Stat 111 Spring 2025 Week 6: Hypothesis Tests (Rice Chapter 9)

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: Based on Type I and Type II errors

· Neyman - Pearson Lemma (Rice pg. 332): Suppose that H_0 and H_1 are simple hypotheses and that the test that rejects H_0 whenever the likelihood ratio is less than c and signifance level is less than or equal to α has power less than or equal to that of the likelihood ratio test.

Type I error: We reject the H_0 when H_0 is true

Type II error: We fail to reject H_0 when H_1 is true

	Reject H_0	Fail to Reject H_0
H_0 is True	Type I Error	Correct Non-Rejection
H_1 is true	Correct Rejection	Type II Error

Court Room Analogy: "Innocent Until proven guilty"

· Can not reject the null hypothesis until we do a test to reject it

H_0: Not guilty

H_1: Guilty

· Type I error: Actually not guilty, but found guilty

· Type II error: Actually guilty, but not guilty

Type II errors are better than Type I

Asymmetry: Can reject one but not accept the other without a new test, Can flip them

· Add null and alternative hypothesis structure

· Definition of tests
  · Test statistic and rejection region
  · Different test statistic mean different test

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.

$H_0 : \theta = 0.1$

· Definition of Significance level and power

$H_1 : \theta > 0.1$

$\mathbb{P}$(rejecting null, when null is true)

$\alpha = 0.06$ (dbinom(

$\mathbb{P}(x \geq 50 \mid \theta = 0.1)$

· Fix Type I error

goes to 0.043 from 51 or over

· Significance level: depends on the test

Power of the test: $\mathbb{P}(x > 0.1 \mid \text{known } \theta)$

$\theta = 0.11 = \dfrac{0.1 - 0.11}{\sqrt{\frac{(0.11)(1-0.11)}{400}}} = z = -0.639$, From z-score table: 0.26

$= 1 - 0.2643 = 0.7357$

$\theta = 0.15 = \dfrac{0.1 - 0.15}{\sqrt{\frac{(0.15)(1-0.15)}{400}}} = -2.8$, From z-score table: 0.0026

$= 1 - 0.0026 = 0.9974$

$\theta = 0.09 = \dfrac{0.1 - 0.09}{\sqrt{\frac{(0.09)(1-0.09)}{400}}} = .698$, From z-score table: .7549

$= 1 - 0.7549 = 0.2451$

Really 0, from alternative hypothesis, wrong direction

limitation

The connection between significance level and power is if we increase the significance level, then the power will increase as well. Since the power is rejecting the null hypothesis when it is false, if we increase the area to reject the null hypothesis (significance level) they will relate to each other.

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?

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

Clear connection to significance level from being the smallest, and we directly compare the p-value to the significance level.

Test statistic, all other values that are in the direction of the alternative

With $n = 400$, we observe 45 defects:

dbinom

$$\frac{0.1125 - 0.1}{\sqrt{\frac{(0.1)(1-0.1)}{400}}} = 0.833$$

Z-score table = 0.7967

p-value = 0.2033

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.

As the sample size increases, the power will increase because an increased sample size reduces the chance of an error, meaning the power will be increased.

Smallest sample size = $0.9 \approx -1.28$

$$-1.28 = \frac{0.1125 - 0.1}{\sqrt{\frac{(0.1)(1-0.1)}{n}}}$$

$n = 943.7184$

p binom : ?

5,067

```
x=rgamma(100000,4,2)
mean(x)
var(x)

y=log(x)
mean(y)
var(y)

log(4/2)-1/(2*4)
digamma(4)-log(2)

trigamma(4)


pbinom()
sum(dbinom(50:400, 400, 0.10))
sum(dbinom(51:400, 400, 0.10))
sum(dbinom(51:400, 400, 0.11))
sum(dbinom(51:400, 400, 0.15))
sum(dbinom(51:400, 400, 0.09))


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


theta0=0.1
theta1vals =seq(.101,.2,.001)

powervals=0*theta1vals
for(i in 1:length(theta1vals))
  powervals[i]=sum(dbinom(50:400,400,theta1vals[i]))

plot(theta1vals,powervals,type="l",xlim=c(0.09,0.2),ylim=c(0,1), main="Power", xlab="Theta", ylab="Power"
points(.1,.05)
points(.15,sum(dbinom(50:400, 400, 0.15)))
points(.11,sum(dbinom(50:400, 400, 0.11)))
points(.09,sum(dbinom(50:400, 400, 0.09)))

1.645*sqrt((.1)*(.9))/0.05+1.282*sqrt((0.15)*(0.85))/(0.05)
19.02^2

sum(dbinom(46:362, 362, 0.15))
sum(dbinom(46:362, 362, 0.1))
```