# What is a testable hypothesis, anyway?

Tools for Samples Continued

Inferences Based on a Single Sample: Tests of Hypothesis

PBAF 527 Winter 2005

1



# Today

1. Tools for Samples (recap)
2. Confidence Intervals and Sample Size: Scallops, Sampling, and the Law
3. Hypothesis Testing
    1. Distinguish Types of Hypotheses
    2. Describe Hypothesis Testing Process
    3. Solve Hypothesis Testing Problems Based on a Single Sample
    4. Explain Power of a Test
    5. Explain p-Value Concept
    6. Summarize the Elements of a Hypothesis Test
    7. Things to Think about When Hypothesis Testing
    8. Special Cases
        - Small Sample Test
        - Large Sample Test of Proportions

2



# Tools for Samples

Point Estimate
- Best guess of a population parameter based upon a sample

Confidence Interval
- Range estimate around point estimate

Hypothesis Test
- Decision rule for rejecting hypothesized population values (null hypotheses)

p-value
- Continuous measure of support for null hypothesis (a probability, $\alpha$)

3

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

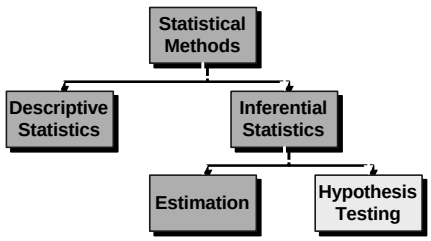
## Tools for Samples

- Point Estimate
  - Best guess of a population parameter based upon a sample
- Confidence Interval
  - Range estimate around point estimate
- Hypothesis Test
  - Decision rule for rejecting hypothesized population values (null hypotheses)
- p-value
  - Continuous measure of support for null hypothesis (a probability, $\alpha$)

4

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Statistical Methods




5

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Hypothesis Testing

6

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Hypothesis Testing

**Population**

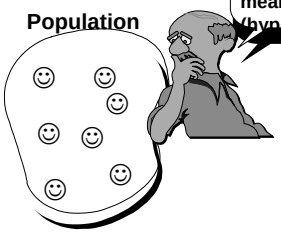
7

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Hypothesis Testing

**I believe the population mean age is 50 (hypothesis)**

**Population**



8

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

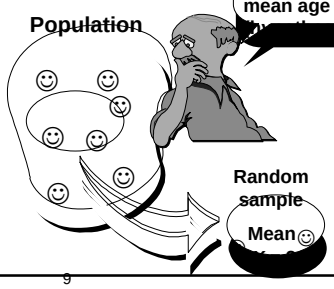
## Hypothesis Testing

**I believe the population mean age is 50**

**Population**

**Random sample**

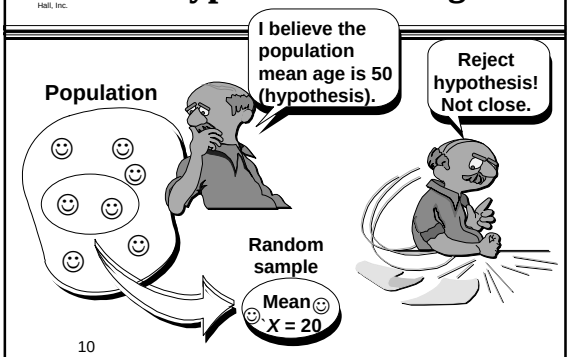
**Mean**



9

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Hypothesis Testing

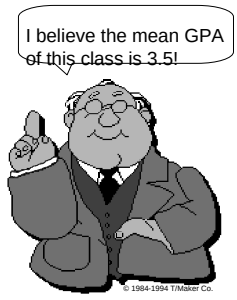



10

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## What's a Hypothesis?

1. A Belief about a Population Parameter
   - Parameter Is **Population** Mean, Proportion, Variance
   - Must Be Stated **Before** Analysis




11

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## A question, often

What is the relationship between being a women and living in poverty?

