ECON 419 Final Exam

Problem One

1. Construct a table that shows the sample mean, standard deviation, and minimum and maximum for birth weights of infants whose mother does not smoke and for infants whose mother does smoke. Make sure that your table is readable and each basic statistic clearly labeled.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Non-Smoker** |  |  |  |  |
|  | **Sample Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| **Birthweight** | 3432.06 Grams | 584.62 Grams | 425.0 Grams | 5755.0 Grams |
| **Smoker** |  |  |  |  |
|  | **Sample Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| **Birthweight** | 3178.83 Grams | 580.01 Grams | 510.0 Grams | 4763.0 Grams |

1. Discuss:

For mothers who do not smoke during pregnancy, the average birthweight of their babies was 3432.06 grams while the average birthweight for women who did smoke was 3178.83 grams. So this difference suggests that, on average, a baby born to a non-smoking mother weights 253.23 grams more and a baby born to a mother who smokes. The standard deviation for both sets of observations are very similar, suggesting variation is more or less the same as well. The range of the observations within the two sets is interesting though. Non-smoking mothers gave birth to not only the heaviest baby but also lightest baby, however this may be due to the larger amount of observations found in the subset of non-smoking mothers compared to the subset of mothers who smoked.

1. A causal relationship cannot be inferred due to possible omitted variable bias. Alcohol usage is most likely correlated with smoking, so it should also be included in the model, in addition to other variables like age and prenatal doctor visits.
2. The number and schedule of prenatal visits would be a good indicator of the baby’s overall health. By doing so, the mother is making sure any that factors contributing to low birthweight can be discussed by her and her gynecologist early and over the course of the entire pregnancy.
3. Looking at the residual plots suggests there is homoscedasticity from the variables alcohol and tripre3. There appears to be a larger variety of residuals for those two variables compared to the others
4. Breusch-Pagan –

Harvey-Godfrey –

White –

1. FGLS – smoking, alcohol, tripre1,2,3
2. FGLS – smoker, alcohol, nprevisit, unmarried

CP, for an unmarried mother, the birthweight of her baby will be, on average, 181.39010 grams less than the baby of a married mother.

Is Unmarried significant at the 95% level?

1. Is being unmarried economically significant?

The coefficient is certainly statistically significant almost equal to the mother being a smoker and the number of prenatal visits. Interesting enough, alcohol usage wasn’t statistically significant at all. Being married may also imply the mother is more financially stable; with the possibility of two incomes. So it would be incorrect to suggest being married is itself is economically significant, and a researcher should also collect the total household income of the mother.

1. Findings

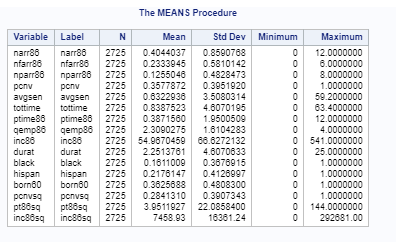
I would not agree with that article. It’s presumptive. The relationship of being married to a higher birthweight does not imply it’ll also cause the baby to healthier, and this variable should be given as much weight as the behavioral traits of not smoking and following prenatal visits guidelines. Also, like I mentioned earlier, income may be a better estimator for average health than whether or not the mother is married.

1. Other states?

If only for 1989, then the analysis should be valid for other states as well. It wasn’t until the late 90s and early 2000s, that smoking bans became widespread in cities and the workplace. Its common knowledge that the smoking rate has drastically decreased in 26 years. So although being a smoker is a very significant predictor for birthweight, other variables should be tested as well to account for this drop in the smoking population. In terms of applying the results to other countries, the physical behavior of smoking should have the same effect on birthweight, however other predictors such as the country’s GDP could be just as significant

Problem Two (Deterrence)

1. Table of Summary Statistics



1. A lot of information can be gleaned from these basic statistics. The variables narr86, nfarr86, and nparr86 all deal with not only the number of arrests but also the nature of the crime, with felonies being more serious and property crimes being non-violent. Comparing the averages of these three variables to their maximums, it appears a very small amount of the population is committing the majority of crimes. Also the variables relating to prison sentences (avgsen and tottime) seem to share similar lengths, suggesting a low rate of recidivism within this sample. Time employed and unemployed shared similar averages as well, which I found interesting. It seems like almost half of those surveyed had experienced a quarter of being unemployed during 1986. That’s difficult to believe though. Also the demographic sample appears to show the percent on black respondents at 16% and Hispanic respondents at 21%. This may be due to the survey being conducted in California. Also I was surprised to see the high number of respondents who were specifically born in 1960, so 26 year olds made up 36% of those surveyed. That may explain with average sentencing and total time served share similar summary statistics.
2. Estimated effect on arr86 if pcnv goes from 0.25 to 0.75

Model:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Intercept | Pcnv | Avgsen | Tottime | Ptime86 | Inc86 | Black | Hispan | Born60 |
| .36098 | -.15438 | .00350 | -.00206 | -.02160 | -.00122 | .16172 | .08926 | .00287 |
| (.01609) | (.02093) | (.00634) | (.00489) | (.00447) | (.00013) | (.02350) | (.02056) | (.01720) |
| [.01668] | [.01893] | [.00588] | [.00422] | [.00275] | [.00012] | [.02549] | [.02103] | [.01713] |

1. Use GLS

y\_hat min: -0.4112080 y\_hat max: 0.5665108, must correct for values less than 0

Model:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Intercept | Pcnv | Avgsen | Tottime | Ptime86 | Inc86 | Black | Hispan | Born60 |
| 0.18052 | - .03589 | -.00286 | .00010 | -.01216 | -.00013 | .28376 | -.00112 | -.08012 |
| .02647 | .01675 | .01072 | .00444 | .00206 | .00006826 | .08920 | .02714 | .01006 |
| .10421 | .08419 | .00703 | .00240 | .00550 | .00024718 | .06984 | .04502 | .06890 |

Its appears the robust standard errors have increased in the new model for the intercept and variables pcnv, avgsen, ptime86, inc86, black, hispan, and born60. Only tottime saw a decrease in the standard error. It terms of variables significance, pcnv, ptime86, inc86, black, and hispan all had pvalues below 0.0001. In the GLS model only ptime and black are significant above the 95% level.

1. Probit Model

1. Joint Significance
2. Likelihood test

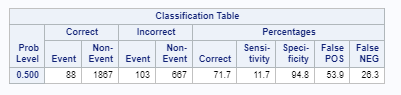
We must run a likelihood test because we can no longer interpret the coefficients as marginal effects. So, we are unable to use an F-test to compute joint hypothesis, since there is no error term.

1. Logit Model

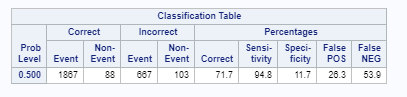
1. Percent Correctly Predicted

This question was very difficult and I had to do a bit of searching. I guess essentially, it is asking for a confusion matrix. This is the best I could to,

For arr86=1:



For arr86=0



Both tables are basically the same.