

# Chapter 9: Hypothesis Testing

## Sections

- 9.1 Problems of Testing Hypotheses
- **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

# Introduction

Statistical Inference: Given a probability model  $f(x|\theta)$  (and possibly a prior  $p(\theta)$ ) we may be interested in

- Parameter estimation - Chapters 7 and 8
- Making decisions - Hypothesis testing, Chapter 9
  - E.g. If the disease affects 2% or more of the population, the state will launch a costly public health campaign.  
Do we have evidence that  $\theta$  is higher than 2% ?
- Other things like, prediction, experimental design, etc.

# Introduction

Statistical Inference: Given a probability model  $f(x|\theta)$  (and possibly a prior  $p(\theta)$ ) we may be interested in

- Parameter estimation - Chapters 7 and 8
- Making decisions - Hypothesis testing, Chapter 9
  - E.g. If the disease affects 2% or more of the population, the state will launch a costly public health campaign.  
Do we have evidence that  $\theta$  is higher than 2% ?
- Other things like, prediction, experimental design, etc.

Some of what lies ahead:

- The concept of testing hypotheses and several tools and notation involving that
- “Famous” tests, such as  $t$ -Test, two-sample  $t$ -Test,  $F$ -Test

**We will now stay in the Frequentist realm until I tell you otherwise**

## Example: After-School

- The Chapel Hill & Carrboro City Schools (CHCCS) operate After-School programs in all the Elementary schools in the district.
- Each month they send invoices to families that have children enrolled in these programs for both the monthly fee and any extra charges, such as care during teacher workdays (eg. election day!)
- Customers are expected to use their own envelopes and stamps to return their payments
- Currently, the time it takes to pay bills has a mean of 24 days and a standard deviation of 6 days.



## Example: After-School, continued

- Suppose the chief financial officer (CFO) believes that including a stamped self-addressed envelope would decrease the amount of time it takes to pay the bills.
- She calculates that the improved cash flow from a 2-day decrease in the payment period would pay for the costs for envelopes and stamps. Further decrease would generate profit.
- To test this she randomly selects 220 customers and includes a stamped self-addressed envelope with their invoice
- She assumes that the time to pay a bill follows the normal distribution  $N(\mu, \sigma^2)$ , both parameters unknown

Using the data from her experiment, how can she conclude whether this plan will be profitable?

# The general setup and some definitions

- Have a probability model that has an unknown parameter  $\theta \in \Omega$
- Let  $\Omega_0$  and  $\Omega_1$  be disjoint and  $\Omega = \Omega_0 \cup \Omega_1$
- *Hypothesis testing*: Inferential method to decide between two complimentary *hypotheses* about a parameter.

$$H_0 : \theta \in \Omega_0$$

Null Hypothesis

$$H_1 : \theta \in \Omega_1$$

Alternative Hypothesis

# The general setup and some definitions

- Have a probability model that has an unknown parameter  $\theta \in \Omega$
- Let  $\Omega_0$  and  $\Omega_1$  be disjoint and  $\Omega = \Omega_0 \cup \Omega_1$
- *Hypothesis testing*: Inferential method to decide between two complimentary *hypotheses* about a parameter.

$H_0 : \theta \in \Omega_0$       Null Hypothesis

$H_1 : \theta \in \Omega_1$       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*  $H_0$

# The general setup and some definitions

- Have a probability model that has an unknown parameter  $\theta \in \Omega$
- Let  $\Omega_0$  and  $\Omega_1$  be disjoint and  $\Omega = \Omega_0 \cup \Omega_1$
- *Hypothesis testing*: Inferential method to decide between two complimentary *hypotheses* about a parameter.

$H_0 : \theta \in \Omega_0$       Null Hypothesis

$H_1 : \theta \in \Omega_1$       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*  $H_0$

After-School example: May be interested in testing the following:

$$H_0 : \mu \geq 22$$

$$H_1 : \mu < 22$$



# Simple vs Composite, One or Two sided

Simple and Composite hypotheses:

- If  $\Omega_i$  contains only a single value,  $\Omega_i = \{\theta_i\}$ , then  $H_i$  is a *simple hypothesis*
- If  $\Omega_i$  contains more than a single value then  $H_i$  is a *composite hypothesis*

One-Sided and Two-Sided (for a one-dimensional  $\theta$ )

- *One-sided* hypotheses:

$$\begin{array}{l}
 H_0 : \theta \geq \theta_0 \\
 H_1 : \theta < \theta_0
 \end{array}
 \quad \text{or} \quad
 \begin{array}{l}
 H_0 : \theta \leq \theta_0 \\
 H_1 : \theta > \theta_0
 \end{array}$$

