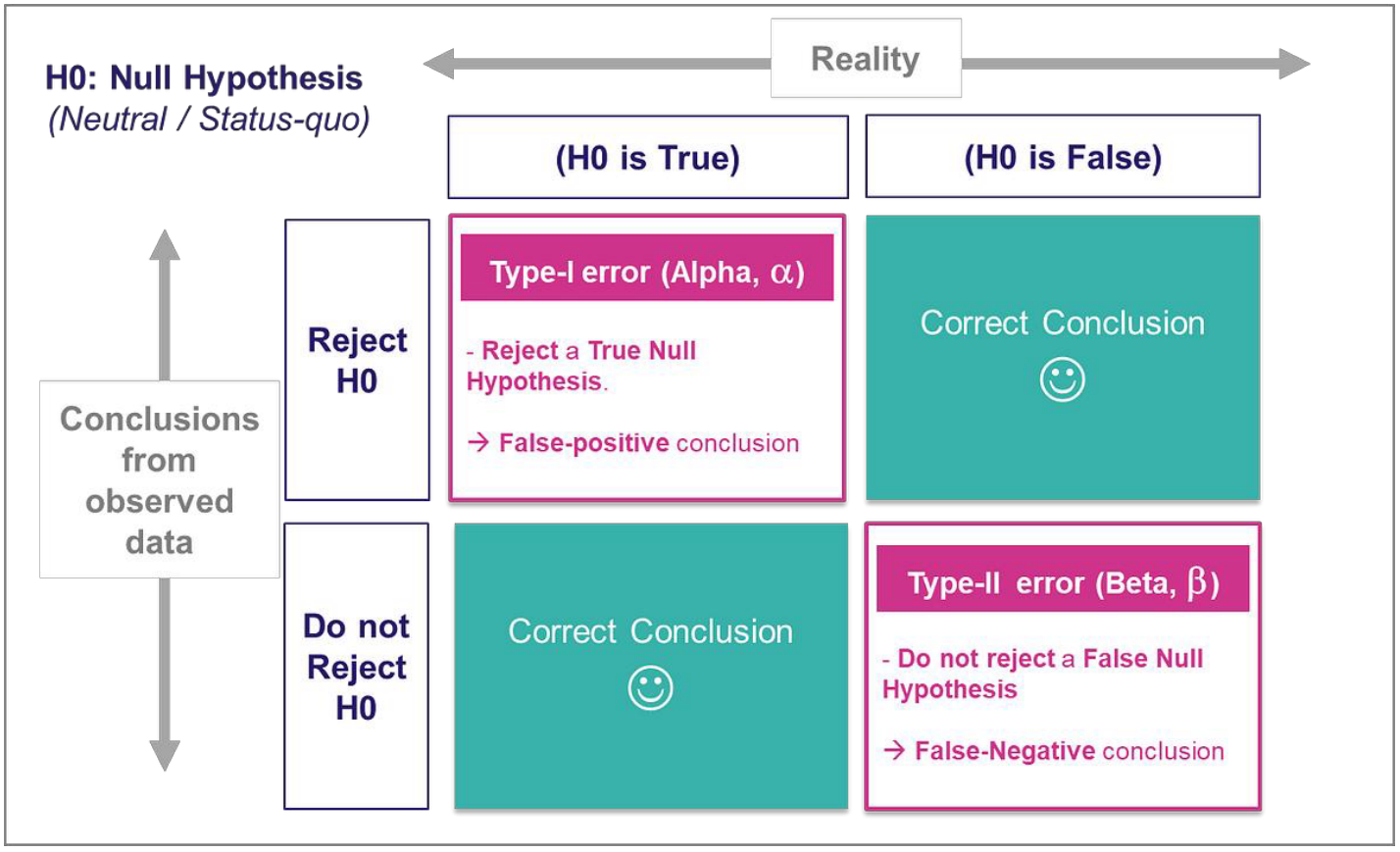
- Type I and Type II error
- Work through selected problem-set questions

