Review of Continuous outcomes and recap of Part II

11 April 2025

MT 2 review

Question 5

In a cardiovascular health program, the cut-off for healthy cholesterol for Test A is set to 240 mg/dL and for Test B at 200 mg/dL. Any values above the threshold are diagnosed with poor cardiovascular health, or positive for disease. Jessica wants to calculate the probability that a person will have unhealthy cholesterol levels according to Test B. Choose the appropriate theoretical distribution. **

Distributions

Distribution	Defined by:	Type of outcome	R notation
Normal	Mean and SD	Continuous	norm
Binomial	number and p	Binary (success or failure in n trials)	binom
Poisson	mean (λ)	Discrete count of events in an interval	pois

Which distribution?

Question 7

Your research team is working on developing a new diagnostic test for bird flu. After running trials with the new test, you calculate the following:

- Sensitivity: 0.85
- Specificity: 0.80
- Prevalence: 0.10

Conditional probabilities of Screening tests

Two characteristics that are conditional on true disease status

- Sensitivity = P(Test positive | Disease)
- Specificity = P(Test negative | No Disease)

Two characteristics that are conditional on test result

- Predictive value positive = P(Disease | Test positive)
- Predictive value negative = P(Not disease | Test negative)

Conditional probabilities of Screening tests

What happens to sensitivity if we are in a context where the disease is more prevalent?

What happens to predictive value positive?

*7.1 [2 points] using Bayes' theorem. Show each step, referencing the sensitivity, specificity, and prevalence. Present your answer as a percentage, and round to two decimal places. **

Question 7

Calculate

$$P(\text{Has virus} \mid \text{Test positive}) = P(V|T)$$

Bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Question 7

Bayes' theorem

$$P(V|T) = \frac{P(T|V)P(V)}{P(T|V)P(V) + P(T|V^c)P(V^c)}$$

Sensitivity=P(T|V)=.85

Specificity=$P(T^c|V^c) = .80$, $P(T^c|V^c) + P(T|V^c) = 1$

Prevalence = P(V) = .10, 1-Prevalence= $P(V^c)$

table method

Test result	Samples with known Disease	Samples without Disease	Totals
Positive			
Negative			
Totals			

- Sensitivity: 0.85
- Specificity: 0.80
- Prevalence: 0.10

Question 10

A clinical research team is investigating the recovery rate of knee replacement surgery among patients at different hospitals. Based on past data, they estimate that approximately 63% of patients who follow their prescribed physical therapy routine after undergoing the surgery will recover within 6 months. Assume that the likelihood of recovery among patients is the same for all hospitals and that the hospital recovery rates are independent of one another.

10.1 [1 point] If the team of researchers is hoping to capture 95% of the data within two standard deviations of the mean adherence, what is the smallest number of hospitals they must randomly select?

Normal approximation for binomial distributions

Suppose that a count X has the binomial distribution with n observations and success probability p. When n is large, the distribution of X is approximately Normal. That is,

$$X \dot{\sim} N(\mu = np, \sigma = \sqrt{np(1-p)})$$

As a general rule, we will use the Normal approximation when n is so large that $np \geq 10$ and $n(1-p) \geq 10$.

It is most accurate for p close to 0.5, and least accurate for p closer to 0 or 1.

Question 11

Abstract: Cognitive functioning has been related to nutritional and physical factors. Your team decides to focus on iron status in young-adult female students given the high rate of iron deficiency reported in this population. You want to assess whether low iron status is associated with a lower grade point average (GPA) than normal iron status among healthy college-aged women.

Methods: You recruit female-identifying students aged 18–35 and record their GPA at the time of measuring their ferritin (iron) level. You move forward with 200 eligible participants, and the average GPAs were compared between the group with Low ($F \leq 12$) iron status ($n = 100$) and the group with Normal ($F \geq 20$) iron status ($n = 100$).

Results: The sample group of students with normal ferritin levels had an average recorded GPA of 3.66 with $\sigma = 0.08$. The sample group of women with low ferritin had an average recorded GPA of 3.39 with $\sigma = 0.09$.

What are the null and alternative hypotheses?

Continuous outcome tests

How do we choose the test to use?

- How many samples/groups?
- Do we know the standard deviation in the population?
- Are the outcomes independent of one another?
- Do we meet the assumptions of the test?

Coffee and race speed

I am interested in testing whether drinking coffee 20 minutes prior to a race increases sprinting speed. I recruit 11 runners from a running club and have them sprint 200 meters. I then give them each a cup of coffee and 20 minutes later ask them to sprint 200 meters again.

Earnings and group

"It is often cited that for the same job position, male employees tend to make more than female employees. Sociologists investigated this, and collected data on yearly salaries from 25 large companies in the tech industry. They collected 2 different samples – a random sample of the yearly salaries of male employees and a random sample of the yearly salaries of female employees matched for the same job position."

Biomarker

Defne is working for a research group that wants to compare levels of a potential biomarker for chronic traumatic encephalopathy between a group of former NFL players and a group of controls. The biomarker ranges continuously from 0.70 to 1.60. They are able to recruit 40 former NFL players, and 80 controls.

viral load

I am following a cohort of infants from birth to one year. All infants are breastfed and exposed to CMV through breast milk. Some of them have transient infections, some have sustained infections with viral shedding and some are uninfected by the end of the year. I want to compare the viral load in the breast milk of their mothers.

Math scores

I suspect that in my county the scores for 8th grade math are lagging behind national scores. I conduct a representative sample of all the 8th graders in my county. I have scores for standardized math tests for 300 students.