Sejuti Mannan
Data Science Bootcamp Week 7 Homework

1) How do you assess the statistical significance of an insight?
To tell if an insight is statistically significant, we use statistical hypothesis testing. Statistical significance is the likelihood that a result is not due to random chance. To do statistical hypothesis testing, you have to find the p-value, which is the probability of an observation occurring by chance given the null hypothesis, and compare it to the significance level. If the p value is greater than the significance level, then we fail to reject the null hypothesis. If the p-value is less than or equal to the significance level, you reject the null hypothesis.

What is the Central Limit Theorem? Explain it. Why is it important?
The Central Limit Theorem says that if sampling is done randomly and independently, the sample means follow a normal distribution as the sample size increases. This occurs regardless of how the underlying population is distributed. The CLT is important because it tells us that the distribution of our given data is not important—we can just focus on the sample means for our statistical analysis like hypothesis testing.

What is the statistical power?
Statistical power is the likelihood of a significance test detecting an effect when there actually is one. High power indicates a large chance of a test detecting a true effect. Low power means that your test only has a small chance of detecting a true effect or that the results are likely distorted by random error.

How do you control for biases?
To control for biases, you must take extra precautions when designing your experiment. For example, you can use randomization, blind the researcher/subjects, and try to make all other variables besides the independent variable constant.

What are confounding variables?
Confounding variables are extraneous factors that influence the independent variable and dependent variable.

What is A/B testing?
A/B testing is a method to compare two versions of the same product to decide which one is more effective.

What are confidence intervals?

A confidence interval gives a range of plausible values for the population parameter. For example, a 95% confidence interval means that in 95 out of 100 repeated samples, the interval will contain the true population parameter.