Type I vs Type II error

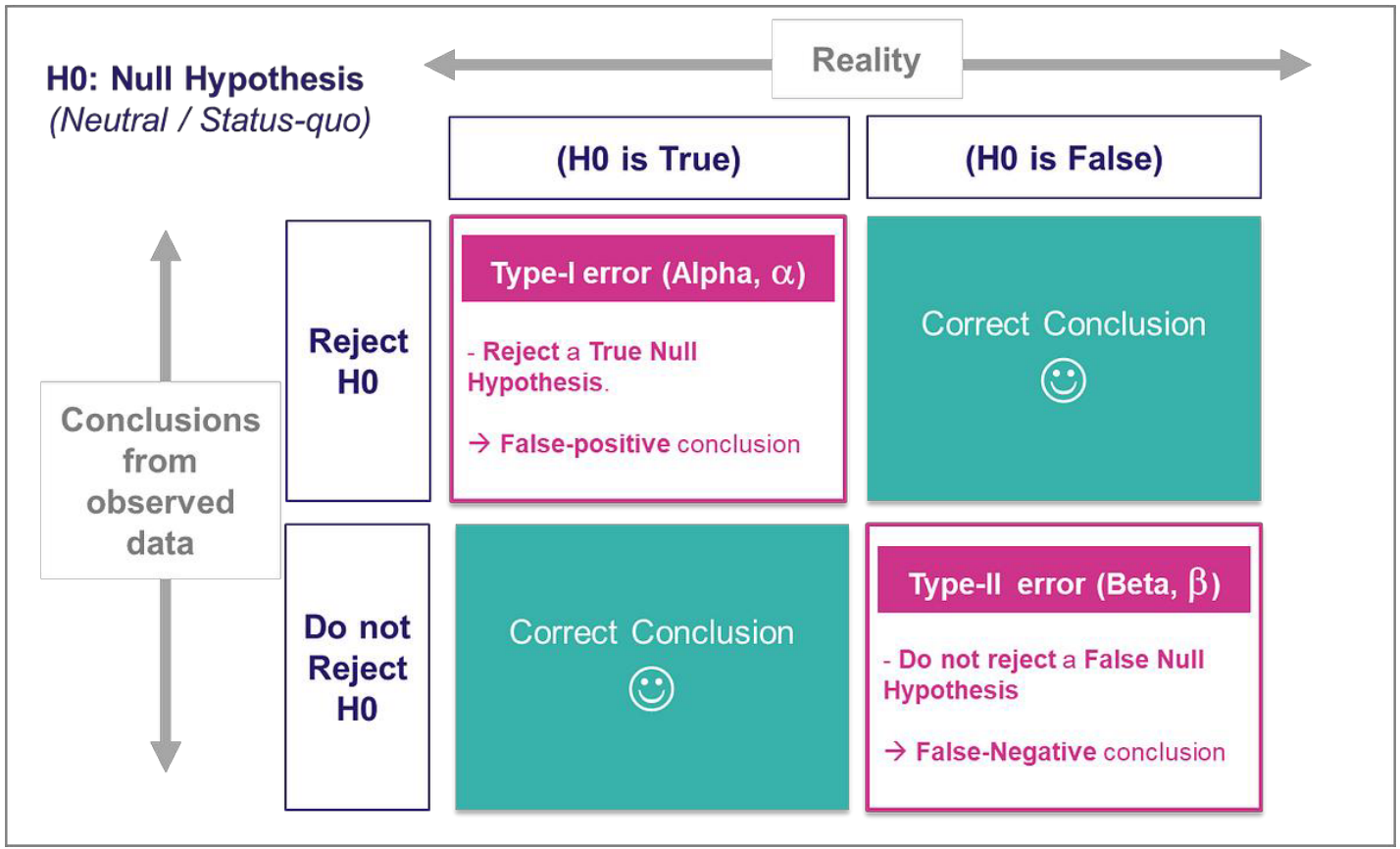
		True State of Nature	
		H_0 Is true	H_a Is true
Conclusion	Support H_0 / Reject H_a	Correct Conclusion	Type II Error
	Support H_a / Reject H_0	Type I Error	Correct Conclusion (Power)


H0: Null Hypothesis
(Neutral / Status-quo)
Reality
(H0 is True)
(H0 is False)
Conclusions from observed data
Reject H0
Do not Reject H0
Type-I error (Alpha, α)
- Reject a True Null Hypothesis.
→ False-positive conclusion
Correct Conclusion
Correct Conclusion
Type-II error (Beta, β)
- Do not reject a False Null Hypothesis
→ False-Negative conclusion

15. Suppose that your decision rule is to never reject the null hypothesis. What is your type I error?



15. Suppose that your decision rule is to never reject the null hypothesis. What is your type I error?

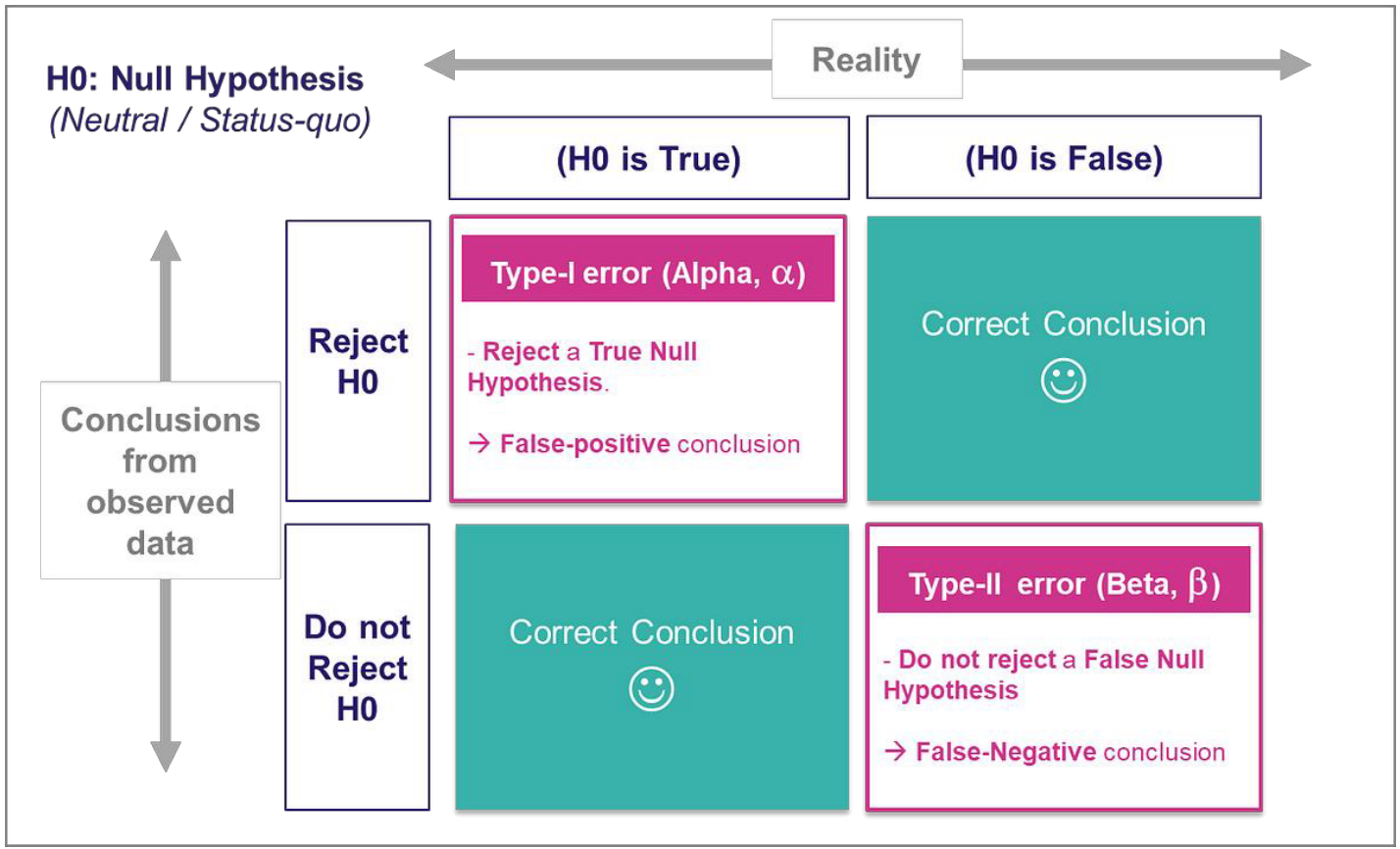


$$\alpha = \Pr(Type\ I\ error)$$
$$= \Pr(\text{reject H0} \mid \text{H0 is true})$$
$$= \frac{\Pr(\text{reject H0} \cap \text{H0 is true})}{\Pr(\text{H0 is true})}$$

