Chapter 9: Hypothesis Testing

Sections

- 9.1 Problems of Testing Hypotheses - we are still here
- Skip: 9.2 Testing Simple Hypotheses
- Skip: 9.3 Uniformly Most Powerful Tests
- Skip: 9.4 Two-Sided Alternatives
- 9.5 The *t* Test
- 9.6 Comparing the Means of Two Normal Distributions
- 9.7 The *F* Distributions
- 9.8 Bayes Test Procedures
- 9.9 Foundational Issues

Hypothesis testing - review from last time

- *Hypothesis testing*: Inferential method to decide between two complimentary *hypotheses* about a parameter.

$$H_0 : \theta \in \Omega_0 \qquad \text{Null Hypothesis}$$
$$H_1 : \theta \in \Omega_1 \qquad \text{Alternative Hypothesis}$$

- Two possible decisions:

Decide that $\theta \in \Omega_0$ i.e. we *do not reject* H_0

Decide that $\theta \in \Omega_1$ i.e. we *reject* H_0

- Test procedure: For data in critical region S_1 or *test statistic* in rejection region R, reject H_0.

Hypothesis testing - more review from last time

- Power-function:

$$\pi(\theta|\delta) = P(\text{reject } H_0|\theta) = P(\mathbf{X} \in S_1|\theta) \quad \text{for } \theta \in \Omega$$

- *Type I error*: Wrongly deciding to reject H_0
  - Rejecting $H_0 : \theta \in \Omega_0$ when in fact $\theta \in \Omega_0$
- *Type II error*: Wrongly deciding not to reject H_0
  - Don't reject $H_0 : \theta \in \Omega_0$ when in fact $\theta \notin \Omega_0$
- Relation to power function:

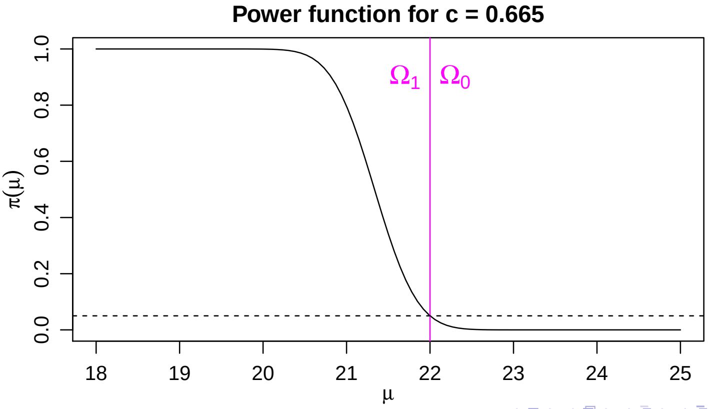
$$\text{probability of type I error} = \begin{cases} \pi(\theta|\delta) & \text{for } \theta \in \Omega_0 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{probability of type II error} = \begin{cases} 1 - \pi(\theta|\delta) & \text{for } \theta \in \Omega_1 \\ 0 & \text{otherwise} \end{cases}$$

Hypothesis testing - More review

Power function for the After-school example

A size $\alpha(\delta) = 0.05$ test for $H_0 : \mu \geq 22$ and $H_1 : \mu < 22$.
Also a level $\alpha_0 = 0.05$ test,
and a level α_0 test for any $\alpha_0 \geq 0.05$



Example: Two-sided Z-test

Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, 1)$, $n = 25$, and suppose we want to test the hypotheses

