Risk and Decision-Making for Data Science and AI

Week 5 Lab 4 Exercise

Question 1:

Do you agree with the below interpretations of p-values and confidence interval and why.

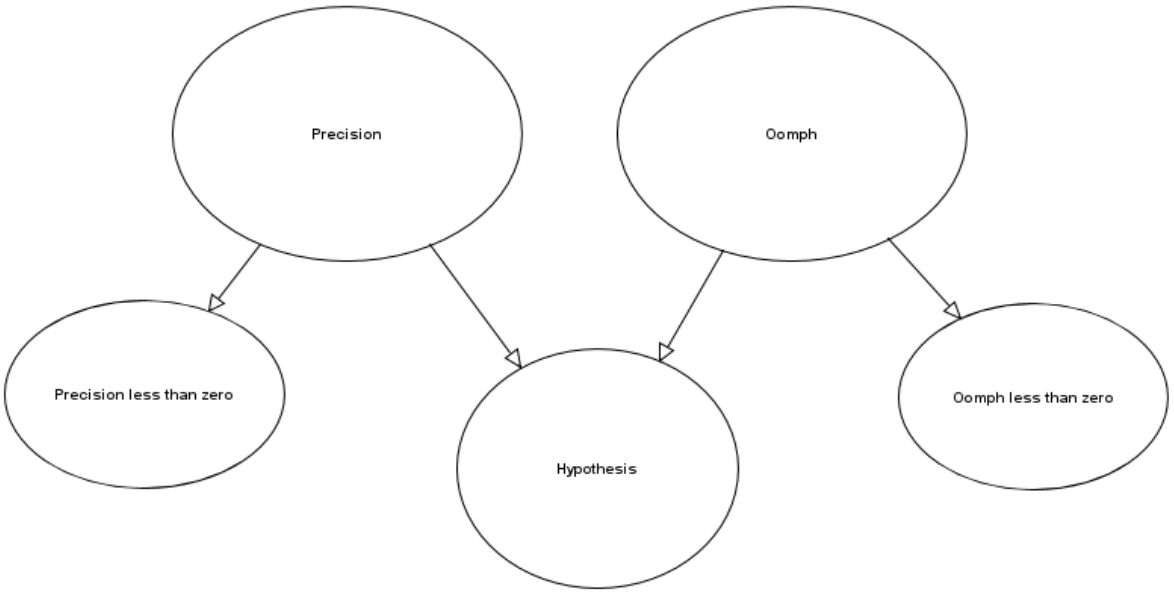
a) "The 95% confidence interval for the average height of students in the school is 150 cm to 160 cm. Therefore, we can be 95% certain that the average height falls within this range."
b) "We computed a 99% confidence interval for the mean score on a standardized test to be [75, 80]. This means that if we were to take repeated samples and calculate confidence intervals from them, we would expect approximately 99% of those intervals to contain the true mean score."
c) "We reject the null hypothesis because the p-value (0.001) is less than the significance level (0.05), indicating strong evidence against the null hypothesis."
d) "Our p-value is greater than 0.05, so we fail to reject the null hypothesis and conclude that there is no effect."

Question 2:

We are interested in comparing two weight-loss drugs to determine which should be sold. Experimental data on the drugs Oomph and Precision have shown that, over a one-month period for a number of test subjects:

- Weight loos for Precision is approximately Normally distributed with mean 5 pounds and variance 1.
- Oomph showed a much higher average wight loss with much more variation and is variation and is estimated as Normal(20,100) distribution.

Create the below BN in a software (agena.ai or python) to answer the question



Hypothesis: if(Precision>Oomph,"Precision","Oomph")

Precision less than zero: if(Precision<0.0,"True","False") (similar for oomph less than zero)

Risk and Decision-Making for Data Science and AI

Week 5 Lab 4 Exercise

Question 3:

Identify the type of missing data in the following scenarios:

a) A researcher collects survey data on a tablet, and due to a technical glitch, some responses are missing.
b) In a medical study, participants with higher incomes may be less likely to disclose their alcohol consumption.
c) An educational researcher is studying the relationship between socioeconomic status (SES) and academic performance. In a school survey, some students choose not to disclose their family income.
d) In a mental health survey, individuals with severe depression may be less likely to provide accurate information about their symptoms.
e) A researcher collects temperature data from weather stations across a region. Due to a random power outage, data from some weather stations are missing.