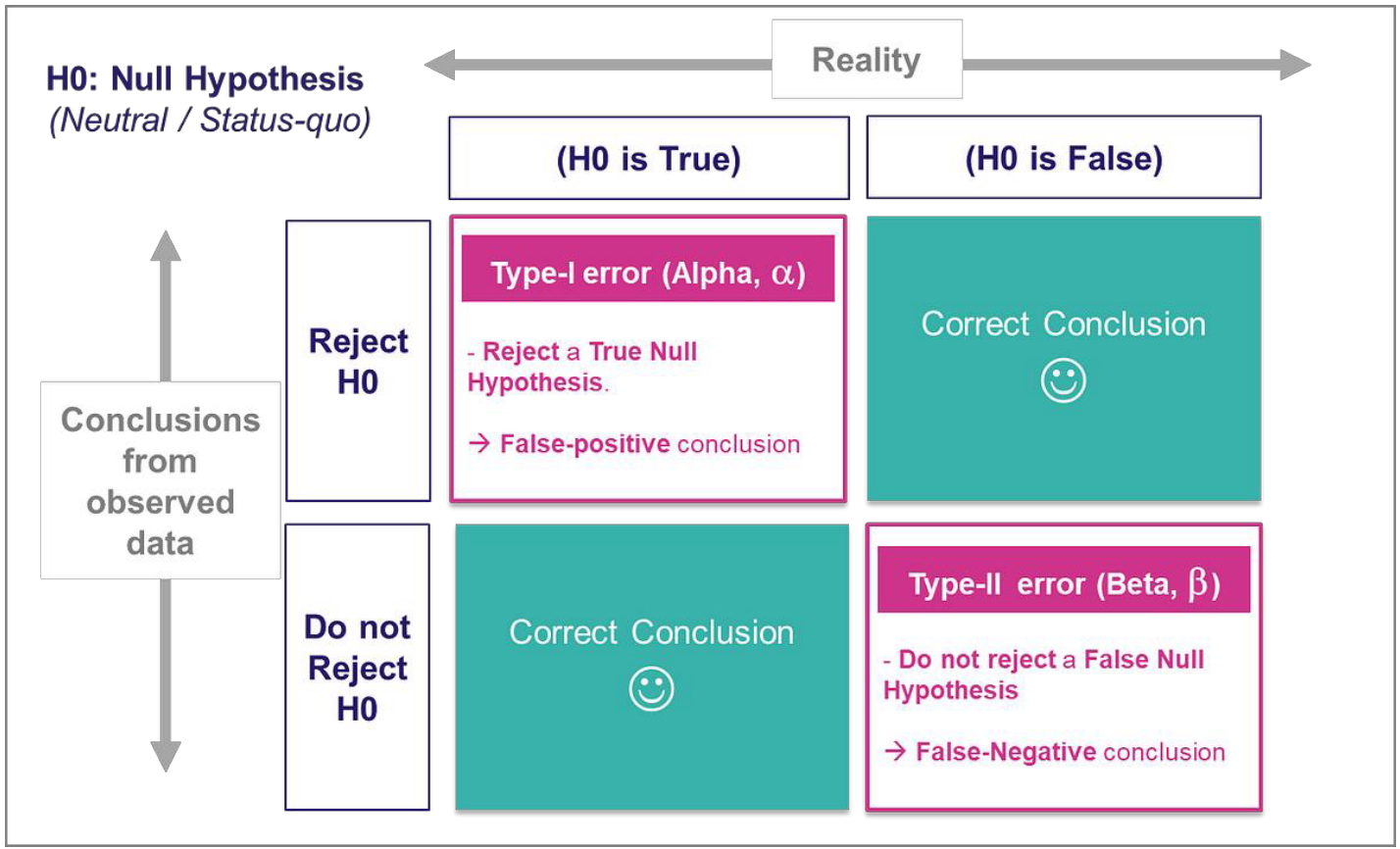
$$\Pr(\text{reject H0}) = 0$$
$$\Pr(\text{reject H0} \cap \text{H0 is true}) = 0$$

