BT2101 Mid-Term Exam

# Problem 1:

# A person has two children. If one of them is a son, what is the probability that the other is also a son? (Assume that there is 50% chance of having a male and a 50% chance of having a female)

# A person has two children. If the first-born child is a son, what is the probability that the other child born later is also a son? (Assume that there is 50% chance of having a male and a 50% chance of having a female)

# Problem 2:

|  |
| --- |
| We want to test hypothesis H0: μ = 68 vs H1: μ ≠ 68, where the μ is the average weight of male students in the gym. We reject the null if the sample mean is the 67 kg or greater than 69 kg. The sample size is 36, and the population standard deviation is known to be 3.6kg. Assume that population follows the normal distribution. Probabilities for Z with the Z-Table can be found below. In this situation, what is the probability of type 1 error? (note: z = (x - μ) / σ. Here, σ = 3.6 / sqrt(n)[[1]](#footnote-1)) |

# 

# Problem 3:

Please read use the DRUG dataset to answer this question. This is a US county-level dataset. Please carefully read the variable description in Table 4.1. The dataset contains county code–level data on opioid overdose death and the number of electronic prescriptions, along with characteristics of the county population, in the US. The idea is to see whether the number of electronic prescriptions has a negative association with opioid overdose death. Assume that this is a perfectly random sample. Assume that you use opioid overdose death as a dependent variable and the number of electronic prescriptions as the independent variable and use all other variables as control.

|  |  |
| --- | --- |
| **Table 4.1 Variable Description** | |
| **Variable** | **Descriptions** |
| Electronic prescriptions | Number of electronic prescriptions in a county |
| Opioid Overdose | Rate of opioid-involved overdose death in a county  The rate of counties with fewer than ten opioid overdose deaths is reported as 0 for privacy protection. |
| Population | Number of individuals in a county |
| Male | Number of males in a county |
| White | Number of non-Hispanic Caucasians in a county |
| Poverty | Number of individuals below the poverty line in a county |
| Income | Median household income in a county |
| Hospitals | Number of hospitals in a county |
| Pharmacies | Number of pharmacies in a county |

3-1) Structure the econometric model to support the idea above

3-2) Use Python and conduct analysis, interpret the coefficient of number of electronic prescriptions

(Note: use commend “smf.ols” to run the model, pre-install package “numpy” “pandas” “matplotlib” “statsmodels.formula.api”,“statsmodels.api”)

3-3) Describe one violation of OLS assumptions when you are using DRUG dataset, describe the criteria of violations and describe the detailed reason why DRUG dataset suffers from such violations.

3-4) Describe two sources of endogeneity when you are using DRUG dataset. Describe the criteria of sources of endogeneity and describe the detailed reason why DRUG dataset suffers from such issue.

# Problem 4:

To measure the association of AI robot adoption on sales improvement in the physical store, the company did lab experiments. Specifically, the company recruited 500 users into the lab. First, they provide a virtual situation that AI robots adopted in the shopping mall, and let users interact with and use AI robot. After that, the company measures the expected spending of participants and AI robot use time. Using the AI user time and expected spending, the company conducted the experiment. Table 1 is the result of the lab experiment.

4-1) Interpret the coefficient of 5.563 in Table 1.

4-2) In the *perspective of validity*, what is the (1) drawback and (2) benefit of adopting the result of table 4.1 (describe only one drawback and only one benefit).

|  |  |
| --- | --- |
| **Table 4.1. Impact of AI robot adoption in the shopping mall** | |
| Dependent variable: | **Expected Spending**  **(by dollar)** |
|  | **(1)** |
| AI use time  (by minutes) | 5.563\*\*\*  (1.611) |
| Observations | 500 |
| R-squared | 0.506 |
| Notes: Robust standard errors are in parentheses.  \* p<0.1, \*\* p<0.05, \*\*\* p<0.01 | |

* Submission: LumiNUS files > Mid Term Submission (same with the assignment)
* After a confirmation that you have successfully submitted your exam paper (through Luminus File), you should chat to TA (via zoom chat function) and send the screenshot of mid-term paper submission time and get confirmation. 1. TA will mark submission time and then allow you to leave, and then you can leave.
* Submit Both Solution and Coded Files
* Solution File Name: Solution\_Name (Name means your name)
* Coded File Name: Coding\_Name
* Solution File Format: .doc/.pdf -Coded File Name: .ipynb/.py

1. σ in the z-score is not the population standard deviation, but it is a standard deviation of our estimator (in this case an average). The sqrt(n) in the denominator of σ is from the central limit theorem result which says the sample average converges to the population mean with the speed of sqrt(n). [↑](#footnote-ref-1)