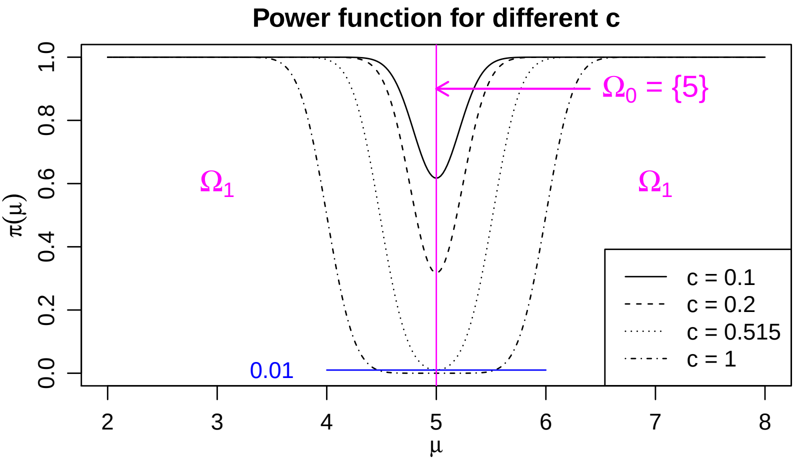
$$H_0 : \mu = \mu_0 \qquad \text{and} \qquad H_1 : \mu \neq \mu_0$$

Let δ_c be the test that rejects H_0 iff $|\overline{X}_n - \mu_0| \geq c$

- Find the power function $\pi(\mu|\delta_c)$
- Find the value of c so that δ_c is of size 0.01

Example: Two-sided Z-test

Power function for $\mu_0 = 5$



Example: Service times

A manager at a local Wells Fargo branch thinks that one of the tellers is working too slowly. The teller protests and claims his average service time is less than 2 minutes. He decides to conduct a hypothesis test to support his claim.

- Assume that the service times, X_i, of n randomly selected customers are i.i.d $\mathrm{Expo}(\theta)$
- He wants to test the hypotheses $H_0 : \theta \leq 0.5$ vs. $H_1 : \theta > 0.5$ (why?)
- He decides to use the test procedure δ_c that rejects H_0 iff $T = \sum_{i=1}^{n} X_i \leq c$

Example: Service times

1. Show that $\pi(\theta|\delta_c)$ is an increasing function of θ
2. Find c so that δ_c is of size α_0
3. He measures service times of 25 randomly selected customers and observes $T = 30.4$ min.
   Will he reject H_0 on the 0.05 significance level?
   What about the 0.01 significance level?

p-values

- Hypothesis testing end in either “reject” or “not reject”.
- Seems inefficient use of data. How close were we to making the other decision? What if we want to use a different level?

Def: p-value

The *p-value* is the smallest level α_0 such that we would reject the null hypothesis at level α_0 after seeing the data

- We reject H_0 if and only if the p-value we get is smaller than the pre-determined level of significance α_0
- We can also say that the observed test statistic is *just significant* at level equal to the p-value

p-value for the After-School example

- $X_1, \ldots, X_n$ i.i.d. $N(\mu, 6^2)$, $n = 220$.
- Test: $H_0 : \mu \geq 22$ and $H_1 : \mu < 22$.
- Hypotheses procedure: Reject H_0 iff $\overline{X}_n \leq 22 - 0.665 = 21.335$.
- We established last time that this is a size 0.05 test.
- Suppose we observe $\overline{X}_n = 21.1$ and hence reject H_0.

Find the p-value for the observed data.

p-value for disease example

- $X_1, \ldots, X_{80}$ i.i.d. $\text{Bernoulli}(p)$
- Hypotheses: $H_0 : p \leq 0.02$ and $H_1 : p > 0.02$
- Test: Reject H_0 if $Y = \sum_{i=1}^{80} X_i > c$.
- Suppose we observe $Y = 6$. Find the p-value for the observed data.

c	1	2	3	4	5	6
$P(Y > c \mid p = 0.02)$	0.477	0.216	0.077	0.022	0.005	0.001

Tests and Confidence intervals

There is a relationship between a confidence interval for θ and a hypothesis of the form

$$H_0 : \theta = \theta_0 \quad \text{and} \quad H_1 : \theta \neq \theta_0$$

- We can obtain a $\gamma = 1 - \alpha_0$ confidence set from an α_0 level test.
- We can obtain an $\alpha_0 = 1 - \gamma$ level test from a $100\gamma\%$ confidence set for θ

For one-sided test, such as

$$H_0 : \theta \leq \theta_0 \quad \text{and} \quad H_1 : \theta > \theta_0$$

we only get one direction (in general):