Never reject H0 $\Rightarrow \alpha = 0$

15. Suppose that your decision rule is to never reject the null hypothesis. What is your type ~~I~~ II error?



15. Suppose that your decision rule is to never reject the null hypothesis. What is your type ~~I~~ II error?

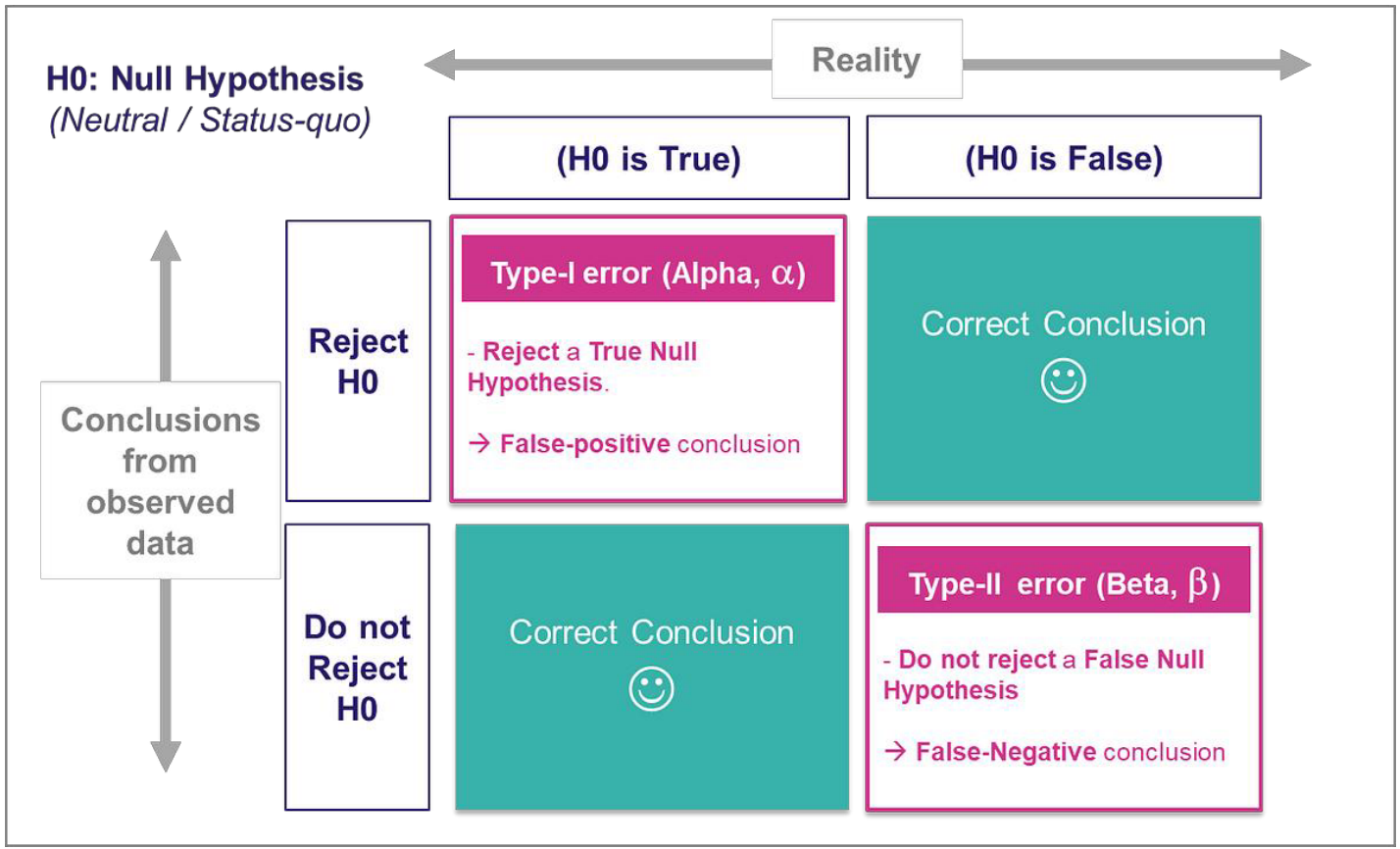


$$\beta = \Pr(Type\ II\ error)$$
$$= \Pr(\text{Don't reject H0} \mid \text{H0 is false})$$
$$= \frac{\Pr(\text{Don't reject H0} \cap \text{H0 is false})}{\Pr(\text{H0 is false})}$$

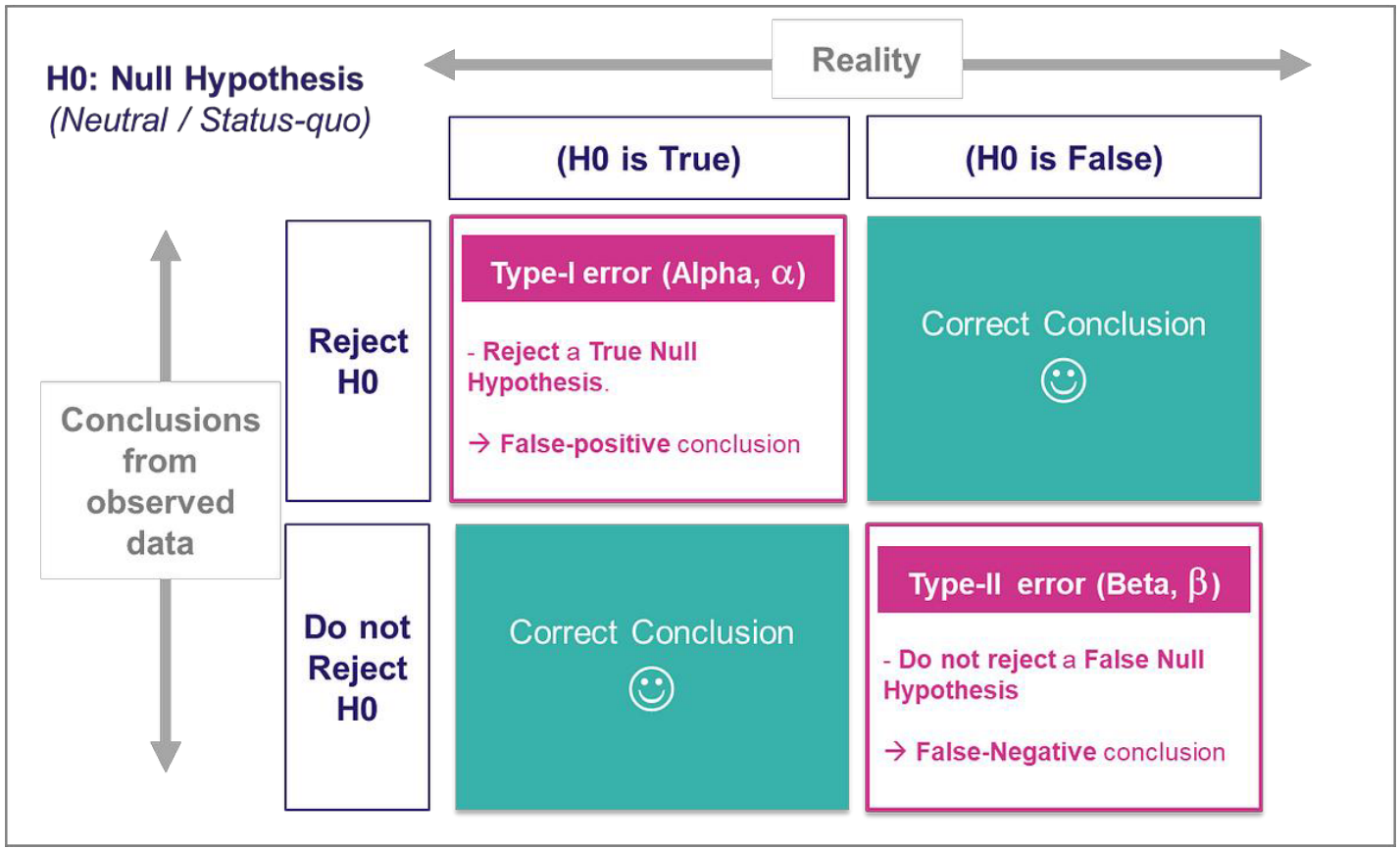
$$\Pr(\text{Don't reject H0} \cap \text{H0 is false}) = \Pr(\text{H0 is false})$$