How fast is the response time to a new drug?

Is the level of pollutant emitted by an industrial facility too high?

12

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## A more specific question!

- Do single female householders in Washington State make, on average, the poverty rate for a family of three?
- Is the time to respond to a new drug 1.2 seconds?
- Is the level of pollutant emitted by an industrial facility in violation of the EPA standard of 55 ppm?

13

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Construct Hypotheses

In the case of the drug trial:

Null hypothesis: $\mu = 1.2$

- This we hold to be true until falsified

Alternative hypotheses: $\mu \neq 1.2$

- Everything but the null hypothesis

14

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Null Hypothesis

1. What Is Tested
2. Has Serious Outcome If Incorrect Decision Made
3. Always Has Equality Sign: =, £, or ³
4. Designated $H_0$
5. Specified as $H_0$: $\mu$ = Some Numeric Value
   - Specified with = Sign Even if £, or ³
   - Example, $H_0$: $\mu = 1.2$

15 Is `**x**` close to this hypothesized value?

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Alternative Hypothesis

1. Opposite of Null Hypothesis
2. Always Has Inequality Sign: ¹, <, or >
3. Designated $H_a$
4. Specified $H_a$: $\mu <$ Some Value
   - Example, $H_a$: $\mu < 3$

Everything but the value of the null hypothesis.

16

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## A hypothesis test…

- Is a decision rule for rejecting or failing to reject the null hypothesis.
- Sets significance level of cut-off ?
  - Often 5% or 1%
- Will reject null if not within CI at same confidence level
- Will reject null if p is less than ?

17

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

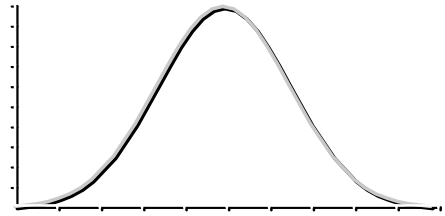
## Basic Idea

18

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Basic Idea

**Sampling Distribution**

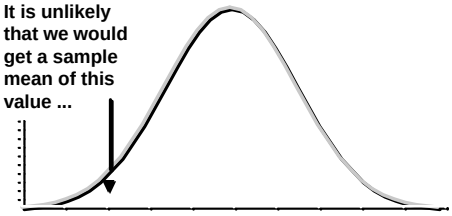



19

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Basic Idea

**Sampling Distribution**

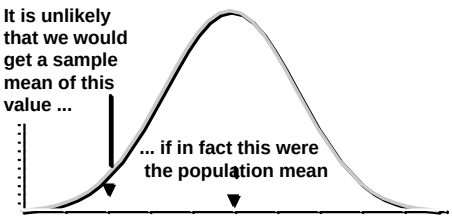



20

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Basic Idea

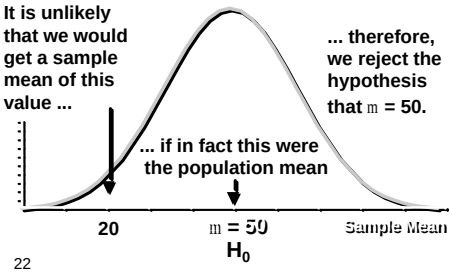
**Sampling Distribution**




21

## Basic Idea

**Sampling Distribution**




22



## 2-Tailed Test About a Mean

Is the observed value the same as the true value?

Is the time to respond to a new drug 1.2 seconds?

Suppose 100 rats were injected with the drug. The mean response time at the end of the experiments was 1.05 seconds, and the standard deviation .5 seconds.

23



## 2-Tailed Test About a Mean

1. Establish hypotheses $H_0$:$\mu$= 1.2 $H_a$: $\mu \neq 1.2$
2. Set the decision rule for the test: if $|z| > z_{\alpha/2}$ then reject the null hypothesis. Draw a picture.

