STAT 111

Recitation 8

Mo Huang

Email: mohuang@wharton.upenn.edu

Office Hours: Wednesdays 3:00 - 4:00 pm, JMHH F96

Slides (adapted from Gemma Moran): `github.com/mohuangx/STAT111-Fall2018`

November 2, 2018

Introduction to Hypothesis Testing

- Statistics is *inductive* (or "bottom-up logic"). We start with finite observations and then attempt to make objective statements about the world or **test hypotheses**.
- For example: Is this coin fair? Is this medicine more effective?
- **Null hypothesis** (H_0)**:** A general statement indicating no significant difference or phenomenon.
  - For example: This coin is fair. This medicine is equally as effective.
- **Alternative hypothesis** (H_1)**:** The hypothesis used that is contrary to the null hypothesis.
  - For example. This coin is unfair. This medicine is more effective.
- **Basic premise of hypothesis testing:** Observe a random sample from a population. If the sample is consistent with H_0, do not reject H_0. If the sample is inconsistent or unlikely under H_0, then reject H_0 in favor of H_1.

Approach to hypothesis testing

1. Define the null hypothesis H_0 and the alternative hypothesis H_1 in terms of parameters.
2. Choose a type I error α.
3. Determine the test statistic.
4. Calculate the critical region/p-value.
5. Reject or not.

1. Null and alternative hypotheses

- Null hypotheses are most often single values. In the fair coin example, $H_0 : \theta = 0.5$.
- Alternative hypotheses are typically ranges and can be one-sided or two-sided.
- For example:
  - One-sided: The coin is more likely to favor heads ($H_1 : \theta > 0.5$).
  - Two-sided: The coin is unfair ($H_1 : \theta \neq 0.5$).
- Whether the alternative is one-sided or two-sided affects the probability calculation.

2. Choose a type I error α

- The type I error α is the probability of rejecting the null hypothesis when the null hypothesis is true. Can think of this as a mistaken rejection.
- The experimenter can set α depending on how willing they are to mistakenly reject H_0. Most common values are 0.01 and 0.05.
- For experiments where we want to be really sure we are not making a mistake, we would want α to be small.
- If $\alpha = 0.05$, this means that in roughly 5 out of every 100 repetitions of the experiment, we will mistakenly reject H_0.

3. Determining the test statistic

- The test statistic is a numerical function of the data we use to reject or accept the null hypothesis.
- Most often, it is the estimate of the parameter.
- For example, the test statistic for the fair coin flip example ($H_0 : \theta = 0.5$) is the proportion of heads.
- We can also use the number of heads, which would involve a different probability calculation.

4-5. Determining significance and rejection

- Two equivalent approaches:
  1. Determine critical region/point at level α and reject H_0 if test statistic falls inside critical region.
  2. Calculate p-value of the test statistic and reject H_0 if less than α.
- Approach 1
  - The **critical region** is calculated such that the probability the test statistic falls in the critical region is α if H_0 is true.
  - One-sided: $P(X \geq A) = \alpha$ or $P(X \leq A) = \alpha$.
    Two-sided: $P(X \leq A) = P(X \geq B) = \alpha/2$.
  - Reject H_0 if test statistic falls inside critical region.
- Approach 2
  - The **p-value** is the probability under H_0 of obtaining a result of more or equal extremeness than the observed test statistic.
  - One-sided: p-value $= P(X \geq x)$ or $P(X \leq x)$.
    Two-sided: p-value $= 2P(X \geq |x|)$ (for symmetric distributions)
  - Reject H_0 if p-value is less than α.

Example

- I want to test the hypothesis a coin is unbiased. I observed 1,070 heads out of 2,000 tosses. Let θ be the probability of tossing a head. Using approach I, calculate the critical point(s) and decide whether to reject H_0 at $\alpha = 0.05$.

Ans: $H_0 : \theta = 0.5$ vs. $H_1 : \theta \neq 0.5$.

Test-statistic: X is the number of heads tossed. $x = 1070$.

Under H_0, $\mu = n\theta = 2000(0.5) = 1000$ and $\sigma^2 = n\theta(1-\theta) = 500$.

We have:

$$P(Z \leq -1.96) = 0.025 \text{ and } P(Z \geq 1.96) = 0.025$$

$$P\left(\frac{X - 1000}{\sqrt{500}} \leq -1.96\right) = 0.025 \text{ and } P\left(\frac{X - 1000}{\sqrt{500}} \geq 1.96\right) = 0.025$$

$$P(X \leq 956) = 0.025 \text{ and } P(X \geq 1044) = 0.025$$

So critical region is $X \leq 956$ and $X \geq 1044$. Since 1,070 is in the critical region, we reject H_0.

Example

- I want to test the hypothesis a coin is unbiased. I observed 1,070 heads out of 2,000 tosses. Let θ be the probability of tossing a head. Calculate the p-value and decide whether to reject H_0 at $\alpha = 0.05$.

Ans: $H_0 : \theta = 0.5$ vs. $H_1 : \theta \neq 0.5$.

Test-statistic: X is the number of heads tossed. $x = 1070$.

Under H_0, $\mu = n\theta = 2000(0.5) = 1000$ and $\sigma^2 = n\theta(1 - \theta) = 500$.

We standardize:

$$z_{obs} = \frac{X - 1000}{\sqrt{500}} = \frac{1070 - 1000}{\sqrt{500}} = 3.13.$$

For a two-sided test:

$$\text{p-value} = 2 \times P(Z > |3.13|) = 2 \times (1 - 0.9991) = 0.0018$$

p-value < 0.05, so we reject H_0.

Example

- I want to test the hypothesis a coin is unbiased. I observed 1,070 heads out of 2,000 tosses. Let θ be the probability of tossing a head. How would the approach be different if we were looking at the *proportion*?

Ans: $H_0 : \theta = 0.5$ vs. $H_1 : \theta \neq 0.5$.

Test-statistic: P is the proportion of heads tossed. $p = 0.535$.

Under H_0, $\mu = \theta = 0.5$ and $\sigma^2 = \frac{\theta(1-\theta)}{n} = \frac{1}{8000}$.

Do the same as before...