1   Q01

A study is conducted that collects data on people who do and do not eat a plant-based diet(Study here). Researchers' collected demographic data on participants and data on two outcomes: if the participant developed cardiovascular disease and if the patient experienced a heart attack within 15 years of being followed.

Let's assume the study followed 100 participants and that 50 particpants followed a plant-based diet and 50 participants did not follow a plant-based diet. Further assume researchers collected the participant: age at the beginning of the study, whether the particpant has a family history of heart disease, whether the patient developed cardiovascular disease, and whether the participant experience a heart attack.

Among those who followed a plant-based diet 20 developed cardiovascular disease and 10 experience a heart attack, and among those who did not follow a plant-based diet 40 developed cardiovascular disease and 15 experienced a heart attack (Note: I made these numbers up.).

A. Define an outcome in this experiment

B. What is the sample space?

C. Describe an event in this sample space

D.Define an outcome in this experiment

E. Define a Binomial random variable called X that describes the number of heart attacks among those participants who followed a plant-based diet. Use as parameters N_{plant} as the number of trials and θ_{plant} as the probability of success.

F. What are the potential values of X?

G. Assuming $\theta_{\text{plant}} = 0.20$, what is expected value of X? what is the variance of X? what is the standard deviation of X?

H. Assuming $\theta_{\text{plant}} = 0.30$, what is the probability that $X = 10$?

2   Q02

Suppose the random variable X has expected value $E(X) = 10$ and variance $(\text{Var}(X) = 2)$, and the random variable Y has expected value $E(Y) = -5$ and varinace $(\text{Var}(Y) = 4)$.

A. Compute E(X+Y)

B. Compute E(X-Y)

C. Compute E(2X+Y)

D. Compute Var(X+Y)

E. Compute Var(X-Y)

3   Q03

Assume there exists a continuous random variable B that can take any of the values in the interval from -2 to 2.

A. What are the inputs to B's cumulative density function (F)?

B. What are the outputs to B's cumulative density function (F)?

C. Can the outputs of F be greater than 1 or less than 0 ? Why or Why not?

D. Can the outputs of B's probability density function be greater than 1 or less than 0? Why or Why not?

4   Q04

You're given the following dataset

Data
2.3
-8.0
0.3
78
-10.78
2

A. Standardize each data point

B. Assume a random variable G generated the above data. Use the central limit theorem to approximate a distribution for $\bar{G}$

C. Compute a 95% confidence interval for the mean (μ) of G

5 Q05

Suppose you study a random variable H with a Bernoulli(θ) distribution. From H you generate 10 values and compute the sample mean, then 100 values and compute the mean, then 1000 values and compute the sample mean. As you collect larger samples, what value do you expect the sample mean to approach, and why?