Never reject H0 $\Rightarrow \beta = 1$

15. Suppose that your decision rule is to ~~never~~ always reject the null hypothesis. What is your type I error?



15. Suppose that your decision rule is to ~~never~~ always reject the null hypothesis. What is your type I error?

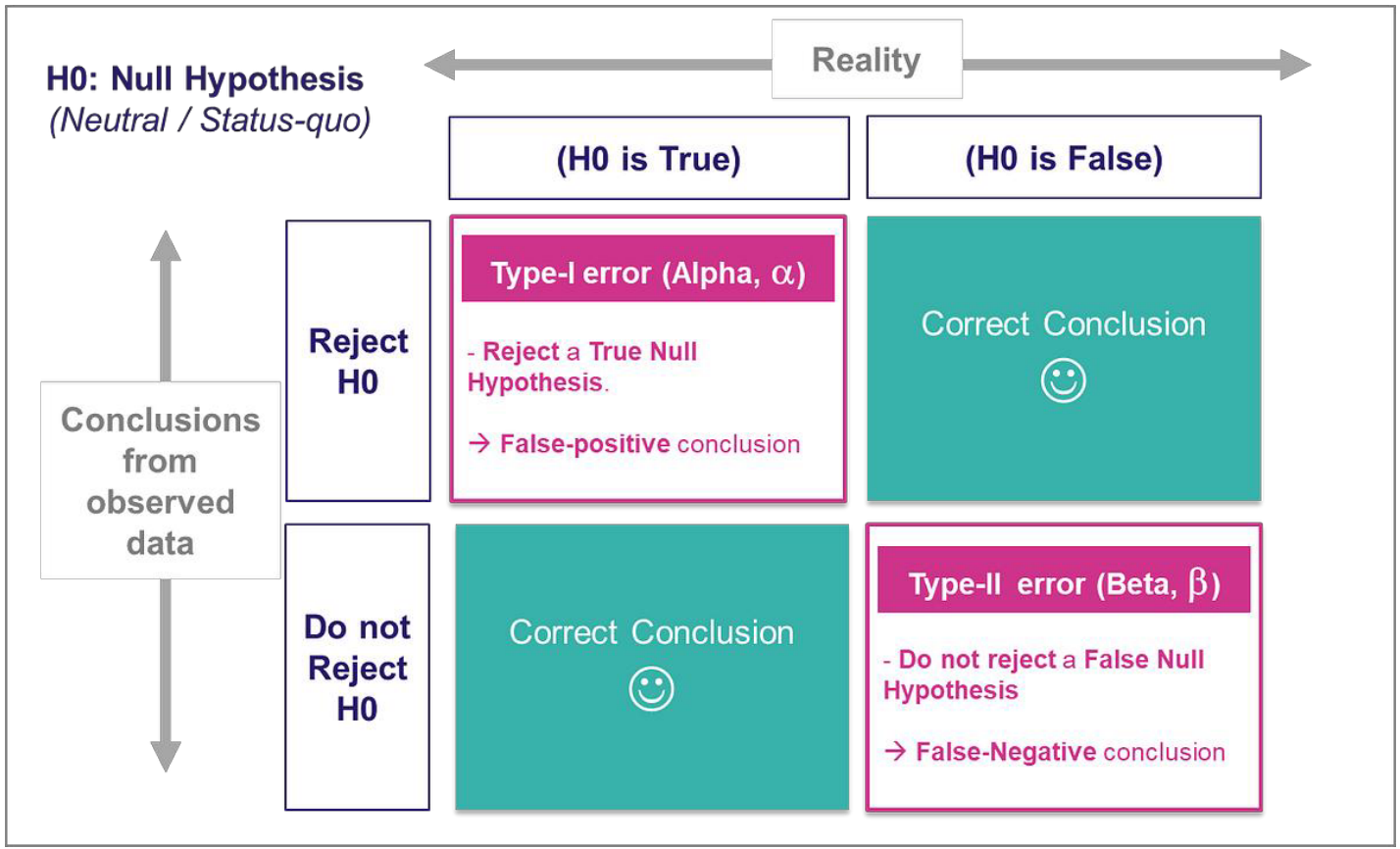


$$\alpha = \Pr(Type\ I\ error)$$
$$= \Pr(\text{reject H0} \mid \text{H0 is true})$$
$$= \frac{\Pr(\text{reject H0} \cap \text{H0 is true})}{\Pr(\text{H0 is true})}$$

