Week 6

Question 1:

1. **The Neyman-Pearson Paradigm**

   a) Describe the Neyman-Pearson paradigm for deciding between a null and alternative hypothesis. Define Type I and Type II errors and make a courtroom analogy (or any other analogy that seems relevant - e.g., replay challenges in sports) to point out the asymmetry in the two hypotheses.

Neyman-Pearson Paradigm (Rice, 331)

Formulated the theory of hypothesis testing by casting it as a decision problem and making the probabilities of the two errors central, thus not needing to specify prior probabilities.

One hypothesis became the null (H_0) and the other the alternative hypothesis (H_A)

Want to maximize the probability of correctly rejecting the null.

Null is set as if there is no change or difference.

Alternative can see change in some direction.

Errors

Type 1 error: Rejecting H_0 (null) when H_0 is true (false positive)

Type 2 error: Accepting the null hypothesis when it is false (β) (false negative)

Court Room Analogy

Null: Not guilty

Alternative: Guilty

Type 1: convicting the defendant, they are actually innocent

Type 2: fail to convict a guilty defendent.

Asymmetry:

In real life convicting an innocent person is often seen as worse than letting a guilty person go. Normally, in law, you are innocent until proven guilty as to avoid wrongfully convicting an innocent person.

Sports: the call on the field is the one you go with unless you have enough info to overturn

Call is confirmed vs call on the field stands vs call overturned

b) Imagine testing the defect rate θ of a manufactured component. The stated rate is $\theta_0 = 0.1$ and you are concerned that it might be higher than this (there's no problem if it is lower than 0.1). Suppose we take a sample of $n = 400$ and sound an alarm if 50 or more of the sampled components are defective. What are the hypotheses and significance level? Report the power of this test if $\theta = 0.11$, $\theta = 0.15$ or $\theta = 0.09$. How would things change if your threshold were 51 or more defectives? Discuss the connections between significance level and power.

θ = defect rate $n = 400$

$\theta_0 = 0.1$ = stated defect rate

Hypotheses:

$H_0: \theta_0 = 0.1$

$H_A: \theta_A > 0.1$

Significance level:

α = P(rejecting null, when null true) $X \sim Binom(400, \theta)$

$\alpha = P(X \geq 50 \mid \theta = 0.1)$

Using R: sum(dbinom(50;400,400,0.1)) } sums the probability of getting 50 successes in 400 trials, given a success rate of θ

$\alpha = 0.06009$

Power = P(null hypothesis is rejected | null is false)

For $\theta = 0.11$

Power = $P(X \geq 50 \mid \theta = 0.11)$

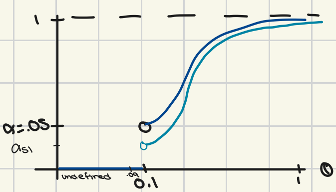
sum(dbinom(50;400,400,0.11)) = 0.1883574

For $\theta = 0.15$

Power = $P(X \geq 50 \mid \theta = 0.15)$

sum(dbinom(50;400,400,0.15)) = 0.9320924

For $\theta = 0.09$



* not $\theta = 0.09$ is not within our alternative hypothesis so the power at $\theta = 0.09$ is 0 or undefined

- this is a limitation to a 1-sided test

If threshold was 51 or more: at 51: $\alpha = .0436 \downarrow$ $\theta = .11 \rightarrow$ power = 0.149 $\downarrow$ $\theta = .15 \rightarrow$ power = 0.91

Significance Level = P(Type I Error), If you make the threshold 51 you are requiring more evidence to reject the null. This will reduce the significance level.

Power = 1 − P(Type II Error), If you increase the threshold to 51 it will be harder to reject the null when the alternative is true. Because the null will require more evidence to reject. So the power will decrease too.

c) Define the P-value of a test and its connection to the significance level. With $n = 400$, suppose you observe 45 defects. What is the P-value?

Def: P-value: the smallest significance level at which the null hypothesis would be rejected (Rice 335)

↳ the probability under the null hypothesis of a result as or more extreme than that actually observed

If p-value $\leq \alpha$, reject H_0

If p-value $> \alpha$, fail to reject

Example) $n=400$ defects $=45$ given $\theta=0.1$

$X \sim \text{Binom}(400, 0.1)$

$P(X \geq 45 \mid \theta=0.1)$

sum(dbinom(45:400, 400, 0.1))

p-value $= 0.22367$

d) Describe the connection between power and sample size in the context of the defect rate example. Show how to solve for the smallest sample size to have power of at least 0.9 for the alternative in part c, when working at $\alpha \approx 0.05$. Use the Normal approximation to the Binomial to get approximate answers, and then use pbinom to make things precise.

Connection between power and sample size

Power = P(correctly rejecting the null)

In context of defect rates:

- If sample size increases, test becomes more sensitive to detecting differences, making power increase.

Makes it easier to detect between the effects and random noise or errors.

Smallest n to have power of 0.9, with $\alpha = 0.05$.

alternative in part c: $\theta_1 = 0.15$ $\theta_0 = 0.1$

$$\hat{\theta} = \frac{\Sigma x_i}{n} \sim N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$$

Rejection Region:

$$\hat{\theta} > \theta_0 + 1.645\sqrt{\frac{\theta_0(1-\theta_0)}{n}}$$

Approx Power:

$$P\left(Z > \underbrace{\frac{\theta_0 + 1.645\sqrt{\frac{\theta_0(1-\theta_0)}{n}} - \theta_1}{\sqrt{\frac{\theta_1(1-\theta_1)}{n}}}}_{\text{Has to be } \leq -1.282 \text{ so that power } \geq 0.9}\right) \geq 0.9$$

Then, $\theta_0 + 1.645\sqrt{\frac{\theta_0(1-\theta_0)}{n}} - \theta_1 \leq -1.282\sqrt{\frac{\theta_1(1-\theta_1)}{n}}$

$$\sqrt{n}(\theta_1 - \theta_0) \geq 1.645\sqrt{\theta_0(1-\theta_0)} + 1.282\sqrt{\theta_1(1-\theta_1)}$$

$$\sqrt{n} \geq \frac{1.645\sqrt{\theta_0(1-\theta_0)} + 1.282\sqrt{\theta_1(1-\theta_1)}}{(\theta_1 - \theta_0)}$$

$\sqrt{n} \geq 19.025$ so $n \geq 361.9624$

So $n \geq 362$ to get a power of at least 0.9

Then if $n = 362$:

New threshold is, $\left(0.1 + 1.645\sqrt{\frac{0.1(0.9)}{362}}\right) 362 = 45.589$

So new significance level sum(dbinom(46:362,362,0.1)) = 0.5528

Power then is 0.90464 (given $\theta_1 = 0.15$), showing that 362 gives power ≥ 0.9

for threshold of 47

$\alpha = 0.03931$ and power = 0.875