- We can obtain a $\gamma = 1 - \alpha_0$ confidence set from an α_0 level test.
- Only in special cases can we obtain a $\alpha_0 = 1 - \gamma$ level test from a one-sided confidence interval

Tests and Confidence intervals

Theorem 9.1.1: Test $\longrightarrow$ Confidence Set

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from a distribution that is indexed by a parameter θ. Let $g(\theta)$ be the parameter of interest and δ_{g_0} be a level α_0 test of the hypothesis

$$H_{0,g_0} : g(\theta) = g_0 \quad \text{and} \quad H_{1,g_0} : g(\theta) \neq g_0$$

Define $\omega(\mathbf{x}) = \{g_0 : \delta_{g_0} \text{ does not reject } H_{0,g_0} \text{ if } \mathbf{X} = \mathbf{x} \text{ is observed } \}$
Then the random set $\omega(\mathbf{X})$ satisfies

$$P(g(\theta) \in \omega(\mathbf{X}) | \theta = \theta_0) \geq \gamma$$

for all $\theta_0 \in \Omega$, i.e. $\omega(\mathbf{X})$ is a $100\gamma\%$ *confidence set for* $g(\theta)$.

Also works for one-sided tests (Theorem 9.1.3)

Tests and Confidence intervals

Theorem 9.1.2: Confidence Set $\longrightarrow$ Test

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from a distribution that is indexed by a parameter θ. Let $g(\theta)$ be the parameter of interest and let $\omega(\mathbf{X})$ be a $100\gamma\%$ confidence set for $g(\theta)$. Let δ_{g_0} be a test of the hypothesis

$$H_{0,g_0} : g(\theta) = g_0 \quad \text{and} \quad H_{1,g_0} : g(\theta) \neq g_0$$

where δ_{g_0} rejects H_{0,g_0} iff $g_0 \notin \omega(\mathbf{X})$.
Then δ_{g_0} is a level $\alpha_0 = 1 - \gamma$ test of the above hypothesis.

Example

Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ and

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \sigma' = \left(\frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{n-1}\right)^{1/2}$$

We know that

$$\left(\overline{X}_n - T_{n-1}^{-1}\left(\frac{\gamma+1}{2}\right)\frac{\sigma'}{\sqrt{n}},\ \overline{X}_n + T_{n-1}^{-1}\left(\frac{\gamma+1}{2}\right)\frac{\sigma'}{\sqrt{n}}\right)$$

is a $100\gamma\%$ confidence interval for μ.

- Construct a level $\alpha_0 = 1 - \gamma$ test of the hypothesis

$$H_0 : \mu = \mu_0 \qquad \text{and} \qquad H_1 : \mu \neq \mu_1$$

Likelihood ratio tests

A popular way of constructing tests

Def: Likelihood Ratio Test (LRT)

The statistic

$$\Lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Omega_0} f_n(\mathbf{x}|\theta)}{\sup_{\theta \in \Omega} f_n(\mathbf{x}|\theta)}$$

is called the *likelihood ratio statistic*. The *likelihood ratio test (LRT)* of

$$H_0 : \theta \in \Omega_0 \quad \text{vs} \quad H_1 : \theta \in \Omega_1$$

is to reject H_0 if $\Lambda(\mathbf{x}) \leq k$ for some constant k

- Note: If $\hat{\theta}$ is the MLE of θ then

$$\sup_{\theta \in \Omega} f_n(\mathbf{x}|\theta) = f_n(\mathbf{x}|\hat{\theta})$$

Example: Z-test as a LRT

- Let $X_1, \ldots, X_2$ be i.i.d. $N(\mu, \sigma^2)$, where σ^2 is known
- Consider the hypotheses

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0$$

- Find the likelihood ratio test of these hypotheses

Criticism on hypothesis testing

- Hypothesis testing is meant to be a decision problem, but yet there is no loss function or utility function involved!
- The meaning of "do not reject" H_0 is not clearly defined. It does not mean that we should accept H_0 as true.
  Some use the phrase "There is no evidence that H_0 is not true".