- pick $\alpha$ $\alpha$=.05 (for two-sided test this is .025 in each tail)
- find $z_{\alpha/2}$ $z_{\alpha/2}$=1.96

3. Find z-statistic (like z-score) $z = \frac{\bar{x} - m}{s_{\bar{x}}} = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}} = \frac{1.05 - 1.2}{\frac{.5}{\sqrt{100}}} = \frac{-.15}{.05} = -3.0$
4. Compare test statistic to critical value.

Since $|z| > z_{\alpha/2}$ we can reject the null hypothesis at a 5% level.
The response time is not actually 1.2 seconds.

24

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## 1-Tailed Test About a Mean

Is the observed value greater then or less than the true value?

The need for action determines the alternative hypothesis. Think of this as the research hypothesis.

Is the sample mean response time of 1.05 seconds *lower than* 1.2 seconds?

25

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## 1-Tailed Test About a Mean: Left-Tailed Test

1. Establish hypotheses $H_0$:$\mu \geq 1.2$ $H_a$: $\mu < 1.2$
   - The need for action is if the observed value is less than 1.2, so that's the alternative hypothesis.
2. Set the decision rule for the test: if $z < z_\alpha$ then reject the null hypothesis. Draw a picture.
   - pick $\alpha$ $\alpha$=.05 (for one-sided test this is .5 in the tail)
   - find $z_\alpha$ $z_\alpha$=-1.645
3. Find test statistic $z = \frac{\bar{x} - m}{s_{\bar{x}}} = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}} = \frac{1.05 - 1.2}{\frac{.5}{\sqrt{100}}} = \frac{-.15}{.05} = -3.0$
4. Compare test statistic to critical value.
   Since $z < z_\alpha$ we can reject the null hypothesis at a 5% level. The response time actually less than 1.2 seconds.

26

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## 1-Tailed Test About a Mean: Right-Tailed Test

The building specifications in a certain city require that the average breaking strength of residential sewer pipe be more than 2,400 pounds per foot of length. Each manufacturer who wants to sell pipe in this city must demonstrate that its product meets the specification.

So, we want to decide whether the mean breaking strength of the pipe exceeds 2,400 pounds per linear foot.

We tested 50 sections of pipe and found the mean and standard deviations of the 50 measurements to be:

27

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## 1-Tailed Test About a Mean: Right-Tailed Test

$\bar{x}$ = 2,460 lbs for linear foot s=200 lbs per linear foot

1. Establish hypotheses
2. Set the decision rule for the test: if z>zα then reject the null hypothesis. Draw a picture.

Pick ?

Find $z_?$

3. Find test statistic $z = \frac{\bar{x} - m}{s_{\bar{x}}} = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}} = \frac{2{,}460 - 2{,}400}{\frac{200}{\sqrt{50}}} = \frac{60}{28.28} = 2.12$
4. Compare test statistic to critical value.

28

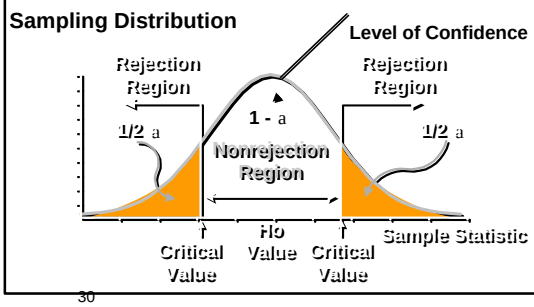
Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Rejection Regions (Two-Tailed Test)

29

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

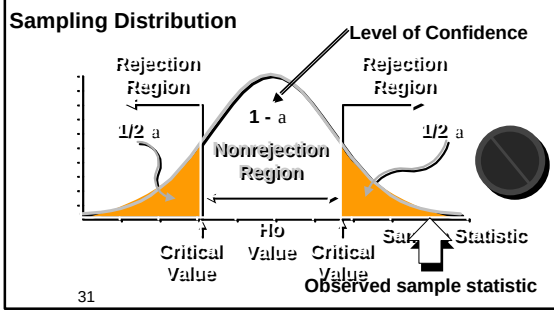
## Rejection Regions (Two-Tailed Test)