- If the Null Hypothesis is simple the alternative is usually *Two-sided*:

$$H_1 : \theta \neq \theta_0$$

## Test procedure in general

Say  $X_1, \dots, X_n$  are i.i.d.  $f(x|\theta)$  and we are interested in testing

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

Test procedure:

- Let  $S$  be the sample space of  $\mathbf{X} = (X_1, \dots, X_n)$
- Let  $S_0$  and  $S_1$  be a partition of  $S$
- Procedure: Reject  $H_0$  if  $\mathbf{X} \in S_1$ , do not reject  $H_0$  if  $\mathbf{X} \in S_0$
- $S_1$  is called the *critical region* of the test

## Test procedure in general

Say  $X_1, \dots, X_n$  are i.i.d.  $f(x|\theta)$  and we are interested in testing

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

Test procedure:

- Let  $S$  be the sample space of  $\mathbf{X} = (X_1, \dots, X_n)$
- Let  $S_0$  and  $S_1$  be a partition of  $S$
- Procedure: Reject  $H_0$  if  $\mathbf{X} \in S_1$ , do not reject  $H_0$  if  $\mathbf{X} \in S_0$
- $S_1$  is called the *critical region* of the test

Test procedure based on a statistic

- Let  $T = r(\mathbf{X})$  be a statistic and  $R \subset \mathbb{R}$
- Procedure: Reject  $H_0$  if  $T \in R$ , do not reject  $H_0$  if  $T \notin R$
- $T$  is called a *test statistic* and  $R$  is called the *rejection region* of the test

## Example: After-School

We had  $X_1, \dots, X_n$  i.i.d.  $N(\mu, \sigma^2)$  and we are interested in testing

$$H_0 : \mu \geq 22 \quad \text{and} \quad H_1 : \mu < 22$$

- It seems reasonable to reject  $H_0$  if  $\bar{X}_n$  is much lower than 22.
- The test procedure would be to reject  $H_0$  if  $\bar{X}_n < 22 - c$  for some positive constant  $c$
- What is the critical region for this test?

It is usually rather difficult to work with a critical region.  
We usually use a test statistic and a rejection region.

## Example: After-School

We had  $X_1, \dots, X_n$  i.i.d.  $N(\mu, \sigma^2)$  and we are interested in testing

$$H_0 : \mu \geq 22 \quad \text{and} \quad H_1 : \mu < 22$$

- We could use the test statistic

$$U = \frac{\sqrt{n}(\bar{X}_n - 22)}{\sigma'}$$

and reject  $H_0$  if  $U$  is very small, say less than some constant  $c$ .

- What is the rejection region for this test ?

## Power function

We can describe the properties of a test procedure by:

### Power function

Let  $\delta$  be a test procedure. The *power function*  $\pi(\theta|\delta)$  (or just  $\pi(\theta)$ ) is the probability of rejecting  $H_0$  for the given  $\theta$ :

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

If we use a test statistic  $T$  we can write

$$\pi(\theta|\delta) = P(T \in R|\theta) \quad \text{for } \theta \in \Omega$$

- Ideal power function:

$$\pi(\theta|\delta) = 1 \quad \text{for } \theta \in \Omega_1$$

$$\pi(\theta|\delta) = 0 \quad \text{for } \theta \in \Omega_0$$

## Power function for after-school

We had  $X_1, \dots, X_n$  i.i.d.  $N(\mu, \sigma^2)$  and we are interested in testing

$$H_0 : \mu \geq 22 \quad \text{and} \quad H_1 : \mu < 22$$

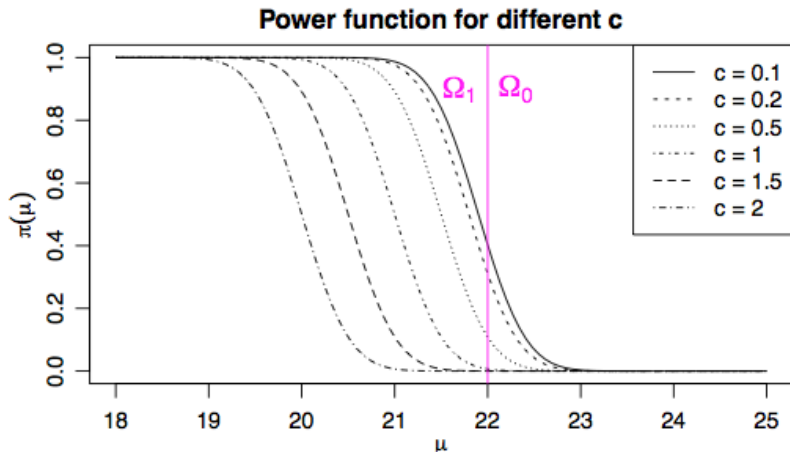
Consider the following test procedure:

- Assume that  $\sigma^2$  is known,  $\sigma^2 = 6^2$
- Reject  $H_0$  if  $\bar{X}_n < 22 - c$ .
  - $\bar{X}_n$  is our test statistic
  - The rejection region is  $(-\infty, 22 - c)$
- The power function is

$$\pi(\theta|\delta) = \Phi\left(\frac{22 - c - \mu}{6/\sqrt{220}}\right)$$

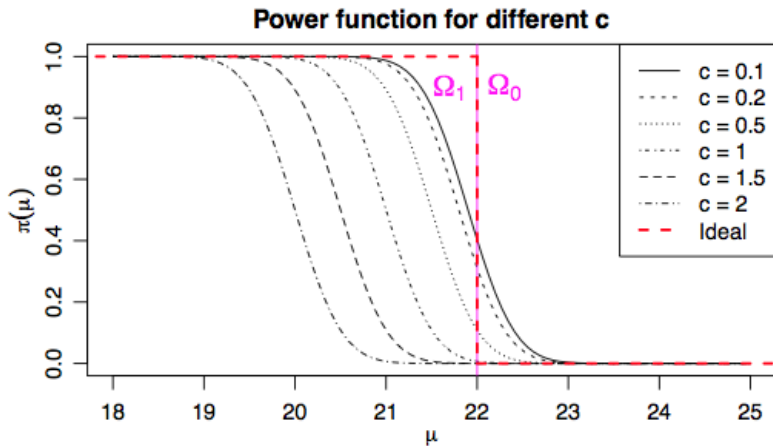
We can then pick  $c$  that gives us the best power function

# Power function for after-school





# Power function for after-school



## Type I and Type II errors

- *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  $\theta \notin \Omega_0$
- What are the type I and type II error for the after school example?

Relation to power function:

- If  $\theta \in \Omega_0$ :  $\pi(\theta|\delta)$  = probability of type I error
- If  $\theta \in \Omega_1$ :  $1 - \pi(\theta|\delta)$  = probability of type II error

## Level and size of tests

Want probability of both types of errors to be small

- Want  $\pi(\theta|\delta)$  to be small for  $\theta \in \Omega_0$  and large for  $\theta \in \Omega_1$
- Generally there is a trade-off between these probabilities
- A popular method: Choose a number  $\alpha_0$  and pick  $\delta$  such that

$$\pi(\theta|\delta) \leq \alpha_0 \quad \text{for } \theta \in \Omega_0$$

That is, we put an upper bound on the probability of type I error.

## Level and size of tests

Want probability of both types of errors to be small

- Want  $\pi(\theta|\delta)$  to be small for  $\theta \in \Omega_0$  and large for  $\theta \in \Omega_1$
- Generally there is a trade-off between these probabilities
- A popular method: Choose a number  $\alpha_0$  and pick  $\delta$  such that

$$\pi(\theta|\delta) \leq \alpha_0 \quad \text{for } \theta \in \Omega_0$$

That is, we put an upper bound on the probability of type I error.

- The test is then called *level  $\alpha_0$  test* or we say that the test has *significance level  $\alpha_0$*

## Level and size of tests

Want probability of both types of errors to be small

- Want  $\pi(\theta|\delta)$  to be small for  $\theta \in \Omega_0$  and large for  $\theta \in \Omega_1$
- Generally there is a trade-off between these probabilities
- A popular method: Choose a number  $\alpha_0$  and pick  $\delta$  such that

$$\pi(\theta|\delta) \leq \alpha_0 \quad \text{for } \theta \in \Omega_0$$

That is, we put an upper bound on the probability of type I error.

- The test is then called *level  $\alpha_0$  test* or we say that the test has *significance level  $\alpha_0$*
- The *size  $\alpha(\delta)$*  of a test is defined as

$$\alpha(\delta) = \sup_{\theta \in \Omega_0} \pi(\theta|\delta)$$

# Level and size of tests

## Corollary 9.1.1: Level and size

A test  $\delta$  is a level  $\alpha_0$  test if and only if  $\alpha(\delta) \leq \alpha_0$

When the null hypothesis is simple ( $H_0 : \theta = \theta_0$ ) then  $\alpha(\delta) = \pi(\theta_0|\delta)$

Example:

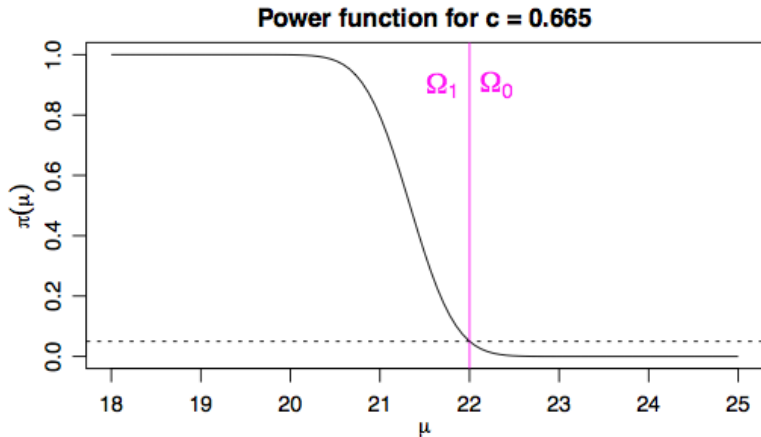
- Consider again the After-School example

We had  $X_1, \dots, X_n$  i.i.d.  $N(\mu, \sigma^2)$  and we are interested in testing

$$H_0 : \mu \geq 22 \quad \text{and} \quad H_1 : \mu < 22$$

- Again, we assume that  $\sigma^2 = 6^2$  and we reject  $H_0$  if  $\bar{X}_n - c$
- Pick a  $c$  so that the test has level 0.05
- What is the size of test?

# Power function for a size 0.05 test for after-school



## Choosing the Null and alternative hypotheses

Usually, the error of type I (rejecting  $H_0$  when it is in fact true) is more serious

After-school example:

- Type I error is deciding that  $\mu < 22$  when in fact  $\mu \geq 22$
- In that case the school district would go ahead and send stamped self-addressed envelopes to all customers and end up losing money
- The type II error here (not rejecting  $H_0$  when it is not true) represents lost opportunity of profit but at least the district will not lose (more) money.

(Do you agree that the first Type I error is more serious in this case?)



# Choosing the Null and alternative hypotheses

Therefore it makes sense to control the probability of type I error

- We usually arrange the Null and Alternative hypotheses so that the type I error is the more serious one.
- Another way to think about it: The statement we are trying to “prove” we set as the alternative hypothesis

## Bernoulli example: Prevalence of a disease

Say we are interested in the prevalence of some disease

- If the disease affects 2% or more of the population, the state will launch a costly public health campaign.
- We can examine a random sample of 80 people
- Let  $X_i$  denote whether person  $i$  has the disease.
- Assume  $X_1, \dots, X_{80}$  are i.i.d. Bernoulli( $p$ )

Which hypothesis do we test?

a)  $H_0 : p \leq 0.02$  and  $H_1 : p > 0.02$

b)  $H_0 : p \geq 0.02$  and  $H_1 : p < 0.02$

# Bernoulli example: Prevalence of a disease

Some issues to ponder:

- If we wrongly reject  $p \leq 0.02$ :

We think that the disease is more common than it is and launch a costly public health campaign that is not as needed as we thought

# Bernoulli example: Prevalence of a disease

Some issues to ponder:

- If we wrongly reject  $p \leq 0.02$ :  
We think that the disease is more common than it is and launch a costly public health campaign that is not as needed as we thought
- If we wrongly reject  $p \geq 0.02$ :  
We think that the disease is less common than it is and do not launch the campaign. This would save money in the short term but that saving could be lost in more healthcare costs. And more people might get sick that could have gotten help after the campaign.
- Which error is worse? Put that statement in the Null Hypothesis

# Bernoulli example: Prevalence of a disease

- Let  $X_i$  denote whether person  $i$  has the disease.
- Assume  $X_1, \dots, X_{80}$  are i.i.d.  $\text{Bernoulli}(p)$
- We want to test the hypothesis

$$H_0 : p \leq 0.02 \quad \text{and} \quad H_1 : p > 0.02$$

- Test: We will reject  $H_0$  if  $Y = \sum_{i=1}^{80} X_i > c$ .
- Find a constant  $c$  so that the test is a level 0.05 test
- Is the test also a size 0.05 test?

# Bernoulli example: Prevalence of a disease

- Let  $X_i$  denote whether person  $i$  has the disease.
- Assume  $X_1, \dots, X_{80}$  are i.i.d. Bernoulli( $p$ )
- We want to test the hypothesis

$$H_0 : p \leq 0.02 \quad \text{and} \quad H_1 : p > 0.02$$

- Test: We will reject  $H_0$  if  $Y = \sum_{i=1}^{80} X_i > c$ .
- Find a constant  $c$  so that the test is a level 0.05 test
- Is the test also a size 0.05 test?

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

E.g. in R: `1-pbinom(c, size=80, prob=0.02)`

# Bernoulli example: Prevalence of a disease

