

Common Misconceptions

Optional (Due Fri, Dec 4)

Prof. Jack Reilly

Fall 2026

1. For each of the following 4 statements, explain what is wrong and why:
 - a. A parameter describes a sample.
 - b. Bias and variability are two names for the same thing.
 - c. The central limit theorem states that, for large n , the population mean is approximately Normal.
 - d. For large n , the distribution of observed values in the sample will be approximately Normal.

Confidence Intervals

2. A group of 36 students were surveyed to determine the average number of hours they spent studying for a final exam. The sample had a mean of 12 with a standard deviation of 3.
 - a. Calculate a 90% confidence interval for the mean.
 - b. Calculate a 95% confidence interval for the mean.
 - c. Calculate a 98% confidence interval for the mean.

A New Teaching Technique

3. A professor hopes to improve student learning in a Statistics course. To do so, she tries an experiment: in one of her sections, she tries a new teaching technique and in her other class, she does not. At the end of the course, she collects the following data: student grade and the class they are part of (Old Pedagogy or New Pedagogy).

	Old Pedagogy	New Pedagogy
A	12	13
B	8	12
C	4	1

The professor asks you to analyze this data for her.

- What is the null hypothesis here?
- What is the alternative hypothesis?
- What kind of analysis would you run to test statistical significance?
- Analyze the data. Is there a statistically significant relationship between the teaching technique and student grades?
- How would you report on the results of the analysis to the professor? Did her new teaching technique work? Is it “statistically significant”?

Fuel Economy

- The EPA is running an audit on car manufacturer Volkswagen to assess their claims about the fuel economy of some of their models. The EPA obtained a sample of 16 vehicles that yielded the following miles per gallon (mpg). The values are as follows:

18.0 15.7 15.8 18.0 18.5 19.8 20.2 20.4 16.9 18.3 19.8 17.2 16.7 17.7 19.5 18.0

Provide a confidence interval for city miles per gallon (mpg) for this model.

Alexis and the Missing P-Values

- Alexis is a friend from grad school who, in order to save money, acquired a pirated version of Stata to do their homework. Unfortunately, the software came with a glitch where it sometimes didn’t show the p-values for a given statistical test. Alexis is doing research on the different levels of noise experienced by people attending fitness classes. They collected data on noise levels in both high-intensity and low-intensity fitness classes across different commercial gyms in Central New York. The hypothesis Alexis is trying to test is whether high-intensity fitness classes have different levels of noise than low-intensity classes. Below is an output produced by Alexis’s shoddy version of Stata.

```
. ttest decibels, by (intensity) unequal
```

```
Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
High	28	92.65714	.5250455	2.778279	91.57984	93.73445
Low	14	86.81429	.7621438	2.851681	85.16778	88.4608
Combined	42	90.70952	.60618	3.928495	89.48532	91.93373
diff		5.842856	.9254923		3.938669	7.747043

```
diff = mean(High) - mean(Low)                                t = 6.3132
H0: diff = 0                                                  Satterthwaite's degrees of freedom = 25.5018

Ha: diff < 0          Ha: diff != 0          Ha: diff > 0
```

As you can see, the table of results does not showcase the p-value for the different hypothesis tests. Based on what you know about how t-tests work and about how to read Stata outputs, what would you tell Alexis? Would you reject the null hypothesis that the two groups have the same means? Indicate what evidence, from what appears in this output, you'd use to justify your decision.

Coffee and Stress

6. A graduate student is interested in the relationship between the number of cups of coffee drunk by students during finals week and their stress level. Specifically, the graduate student thinks that more cups of coffee will lead to higher stress levels. Presume that coffee is measured by "cups per day" and stress is measured by a self-reported 10 point scale where 10 indicates "high stress" and 0 indicates "low stress", and that the graduate student surveys 100 students.
 - a. If the graduate student is analyzing this relationship with regression, what regression equation would they be estimating?
 - b. Let's say the graduate student collected data, and returned the following information:

	Coefficient	Standard Error	T-Ratio
Cups of Coffee	.5	.2	
Intercept	5	1	

Calculate the T-statistic and fill it in the T-Ratio column.

- c. Interpret the slope and intercept coefficients. What do they mean?
- d. Let's say the model sum of squares for this regression was 200 and the residual sum of squares was 300. What is the R^2 ? Calculate and interpret this number.
- e. Is there a statistically significant relationship between Cups of Coffee and stress levels, according to this regression? What does it mean to be "statistically significant" in this context?
- f. What is the 95% confidence interval for "Cups of Coffee"? (You may round) What does this mean?
- g. What is the expected stress level for a student who drinks 1 cup of coffee per day? 3 cups of coffee per day? 10 cups of coffee per day?

- h. Let's say the graduate student added another variable to the regression - hours of sleep per day. The results of the new regression may now be found below.

	Coefficient	Standard Error	T-Ratio
Cups of Coffee	.1	.2	
Hours of Sleep	-.4	.1	
Intercept	8	1	

Calculate T-Ratios for the graduate student again.

- i. Interpret the slope and intercept coefficients. What do they mean?
- j. What happened to the coefficient for "cups of coffee" between the first and second regression? Why? What does this mean?
- k. Let's say the total sum of squares for this regression was 500 and the residual sum of squares was 200. What is the R^2 ? Calculate and interpret this number.
- l. Is there a statistically significant relationship between Cups of Coffee and stress levels, according to this regression? What about between sleep and stress? What does it mean to be "statistically significant" in this context?
- m. What is the 95% confidence interval for the coefficient for "Cups of Coffee" and "Hours of Sleep"? (You may round). What does this mean?
- n. What is the expected stress level for a student who drinks 1 cup of coffee per day and gets 8 hours of sleep? 3 cups of coffee per day and 4 hours of sleep? 10 cups of coffee per day and 8 hours of sleep?
- o. Considering the information in both regressions, what would you conclude about the relationship between stress, coffee, and sleep?