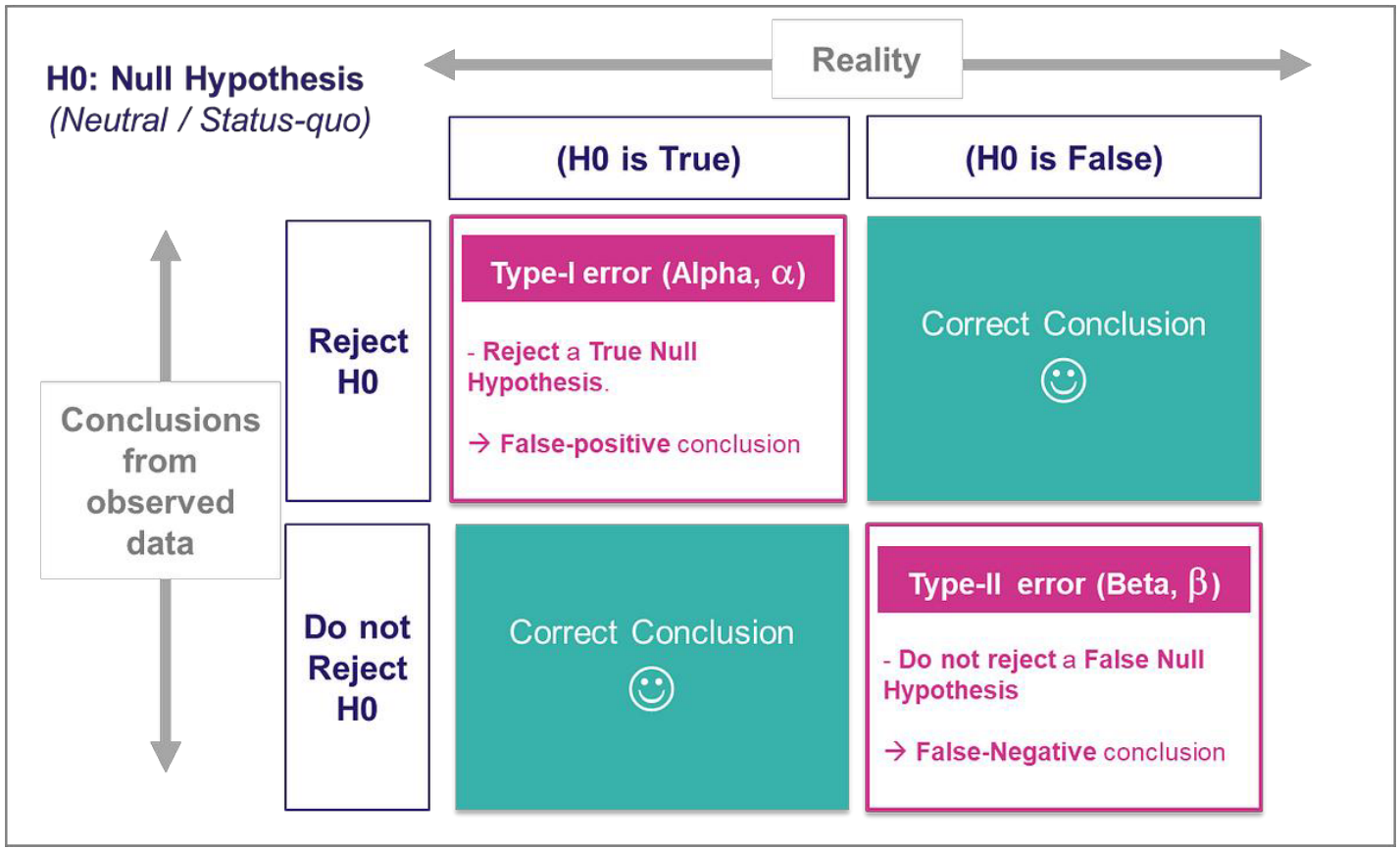
$$\Pr(\text{reject H0} \cap \text{H0 is true}) = \Pr(\text{H0 is true})$$

Always reject H0 $\Rightarrow \alpha = 1$

15. Suppose that your decision rule is to ~~never~~ always reject the null hypothesis. What is your type ~~I~~ II error?



15. Suppose that your decision rule is to ~~never~~ always reject the null hypothesis. What is your type ~~I~~ II error?



$$\beta = \Pr(Type\ II\ error)$$
$$= \Pr(\text{Don't reject H0} \mid \text{H0 is false})$$
$$= \frac{\Pr(\text{Don't reject H0 \& H0 is false})}{\Pr(\text{H0 is false})}$$

Always reject H0 $\Rightarrow \beta = 0$

4. Consider the MLR model with three independent variables, under the classical linear assumptions MLR.1 through MLR.6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

You would like to test the null hypothesis $H_0 : \beta_1 - 3\beta_2 = 1$.

(a) Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the OLS estimators of β_1 and β_2. Find $V(\hat{\beta}_1 - 3\hat{\beta}_2)$ in terms of the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ and the covariance between them. What is the standard error of $\hat{\beta}_1 - 3\hat{\beta}_2$?.

$$\mathrm{Var}(\hat{\beta}_1 - 3\hat{\beta}_2) = \mathrm{Var}(\hat{\beta}_1) + 9\,\mathrm{Var}(\hat{\beta}_2) + 2\mathrm{Cov}(\hat{\beta}_1, -3\hat{\beta}_2)$$

$$= \mathrm{Var}(\hat{\beta}_1) + 9\,\mathrm{Var}(\hat{\beta}_2) - 6\,\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2)$$

$$\widehat{\mathrm{Var}}(\hat{\beta}_1 - 3\hat{\beta}_2) = \widehat{\mathrm{Var}}(\hat{\beta}_1) + 9\widehat{\mathrm{Var}}(\hat{\beta}_2) - 6\,\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$$

$$\mathrm{Se}(\hat{\beta}_1 - 3\hat{\beta}_2) = \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_1) + 9\widehat{\mathrm{Var}}(\hat{\beta}_2) - 6\,\widehat{\mathrm{Cov}}(\hat{\beta}_1, \hat{\beta}_2)}$$