30

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

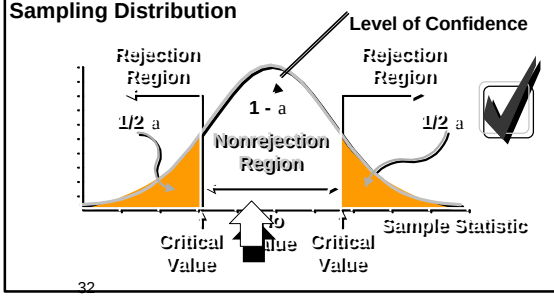
## Rejection Regions (Two-Tailed Test)




31

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

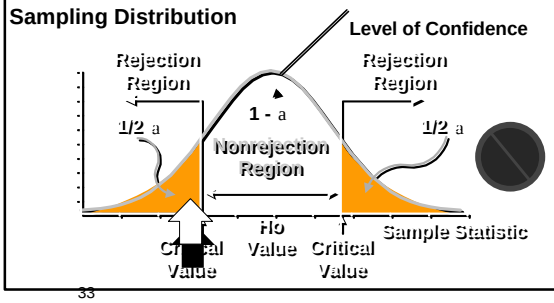
## Rejection Regions (Two-Tailed Test)




32

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Rejection Regions (Two-Tailed Test)




33

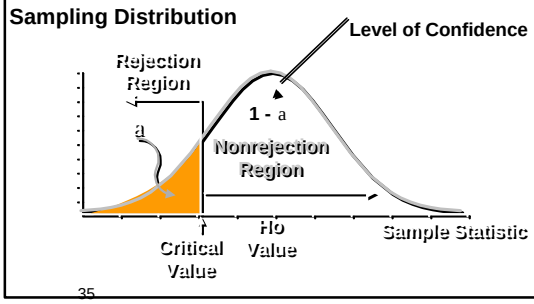
Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Rejection Region (One-Tail Test)

34

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

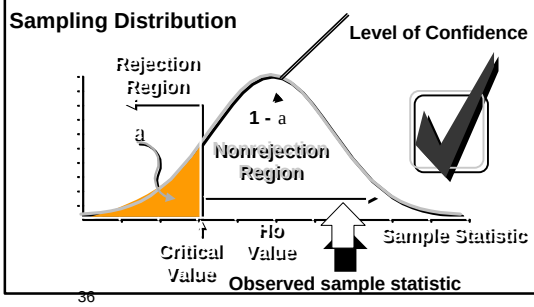
## Rejection Region (One-Tail Test)




35

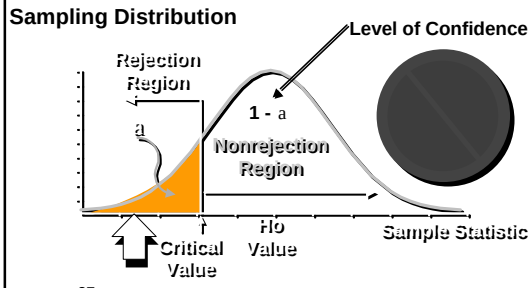
Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Rejection Region (One-Tail Test)




36

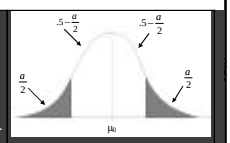
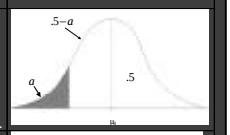
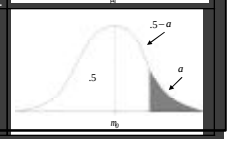
## Rejection Region (One-Tail Test)




