STAT 713 sp 2023 Lec 10 slides

Tests of hypotheses, size, power, and p-values

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.

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 $\Theta_1 \cap \Theta_0 = \emptyset$ and $\Theta_0 \cup \Theta_1 = \Theta$.

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

A *statistical inference* is a decision, based on the data, to

1. reject H_0 and conclude that H_1 is true.
2. not reject H_0 and conclude nothing.

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.

Hypothesis test/test of hypotheses

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

Given data $\mathbf{X}$ and hyps. H_0 and H_1, tests of hypotheses take the form

$$\text{Reject } H_0 \text{ iff } T(\mathbf{X}) \in \mathcal{R}.$$

The function $T(\mathbf{X})$ of the data is called the *test statistic*.

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

A test as a binary function

We sometimes express a test as a function $\phi(\mathbf{X}) = \mathbf{1}(T(\mathbf{X}) \in \mathcal{R})$. Then

$$\begin{aligned} \phi(\mathbf{X}) = 1 \quad &\implies \quad \text{Reject } H_0 \\ \phi(\mathbf{X}) = 0 \quad &\implies \quad \text{Fail to reject } H_0 \end{aligned}$$

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.

Make table...

Power function, size, and level of a test

Consider a test of H_0: $\theta \in \Theta_0$ vs H_1: $\theta \in \Theta_1$ which rejects H_0 when $T(\mathbf{X}) \in \mathcal{R}$.

1. The *power function* of the test is given by

$$\gamma(\theta) = P(\text{Reject } H_0 \text{ when true value of parameter is } \theta) = P_\theta(T(\mathbf{X}) \in \mathcal{R}).$$

2. The *size* of the test is

$$\sup_{\theta \in \Theta_0} \gamma(\theta) = \text{maximum probability of a Type I error}$$

3. The test is said to have *level* α if its size is no greater than α.

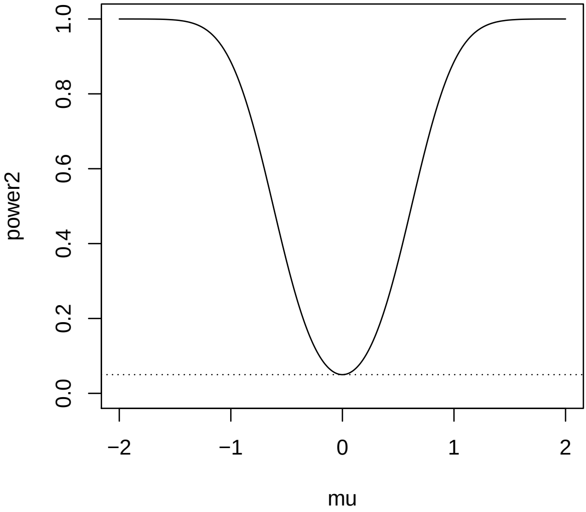
Discuss: Interpretations of $\gamma(\theta)$ for $\theta \in \Theta, \Theta_0, \Theta_1$.

Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$ unknown, and

$$H_0\colon \mu = \mu_0 \text{ versus } H_1\colon \mu \neq \mu_0.$$

Consider the test $\phi(\mathbf{X}) = 1 \iff \sqrt{n}|\bar{X}_n - \mu_0|/\sigma > z_{\alpha/2}$, for some $\alpha \in (0, 1)$.

1. Find an expression for the power function.
2. Calculate the size of the test.
3. Make a plot of the power function under $\alpha = 0.05$, $\sigma = 1$, $\mu_0 = 0$, and $n = 10$.


1.0
0.8
0.6
0.4
0.2
0.0
power2
−2
−1
0
1
2
mu

Exercise: Let $Y \sim \text{Binomial}(n, p)$. Test H_0: $p \leq p_0$ vs H_1: $p > p_0$ with the decision rule $Y > B_{n,p_0,\alpha}$, where $B_{n,p,\alpha} = \inf\{y : P_p(Y \leq y) \geq 1 - \alpha\}$.

1. Give an expression for the power function of the test.
2. Give an expression for the size of the test.
3. Argue that the test has level α for any p_0, n, and α.
4. Plot power function and give size when $\alpha = 0.05$, $p_0 = 1/2$, and $n = 20$.

All rejection rules can be written as $T(\mathbf{X}) > c$ for some $T(\mathbf{X})$ and c.

Theorem (Calibrating the rejection region to control the level)

We may define a level-α test by the rejection rule $T(\mathbf{X}) > c_\alpha$, where

$$c_\alpha = \inf\left\{c : \sup_{\theta \in \Theta_0} P_\theta(T(\mathbf{X}) > c) \leq \alpha\right\}.$$

If $T(\mathbf{X})$ is continuous for all θ then the test has size α.

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

$$H_0\colon \lambda \leq 2 \text{ versus } H_1\colon \lambda > 2.$$

Consider the test with rejection rule $\bar{X}_n > c$.

1. Find the value c such that the test has size $\alpha = 0.05$.
2. Give the power function under this choice of c.

The power function helps with sample size calculations.

Sample size calculation

For testing $H_0\colon \theta \in \Theta_0$ vs $H_1\colon \theta \in \Theta_1$, ask:

1. What is a meaningful alternative (meaningful non-nullity)? Choose $\theta_1 \in \Theta_1$.
2. With what probability do you want to reject H_0? Choose $\gamma^* \in (0, 1)$.

Suppose your test has power function $\gamma_n(\theta)$ under the sample size n.

Then find the smallest n such that $\gamma_n(\theta_1) \geq \gamma^*$.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, 4)$. Reject $H_0\colon \mu \geq 500$ if

$$\sqrt{n}(\bar{X}_n - 500)/2 < -z_{0.05}$$

Find n required to reject H_0 with probability at least 0.80 for any $\mu \leq 499$.

The p-value or observed significance level

For a test of H_0: $\theta \in \Theta_0$ vs H_1: $\theta \in \Theta_1$ with rej. rule $T > c$, define the function

$$p(t) = \sup_{\theta \in \Theta_0} P_\theta(T \geq t) = \text{size of test with rejection rule } T \geq t.$$

Then $p(T)$ is the *p-value* corresponding to the test statistic value T.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$. Reject H_0: $\mu = \mu_0$ if

$$\left|\sqrt{n}(\bar{X}_n - \mu_0)/\sigma\right| > c \quad \text{for some } c > 0.$$

Give the formula for computing the p-value based on some observed data.

Theorem (The p-value as a test statistic)

If $p(T)$ is a p-value for testing $H_0\colon \theta \in \Theta_0$ vs $H_1\colon \theta \in \Theta_1$, then

$$P_\theta(p(T) \leq \alpha) \leq \alpha \text{ for all } \alpha \in (0,1) \text{ and all } \theta \in \Theta_0.$$

Therefore

1. $p(T) \leq \alpha$ *gives an* α*-level test.*
2. *If* T *is continuous the test has size* α *and* $p(T) \sim \text{Uniform}(0,1)$.

Exercise: Go through proof.