Note:

$$se(\hat{\beta}_1) = \sqrt{\widehat{Var}(\hat{\beta}_1)}$$

$$se(\hat{\beta}_2) = \sqrt{\widehat{Var}(\hat{\beta}_2)}$$

$$Se(\hat{\beta}_1 - 3\hat{\beta}_2) = \sqrt{se^2(\hat{\beta}_1) + 9se^2(\hat{\beta}_2) - 6\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$$

$$= \sqrt{se^2(\hat{\beta}_1) - 6\underbrace{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}_{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) \neq Se(\hat{\beta}_1)\, Se(\hat{\beta}_2)} + 9se^2(\hat{\beta}_2)}$$

$$\neq Se(\hat{\beta}_1) - 3\,Se(\hat{\beta}_2)$$

★ Can't directly use the se showed from the multi-regressor regression to test the hypothesis

Y	Coeff	Se	t	P	CI
_const	$\hat{\beta}_0$	$Se(\hat{\beta}_0)$	t_0	$p>..$	(,)
X_1	$\hat{\beta}_1$	$Se(\hat{\beta}_1)$	t_1	$p>..$	(,)
X_2	$\hat{\beta}_2$	$Se(\hat{\beta}_2)$	t_2	$p>...$	(,)

4. Consider the MLR model with three independent variables, under the classical linear assumptions MLR.1 through MLR.6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

You would like to test the null hypothesis $H_0 : \beta_1 - 3\beta_2 = 1$.

(b) Write the t statistic for testing $H_0 : \beta_1 - 3\beta_2 = 1$.

$H_0 : \mu = \mu_0 \Rightarrow$ t-statistic is always : $t = \frac{\mu - \mu_0}{se(\hat{\mu})}$

$$\text{t-statistic} = \frac{\beta_1 - 3\beta_2 - 1 - 0}{se(\widehat{\beta_1 - 3\beta_2 - 1})}$$

4. Consider the MLR model with three independent variables, under the classical linear assumptions MLR.1 through MLR.6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

You would like to test the null hypothesis $H_0 : \beta_1 - 3\beta_2 = 1$.

(c) Define $\theta_1 = \beta_1 - 3\beta_2$, and $\hat{\theta}_1 = \hat{\beta}_1 - 3\hat{\beta}_2$ Write down a regression equation involving β_0, θ_1, β_2 and β_3 that allows you to directly obtain $\hat{\theta}_1$ and its standard error.

part (a) shows that $se(\hat{\beta}_1 - 3\hat{\beta}_2)$ can't be achieved by combining the $se(\hat{\beta}_1)$ & $se(\hat{\beta}_2)$ that we have already known from the regression table

⇒ Construct the equation in a way s.t it gives se specifically for $(\hat{\beta}_1 - 3\hat{\beta}_2)$

⇒ Construct the equation that includes the coefficient

$(\beta_1 - 3\beta_2) = \theta_1$ in it & estimate it

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$= \beta_0 + \theta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u + ?$$

what ? should be to make the equation hold.

$$\beta_0 + (\beta_1 - 3\beta_2) x_1 + \beta_2 x_2 + \beta_3 x_3 + u + ?$$

$$= \beta_0 + \beta_1 x_1 - 3\beta_2 x_1 + \beta_2 x_2 + \beta_3 x_3 + u + ? \quad (+3\beta_2 x_1)$$

compare with

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$\Rightarrow y = \beta_0 + \theta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + 3\beta_2 x_1 + u$$

$$= \beta_0 + \theta_1 x_1 + \beta_2 (x_2 + 3x_1) + \beta_3 x_3 + u$$

We are going to regress y on x_1, $x_2 + 3x_1$, x_3

In stata. you will need to

① gen new_var = $x_2 + 3x_1$

② reg y x_1 new_var x_3 $\Rightarrow$

y	coeff	se
x_1	$\hat{\theta}_1$	$se(\hat{\theta}_1)$
new_var	$\hat{\beta}_2$	$se(\hat{\beta}_2)$
x_3	$\hat{\beta}_3$	$se(\hat{\beta}_3)$

5. **STATA Exercise** The following model can be used to study whether campaign expenditures affect election outcomes:

$$\text{voteA} = \beta_0 + \beta_1 \log(\text{expendA}) + \beta_2 \log(\text{expendB}) + \beta_3 \text{prtystrA} + u,$$

where voteA is the percentage of the vote received by Candidate A, expendA and expendB are campaign expenditures by Candidates A and B, and prtystrA is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

(a) What is the interpretation of β_1?

Level-log regression:

1% increase in x is associated to $\frac{\beta_1}{100}$ units change in y.

A 1% increase in expendA is associated to $\frac{\beta_1}{100}$ units change in voteA in the unit of percentage change.

What does it mean ?

1% ↑ in expendA ⇒ voteA to increase, e.g.

from 2% → $(2+\frac{\beta_1}{100})$%

This is an increase of $\frac{\beta_1}{100}$ percentage point

If we are saying $\frac{\beta_1}{100}$ percentage,

it means voteA increase from 2% to α% where

$$\left(\frac{\alpha-2}{2}\right) \times 100\% = \frac{\beta_1}{100}\%$$

$$\frac{\alpha-2}{2} = \frac{\frac{\beta_1}{100}\%}{100\%} = \frac{\beta_1}{10000}\%$$

$$\alpha - 2 = \frac{2\beta_1}{10000}\%$$

$$\alpha = 2 + \frac{2\beta_1}{10000}\%$$

5. **STATA Exercise** The following model can be used to study whether campaign expenditures affect election outcomes:

$$\text{voteA} = \beta_0 + \beta_1 \log(\text{expendA}) + \beta_2 \log(\text{expendB}) + \beta_3 \text{prtystrA} + u,$$

where voteA is the percentage of the vote received by Candidate A, expendA and expendB are campaign expenditures by Candidates A and B, and prtystrA is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

(b) In terms of the parameters, state the null hypothesis that a 1% increase in A's expenditures is offset by a 1% increase in B's expenditures.

effect of 1% increase in expand A on Vote A $\Rightarrow \frac{\beta_1}{100}\%$

effect of 1% increase in expandB on Vote A $\Rightarrow \frac{\beta_2}{100}\%$

$H_0 : \frac{\beta_1}{100}\% + \frac{\beta_2}{100}\% = 0$

$\Rightarrow H_0 : \beta_1 + \beta_2 = 0$

5. **STATA Exercise** The following model can be used to study whether campaign expenditures affect election outcomes:

$$\text{voteA} = \beta_0 + \beta_1 \log(\text{expendA}) + \beta_2 \log(\text{expendB}) + \beta_3 \text{prtystrA} + u,$$

where voteA is the percentage of the vote received by Candidate A, expendA and expendB are campaign expenditures by Candidates A and B, and prtystrA is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

(c) Estimate the given model using the data in VOTE1 and report the results in usual form. Do A's expenditures affect the outcome? What about B's expenditures? Can you use these results to test the hypothesis in part (b)?

