Practice Quiz 1
G572
Spring 2023

The quiz will be 10 questions (numerical, true/false, multiple choice). I recommend when taking the practice quiz, also coming up with a "why?" answer where relevant, even though those kinds of questions won't be asked on the quiz.

Practice questions:

1. Experimental data will never include information gathered during interactions with people in their everyday lives. (T/F)
   False – firms regularly conduct randomized, controlled trials through A/B testing versions of their websites, sending out discounts (if randomized, not the targeted ones), etc. – all while interacting with consumers who are going about their everyday lives.

2. Given the following output, how many standard errors away from a (correlational) effect of 0.0 is the (correlational) effect of Age on SickTime (holding education and tenure fixed)? (round to the nearest one-tenth)

```
. reg SickTime Education Age TenurewithFirm
```

Source	SS	df	MS	Number of obs	=	8,093
				F(3, 8089)	=	3323.73
Model	82746.9846	3	27582.3282	Prob > F	=	0.0000
Residual	67127.3728	8,089	8.29859967	R-squared	=	0.5521
				Adj R-squared	=	0.5519
Total	149874.357	8,092	18.5212997	Root MSE	=	2.8807

SickTime	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Education	.0013832	.0159338	0.09	0.931	-.0298512	.0326176
Age	.1501507	.0029555	50.80	0.000	.1443571	.1559442
TenurewithFirm	.2992801	.0061051	49.02	0.000	.2873126	.3112476
_cons	7.198536	.2715687	26.51	0.000	6.666192	7.730881

50.8 (the number of standard errors above or below the null hypothesis is the t statistic, by definition)

3. Given the following output, at a 95% confidence level, would you reject the null hypothesis for the coefficient on Education?

```
. reg SickTime Education Age TenurewithFirm
```

Source	SS	df	MS	Number of obs	=	8,093
				F(3, 8089)	=	3323.73
Model	82746.9846	3	27582.3282	Prob > F	=	0.0000
Residual	67127.3728	8,089	8.29859967	R-squared	=	0.5521
				Adj R-squared	=	0.5519
Total	149874.357	8,092	18.5212997	Root MSE	=	2.8807

SickTime	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Education	.0013832	.0159338	0.09	0.931	-.0298512	.0326176
Age	.1501507	.0029555	50.80	0.000	.1443571	.1559442
TenurewithFirm	.2992801	.0061051	49.02	0.000	.2873126	.3112476
_cons	7.198536	.2715687	26.51	0.000	6.666192	7.730881

No, if using a 95% confidence level, we would only reject if the p-value was less than 0.05.

4. A random sample ensures our estimate of the causal impacts of X1 on Y will be unbiased.
    a. True
    b. False

b. It only ensures we can extrapolate whatever results we get from the sample out to the population. It does nothing to prevent confounding factors.

5. The standard error estimates the standard deviation of which of the following:
    a. The population
    b. The sampling distribution
    c. Both of the above
    d. None of the above

The sample standard deviation estimates the standard deviation of the population. The standard error's actual name is the **estimate of the standard deviation of the sampling distribution** (which is long, so that's likely why it's called standard error).

6. For a given population (with variance > 0) of hours salespeople spend making calls per day, compared to a sampling distribution for sample means of sample size 25, a sampling distribution of sample size 125 will definitely:
    a. Have the same standard deviation
    b. Have a lower standard deviation
    c. Have a higher standard deviation

The sampling distribution is narrower/tighter for a larger sample size so its standard deviation is clearly lower – note we don't know what the standard deviation is, we only estimate it (with the standard error).

7. If we reject the null hypothesis (let's say using a 99% confidence level) that the true population mean of daily profits for companies in Indiana is $10,000, it means we have proven that $10,000 is not the true population mean. (T/F)

    No, there's still a chance (albeit small) that the null hypothesis is actually true but we randomly sampled a really extreme sample. This is clearly unlikely, but for this reason we don't say we've "proven" the null to be false. We just reject it.

8. An R-squared above 0.9 implies that coefficient estimates can be interpreted as causal effects. (T/F)

    False – R-squared just tells us how Y moves with the X variable(s) and doesn't provide us with any information about causality.

9. Suppose Kroger offers a discount to some customers and not to others. If they wanted to conduct a hypothesis test on the outcomes for these two groups (we can do other types of analysis), which type of hypothesis test would they want to conduct?
    a. Single population hypothesis test
    b. Difference in means hypothesis test

Since there are two different groups and we are interested in how the treatment may have impacted one of them, we cannot use the single sample/pop'n t test. We could use the two sample t test (aka the difference in means test).

10. What is typically the null hypothesis for a two sample t test? (numerical answer, but definitely consider the "why?" here)

Zero – most commonly we are asking if the two populations (sometimes one with the treatment and one without) are equal, which means we're asking if their difference is 0.