**Qns 1-1)**

The probability is 50%

**Qns 1-2)**

The probability is 50%. The variables are independently and identically distributed.

**Qns 2)**

This is a two-tailed test. The probability of a type 1 error has a 4.75% chance of occurring from the Z-test being equal to -1.67 if the sample mean is 67kg.

**Qns 3-1)**

The model is as follows:

Opioid Overdoses = X0 - X1 (Electronic Prescriptions)

**Qns 3-2)**

The coefficient of the number of electronic prescriptions is 0.0005, where X1 = 0.0005

**Qns 3-3)**

One clear violation of the assumptions would be exogeneity and endogeneity issues, where there are omitted independent variables that affect both the number of opioid overdoses and electronic prescriptions, resulting in the violation of exogeneity as there is an endogeneity issue where there are omitted variables that could make the OLS model more accurate and efficient.

This is because the DRUG dataset and the OLS model did not collect enough data that could affect the number of electronic prescriptions. For example, the DRUG dataset fails to account for other factors like number of firms in the country which can produce these electronic prescriptions as a higher number of firms producing these prescriptions could have a positive association with the number of prescriptions taken and a negative association with the number of opioid overdoses.

**Qns 3-4)**

One such endogeneity issues would Income and Number of Pharmacies in terms of omitted variable bias and measurement error. The criterion for endogeneity is that the amount of income could affect the number of electronic prescriptions, as a higher income may result in the ability to afford these electronic prescriptions and thus resulting in a positive association between electronic prescriptions and income. Moreover, what is considered as a pharmacy may vary between country to country depending on the country’s law, resulting in an in accurate number of pharmacies collected for each country.

The DRUG dataset and the OLS model suffers from these issues as the model suffers from an omitted variable bias where the model fails to account for these control factors that could affect the validity and accuracy of the OLS model and measurement errors where the data collected is not valid and inaccurate leading to an inaccurate model.

**Qns 4-1)**

The coefficient of 5.563 indicates that there is a strong positive association between expected shopping and AI use time subject to a significance value of 1%.

**Qns 4-2)**

One drawback would be that there could be errors in variables bias in the dataset as seen from the table. For example, during the collection of how much time everyone uses the AI bot, there could be scenario where when a user doesn’t use the AI bot at all, the company might put the value of 0 instead of the average, resulting in the data being skewed towards one area. Furthermore, the company did not consider ambiguity like the level of income of these shoppers, which could affect the expected spending. This means that the collection of the data could be invalid as very little thought could be placed towards omitted variables or false answers.

One benefit would be that there is very little occurrence of reverse causality, where the dependent variable affects the independent variable, and the independent variable affects the dependent variable resulting in a circular dependency. This is not the case in this result, as expected spending does not affect AI use time since the users where first told to interact with the robot before going to the shopping mall to spend, meaning to say AI usage time cannot be influenced by spending amount as the users could not interact with the AI before spending. Thus, with lower chance of reverse causality, the model is valid.