Source	SS	df	MS	Number of obs	=	173
				F(3, 169)	=	215.15
Model	38402.42	3	12800.8067	Prob > F	=	0.0000
Residual	10054.8285	169	59.4960267	R-squared	=	0.7925
				Adj R-squared	=	0.7888
Total	48457.2486	172	281.728189	Root MSE	=	7.7134

voteA	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
logexpendA	6.081355	.3821139	15.92	0.000	5.327024	6.835686
logexpendB	-6.615631	.378889	-17.46	0.000	-7.363596	-5.867666
prtystrA	.1520142	.0620259	2.45	0.015	.0295689	.2744595
_cons	45.08788	3.926795	11.48	0.000	37.33599	52.83977

To test $H_0: \beta_1 + \beta_2 = 0$

t-statistic : $\dfrac{\beta_1 + \beta_2 - 0}{se(\widehat{\beta_1 + \beta_2})}$

From Q4, we learned that

$se(\widehat{\beta_1 + \beta_2}) \neq se(\hat{\beta_1}) + se(\hat{\beta_2})$

So we are not able to do this.

5. **STATA Exercise** The following model can be used to study whether campaign expenditures affect election outcomes:

$$\text{voteA} = \beta_0 + \beta_1 \log(\text{expendA}) + \beta_2 \log(\text{expendB}) + \beta_3 \text{prtystrA} + u,$$

where voteA is the percentage of the vote received by Candidate A, expendA and expendB are campaign expenditures by Candidates A and B, and prtystrA is a measure of party strength for Candidate A (the percentage of the most recent presidential vote that went to A's party).

(d) Estimate a model that directly gives the t statistic for testing the hypothesis in part (b). What do you conclude? (Use a two-sided alternative.)

$$\begin{cases} H_0 : \beta_1 + \beta_2 = 0 \\ H_a : \beta_1 + \beta_2 \neq 0 \end{cases}$$

We care about $\beta_1 + \beta_2$, denote $\beta_1 + \beta_2$ as θ_1

Rearrage the original mode sit it includes θ_1 as the coefficient directly.

Originally:

$$VoteA = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u. \quad ①$$

We change it s.t

$$VoteA = \beta_0 + \theta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u + ? \quad ②$$

Needs to find ? s.t equation ① & ② equal.

Expands ② gives

$$VoteA = \beta_0 + (\beta_1 + \beta_2) x_1 + \beta_2 x_2 + \beta_3 x_3 + u + ?$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_1 + \beta_2 x_2 + \beta_3 x_3 + u + ? \quad -\beta_2 x_1$$

Compare it with ①

$$Vothe\ A = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

⇒ Our rearranged equation should be.

$$VoteA = \beta_0 + \theta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u - \beta_2 x_1$$

$$= \beta_0 + \theta_1 x_1 + \beta_2 (x_2 - x_1) + \beta_3 x_3 + u.$$

which means we just need to run:

```
gen new_var = lexpandA - lexpandB
reg VoteA x1 newvar x3
```

⇒ `reg voteA lexpandA new_var prtystA`.

3. Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (*roe*, in percentage form), and return on the firm's stock (*ros*, in percentage form):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u.$$

(a) In terms of the model parameters, state the null hypothesis that, after controlling for *sales* and *roe*, *ros* has no effect on CEO salary. State the alternative that better stock market performance increases a CEO's salary.

$H_0 : \beta_3 = 0$

$H_a : \beta_3 > 0$

3. Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (*roe*, in percentage form), and return on the firm's stock (*ros*, in percentage form):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u.$$

(b) Using the data in CEOSAL1, the following equation was obtained by OLS:

$$\widehat{\log(salary)} = \underset{(.32)}{4.32} + \underset{(.035)}{.280}\log(sales) + \underset{(.0041)}{.0174}\,roe + \underset{(.00054)}{.00024}\,ros,$$

with $n = 209$, and $R^2 = .283$.. By what percentage is salary predicted to increase if *ros* increases by 50 points? Does *ros* have a practically large effect on salary ?

log-level : 1 unit ↑ in x ~ $100\beta_1$% ↑ in y.

1 unit ↑ in ros ~ 100×0.00024 % ↑ in y

50 units ↑ in ros ~ $50 \times 100 \times 0.00024$% ↑ in y

3. Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (*roe*, in percentage form), and return on the firm's stock (*ros*, in percentage form):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u.$$

(c) Test the null hypothesis that *ros* has no effect on salary against the alternative that *ros* has a positive effect. Carry out the test at the 10% significance level.

$H_0: \beta_3 = 0$

$H_a: \beta_3 > 0$

single side test.

$$t = \frac{\beta_3 - 0}{se(\beta_3)} = \frac{0.00024}{0.00054} = \frac{24}{54} = 0.44.$$

$Pr(t > 1.28) = 0.1$

$0.44 < 1.28$ Fail to reject. H_0

3. Consider an equation to explain salaries of CEOs in terms of annual firm sales, return on equity (*roe*, in percentage form), and return on the firm's stock (*ros*, in percentage form):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u.$$

(d) Would you include *ros* in a final model explaining CEO compensation in terms of firm performance? Explain.

No, we shouldn't remove a regressor just because it is insignificant

We includes regression based on whether it makes sense to have them in the economic model or not

Insignificant result here can just mean that the variation of ros in this sample is not big enough for us to have power to test its effect.