37



Rejection Regions Summarized

| | |
|---|---|
| Form of $H_a$: $\mu \neq \mu_0$<br>2-tail hypothesis<br>Is $\lvert z \rvert > z_{\alpha/2}$?<br>Then reject the null hypothesis. |  |
| Form of $H_a$: $\mu < \mu_0$<br>1-tail hypothesis<br>Is $z < z_\alpha$?<br>Then reject the null hypothesis. |  |
| Form of $H_a$: $\mu > \mu_0$<br>1-tail hypothesis<br>Is $z > z_\alpha$?<br>Then reject the null hypothesis |  |

38



## Errors in Making Decision

1. Type I Error
   - Probability of Rejecting True Null Hypothesis
   - Has Serious Consequences
   - Probability of Type I Error Is $\alpha$ (Alpha)
     - Called Level of Significance
2. Type II Error
   - Probability of Failing to Reject a False Null Hypothesis
   - Probability of Type II Error Is $\beta$ (Beta)

39

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Decision Results

**$H_0$: Innocent**

| Jury Trial | | |
|---|---|---|
| | Actual Situation | |
| Verdict | Innocent | Guilty |
| Innocent | Correct | Error |
| Guilty | Error | Correct |

40

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Decision Results

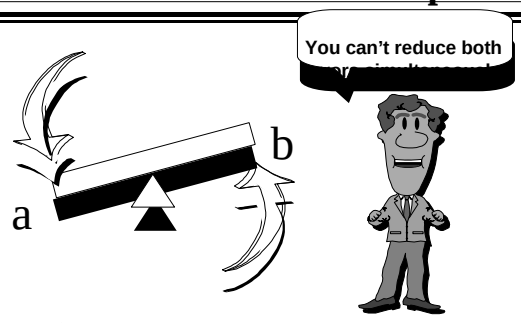
**$H_0$: Innocent**

| Jury Trial | | | $H_0$ Test | | |
|---|---|---|---|---|---|
| | Actual Situation | | | Actual Situation | |
| Verdict | Innocent | Guilty | Decision | $H_0$ True | $H_0$ False |
| Innocent | Correct | Error | Accept $H_0$ | 1 - a | Type II Error (b) |
| Guilty | Error | Correct | Reject $H_0$ | Type I Error (a) | Power (1 - b) |

41

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## a & b Have an Inverse Relationship




42

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Pipe Example, Revisited

What if the sample mean breaking strength for the pipe turned out to be 2,430 pounds, but the sample standard deviation remains 200 pounds per linear foot.

$\bar{X}$ = 2,430 lbs for linear foot s=200 lbs per linear foot

1. Establish hypotheses
2. Set the decision rule for the test: if z>zα then reject the null hypothesis. Draw a picture.
   - Pick ?
   - Find $z_?$
3. Find test statistic $z = \frac{\bar{x} - m}{s_{\bar{x}}} = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}} = \frac{2{,}430 - 2{,}400}{\frac{200}{\sqrt{50}}} = \frac{30}{28.28} = 1.06$
4. Compare test statistic to critical value.

43

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Level of Significance

1. Probability
2. Defines Unlikely Values of Sample Statistic if Null Hypothesis Is True
   - Called Rejection Region of Sampling Distribution
3. Designated α (alpha)
   - Typical Values Are .01, .05, .10
4. Selected by Researcher at Start

44

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

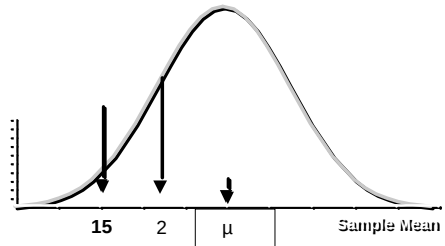
## P-Value

- Probability that you'd get a sample value this far from the mean or more *IF* the null hypothesis were true.
- A continuous measure of "strength" of evidence for null hypothesis
- Higher p value means less evidence that Ha is right (more support for Ho)
- 1-sided p values give value for one end of the distribution; 2 sided includes both.

