STAT 513 fa 2020 Lec 01 slides

Inference, tests of hypotheses, power and size

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Table of Contents

Consider a parameter of interest $\theta \in \Theta$.

Null and alternate hypotheses

Consider *null and alternative hypotheses* H_0 and H_1 of the form

$$H_0\colon \theta \in \Theta_0 \text{ versus } H_1\colon \theta \in \Theta_1,$$

where

$$\Theta_1 = \Theta \setminus \Theta_0 = \Theta \cap \Theta_0^c$$

so that $\Theta_0 \cup \Theta_1 = \Theta$.

Call Θ_0 the *null space* and Θ_1 the *alternate space*.

A *statistical inference* is a decision to reject or not reject H_0 based on data.

If we reject H_0, we conclude that H_1 is true. If we fail to reject H_0, we do not make any conclusion.

Examples:

- For $p \in (0,1)$, might test H_0: $p = 1/2$ versus H_1: $p \neq 1/2$.
- For $\mu \in (0,\infty)$, might test H_0: $\mu \leq 2$ versus H_1: $\mu > 2$.
- For $\delta \in (-\infty,\infty)$, might test H_0: $\delta = 0$ versus H_1: $\delta \neq 0$.

Identify Θ, Θ_0, and Θ_1 for the above examples.

Simple and composite hypotheses

- *Simple hypotheses* specify a single value for a parameter.
- *Composite hypotheses* specify multiple possible values for a parameter.

Consider previous examples.

Hypothesis test/test of hypotheses

A *hypothesis test* is a rule for deciding whether or not to reject H_0 based on data.

Given a rs $X_1, \dots, X_n$ and hyps. H_0 and H_1, tests of hypotheses take the form

$$\text{Reject } H_0 \text{ iff } T(X_1, \dots, X_n) \in \mathcal{R}.$$

The function $T(X_1, \dots, X_n)$ of the sample values called the *test statistic*.

The set $\mathcal{R}$ is called the *rejection region*.

Exercise: Identify (i) H_0 and H_1, (ii) T, and (iii) $\mathcal{R}$ for the following:

1. Based on 10 coin tosses with the probability of "heads" equal to p, call the coin unbalanced if more than six or fewer than four "heads" are observed.
2. Conclude that the mean monthly rent μ paid by USC students is greater than 600 if mean rent of 20 randomly sampled students exceeds 650.
3. Infer that the standard deviation σ of heights of 2-yr-old trees on a tree farm is less than 3 feet if the standard deviation of the heights of 10 randomly sampled trees is less than 2 feet.

Type I and Type II errors

- A *Type I error* is rejecting H_0 when H_0 is true
- A *Type II error* is failing to reject H_0 when H_0 is false.

Draw that one table...

Jeopardy-style checkpoint:

1. Statistical inference boils down to making decisions concerning these two opposing statements about the data generating process.
2. The null and alternate hypotheses partition the parameter space into these two non-overlapping spaces.
3. This type of hypothesis specifies only a single value for the parameter.
4. This type of inferential error is called a Type I error.
5. This type of inferential error is called a Type II error.
6. A decision whether to reject or not reject the null hypothesis is made based on the value of this quantity.
7. If the test statistic lies in this set, the null hypothesis is rejected.

Power of a test

The *power* of a test is the probability that it will lead to a rejection of H_0.

The power of a test depends on the true value of the parameter θ.

If H_0 is true, we want the power to be ________________(small/large).
If H_0 is false, we want the power to be ________________(small/large).

Recall that tests of hypotheses have the form

$$\text{Reject } H_0 \text{ iff } T(X_1, \ldots, X_n) \in \mathcal{R}.$$

Power function

The *power function* for a test of hyps. about a parameter θ is given by

$$\begin{aligned}\gamma(\theta) &= P(\text{Reject } H_0 \text{ when true value of parameter is } \theta) \\ &= P_\theta(T(X_1, \ldots, X_n) \in \mathcal{R}).\end{aligned}$$

Use P_θ to denote probability computed when the parameter takes the value θ.

Discuss: Interpretations of $\gamma(\theta)$ for $\theta \in \Theta, \Theta_0, \Theta_1$.

Exercise: Based on 10 coin tosses, call the coin unbalanced if more than six or fewer than four "heads" are observed.

1. Identify H_0, H_1, T and $\mathcal{R}$.
2. In the following scenarios, what decision do you make and is it a correct decision, a Type I error, or a Type II error?
   - The true probability of "heads" is 0.6 and you roll 7 heads.
   - The true probability of "heads" is 0.6 and you roll 5 heads.
   - The true probability of "heads" is 0.5 and you roll 7 heads.
   - The true probability of "heads" is 0.5 and you roll 4 heads.
3. Find the power of the test when the true probability of getting "heads" is 0.6.
4. Plot the power $\gamma(p)$ against p for $p = 0.01, 0.02, \dots, 0.99$.
5. Suppose the coin is balanced. What is the probability of a Type I error?
6. Give the probability of a Type II error if the true probability of "heads" is $1/3$.

Size of a test

The *size* of a test, denoted by α, is defined as

$$\alpha = \sup_{\theta \in \Theta_0} \gamma(\theta),$$

which we read as "the supremum of the power $\gamma(\theta)$ over all $\theta \in \Theta_0$".

- The size is the maximum power over the null space.
- The size is the largest probability of a Type I error over all $\theta \in \Theta_0$.
- If the null space contains a single point, say $\Theta_0 = \{\theta_0\}$, then the size is

$$\alpha = \sup_{\theta \in \{\theta_0\}} \gamma(\theta) = \gamma(\theta_0).$$

Exercise (cont): Based on 10 coin tosses, call the coin unbalanced if more than six or fewer than four "heads" are observed. What is the size of the test?

Exercise: Let $X_1, \dots, X_{10} \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$, $p \in (0,1)$ unknown, and

$$H_0\colon p \leq 1/4 \text{ versus } H_1\colon p > 1/4.$$

Consider the test

$$\text{Reject } H_0 \text{ iff } X_1 + \cdots + X_{10} > 5.$$

1. Find an expression for the power function.
2. Calculate the size of the test.
3. Make a plot of the power function.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, 1)$, $\mu \in (-\infty, \infty)$ unknown, and

$$H_0\colon \mu = 0 \text{ versus } H_1\colon \mu \neq 0.$$

Consider the test

$$\text{Reject } H_0 \text{ iff } |\sqrt{n}\bar{X}_n| > 2.$$

1. Find an expression for the power function.
2. Calculate the size of the test.
3. Make a plot of the power function.

Exercise: Let $X_1, \dots, X_{25} \stackrel{\text{ind}}{\sim} \text{Exponential}(\lambda)$, $\lambda \in (0, \infty)$ unknown, and

$$H_0\colon \lambda \geq 2 \text{ versus } H_1\colon \lambda < 2.$$

Consider the test

$$\text{Reject } H_0 \text{ iff } \bar{X}_{25} < 1.5.$$

1. Find an expression for the power function.
2. Calculate the size of the test.
3. Make a plot of the power function.

- We often choose the rejection region in order to control the size of the test.
- We cannot directly control the Type II error rate.
- The smaller the size of a test, the stronger the evidence against H_0 must be in order for the test to reject H_0.

Exercise: Revisit each of the last three exercises and

1. Modify the rejection rejection region to make size ≤ 0.01.
2. Plot new power curves alongside old power curves.