

PS7_{spohn}

adam spohn

March 2024

1 Wages information

Logwage is missing at a rate of 25.12

I ran a t test on the relationship between whether a value in logwage was missing, and the values of the other variables. This relationship was found to be significant for hgc, college, and tenure. Becuase this suggests a systematic difference between logwage being different or not, I beleive that they are missing not at random.

	(1)	(2)	(3)	(4)
(Intercept)	0.761***	0.965***	0.761***	0.704***
	(0.130)	(0.102)	(0.098)	(0.112)
hgc	0.062***	0.049***	0.062***	0.066***
	(0.005)	(0.004)	(0.004)	(0.005)
college	-0.146***	-0.160***	-0.146***	-0.148***
	(0.035)	(0.026)	(0.025)	(0.028)
tenure	0.023***	0.015***	0.023***	0.022***
	(0.002)	(0.001)	(0.001)	(0.001)

Figure 1: Regression Table for PS7

True value of B1 is 0.093.

All four of these were consistently lower than this. I would not say they are wildly inaccurate, but not as accurate as one might hope. The third method

had the same prediction as the first one. Perhaps this is because it used the coefficients from the list wise method. The mice method is closer to the real value, so I thought that this method may be more accurate. However, I noticed when ran it multiple times it got different results, some of which were farther than the true value than the third method, so maybe it is not always the better method.

2 Final Project

I have been a bit unsure about my current topic. I have been having issues finding good data that would accurately represent chat.gpt usage, and feel like would then be difficult to relate that to employment changes or perhaps productivity differences, just because there can be so many confounding factors.

I have planned to change my topic. My plan is to relate carbon emission and or average yearly temperatures to tornado numbers and locations. I often hear that tornado alley has shifted, so the idea is to see if global warming and or carbon emissions has played a role in this shift. I have found data sets that mark tornado locations on a map, though I do not have experience with geographical data like that. Instead I am planning to use a different data set I found that shows how many each state had per year, then can look at, for example, is the amount Oklahoma has as a percentage of the US increasing or decreasing. I have found a few data sets that show emissions and temperatures, I am still working on which one I want to use exactly and how I want to approach what I want from the data exactly. I am not sure what kind of model may want to use yet, perhaps a fixed effect or random effects model.