Day 8

Outline

1. A History Lesson
2. Neyman-Pearson Hypothesis Testing

History Lesson

Major Players

- Karl Pearson
- Egon Pearson
- Jerzy Neyman
- Ronald Fisher

TL;DR

This test will allow us to make preemptive decisions based on conditions presented before the study is conducted. These are the theoretical outcomes WITHOUT taking any sample data

Neyman-Pearson Hypothesis Testing

TL;DR version

1. Define a boundary used to inform a decision
2. Obtain data and see which side of the boundary it falls on
3. Make decision

Example

We have a coin and it is weighted but we don't know if it's weighted to be 60% heads or 60% tails.

Define a parameter to describe the situation

Let `P` represent probability of getting heads ("population proportion of heads")

Define two competing "hypothesis" involving the parameter.

(heads)

- H_0 : `P = 0.6` [null hypothesis: "nothing unexpected"]
- H_1 : `P = 0.4` [alternative hypothesis: "something is happening, we should change our minds"]

Define a "critical region" based on our sample data

1. Define a test statistic `T` whose value can be computed from the sample data
2. Define the sampling distribution of `T` under H_0 and H_1
3. Based on the sampling distribution under H_0, define:
    - $\alpha = P(we\, claim\, H_1\, is\, true | H_0\, is\, true)$ and find the region in the sampling distribution under H_0 corresponding to that α value.
4. If the observed value of `T` is in that region, conclude H_1 is true. Otherwise, conclude H_0 is true

Critical region: a range of values that corresponds to the rejection of the null hypothesis at some chosen probability level.

Example

Our decision rule:

- If we get 4 or fewer heads in 10 flips: conclude H_1 is true.
- If more than 4 heads in 10 flips: conclude H_0 is true.

"Critical region": Let $X = number\, of\, heads\, in\, 10\, flips$

- $X \leq 4$

Recall:

Gender compared to handedness

	Handed		
	Left	Right	
Female	7	46	53
Male	5	63	68
	12	109	121

Now apply this to Neyman-Pearson rules:

	Do not reject H_0	Reject H_0
H_0 is true	Correct Decision	Incorrect Decision: Type I error α
H_0 is false	Incorrect Decision: Type II error β	Correct Decision

Under N-P Rules

Type 1 Error is "worse" than Type 2 Error. However, if P(Type 1 Error) is too low, P(Type 2 Error) balloons.

$\alpha = P(1)$ - P(Concluded H_1 | H_0 is true)

$\beta = P(2)$ - P(Concluded H_0 | H_1 is true)

Power of test $= 1 - \beta$

- = P(concluded H_1 | H_1 is true)

Example [Continued from Above]

Let $X = number\,of\,heads\,in\,10\,flips$

- Under H_0: $X \sim B(10, 0.6)$
- Under H_1: $X \sim B(10, 0.4)$

For critical region X $\leq$ 4:

- $\alpha = P(X \leq 4\,|\,p = 0.6) = 0.166$
- $\beta = P(X > \,|\,p = 0.4) = 0.367$

Power $= P(X \leq 4\,|\,p = 0.4) = 0.633$

Traditionally, set $\alpha = 0.05$ or $\alpha = 0.01$

- α refers to the probability of making a Type I Error.

Find the critical region giving a Type 1 Error rate of at most α

(Find x such that P(X $\leq$ x | H_0 is true) $\leq \alpha$)

P(x $\leq$ 2 | H_0 is true) = 0.0123

P(X $\leq$ 3 | H_0 is true) = 0.0548

Critical region corresponding to $\alpha = 0.05$: $x \leq 2$

What is β for this critical region?

- $\beta = P(x > 2|p = 0.4) = 0.833$

In most fields, we use power instead

Power $= P(X \leq 2|p = 0.4) = 0.167$

Rules of Thumb

1. $\alpha < \beta$. If $\alpha \leq \beta$, either decrease α or switch H_0 or H_1
2. At your "given" α value, $\beta \leq 2$ or equivalently, power ≥ 0.8 (80% power). If power < 0.8, plan to collect more data!

Power must be at least 80 percent

In Practice

1. The idea of "nothing weird happening" should give us the value of the parameter.
2. We define a clinically signifiant/practically signifiant difference in parameter values ("minimum effect size")

What we need at each step

1. To compute the critical region:

- need α, H_0 (value of P under H_0)
- sampling distribution of test statistic under H_0

2. To compute power:

- need critical region, H_1 (value of P under H_1)
- sampling distribution of test statistic under H_1