45

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Is z=15 the same as z=2?




46

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## P-Value Example

What is the probability of getting an observed mean of 1.05 seconds to respond to the drug if the true mean is 1.2 seconds? There were 100 rats were injected with the drug, the mean response time at the end of the experiments was 1.05 seconds, and the standard deviation was .5 seconds.

47

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## P-Value Example

To find P value:

1. Set up hypotheses: $H_0$:μ= 1.2 $H_a$: μ≠1.2
2. Find estimate of mean and standard error:

$$\bar{x} = 1.05 \qquad \hat{s}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{.5}{\sqrt{100}} = .05$$

3. Find test statistic (z score):

$$z = \frac{\bar{x} - m}{s_{\bar{x}}} = \frac{1.05 - 1.2}{.05} = -3.0$$

4. Find p value (probability) associated with test statistic from t or z table (large samples, use z-table)

p=P(|z|>3|$H_0$ true)=2(.0013)=.0026 is the probability that you'd get a value of 1.05 if the actual mean were 1.2.

48

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Summary: Elements of a Hypothesis Test (1)

Null Hypothesis (H0)
- A theory about the values of one or more population parameters. The status quo.

Alternative Hypothesis (Ha)
- A theory that contradicts the null hypothesis. The theory generally represents that which we will accept only when sufficient evidence exists to establish its truth.

Test Statistic
- A sample statistic used to decide whether to reject the null hypothesis. In general,

$$\text{test statistic} = \frac{\text{Estimate - Hypothesized Parameter}}{\text{Standard Error}}$$

49

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Summary: Elements of a Hypothesis Test (2)

Critical Value
- A value to which the test statistic is compared at some particular significance level. (usually apha=.01, .05, .10)

Rejection Region
- The numerical values of the test statistic for which the null hypothesis will be rejected.
- The probability is $\alpha$ that the rejection region will contain the test statistic when the null hypothesis is true, leading to a Type I error. $\alpha$ is usually chosen to be small (.01, .05, .10) and is the level of significance of the test.

Experiment and calculation of test statistic
- Sample from the population and determine the numerical value of the test statistic.

50

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Some things to think about

1. What happens when n is large or s is small?
2. When can we automatically accept the null hypothesis?
3. What if we misspecify the null hypothesis?
4. Practical significance is not the same thing as statistical significance.
5. Always formulate $H_0$ and $H_a$ BEFORE you analyze the data. No snooping!

51

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Hypothesis Testing When n is Small and s Unknown

Because the sample is small

- Cannot assume normality
- Cannot assume s is a good approximation for s

So, use t-distribution:

$$t = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}}$$ with n-1 degrees of freedom

52

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Small Sample t-test Example 1 (1)

Most water treatment facilities monitor the quality of their drinking water on hourly basis. One variable monitored it is pH, which measures the degree of alkalinity or acidity in the water. A pH below 7.0 is acidic, one above 7.0 is alkaline, and a pH of 7.0 is neutral. One water treatment plant has a target pH of 8.5 (most try to maintain a slightly alkaline level). The mean and standard deviation of 1 hour's test results, based on 17 water samples at this plant are:

$\bar{x} = 8.24$ s=.16

Does this sample provide sufficient evidence that the mean pH level in the water differs from 8.5?

53

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Small Sample t-test Example 1 (2)

1. Establish hypotheses
2. Set the decision rule for the test:

pick $\alpha$

find $t_\alpha$ at n-1 df

3. Find test statistic $t = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}} = \frac{8.42 - 8.5}{\frac{.16}{\sqrt{17}}} = \frac{-.08}{.039} = -2.05$
4. Compare test statistic to critical value.

54

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Small Sample t-test Example 2 (1)

A major car manufacturer wants to test a new engine to determine whether it meets new air-pollution standards. The mean emission μ for all engines of this type must be less than 20 parts per million of carbon. 10 engines are manufactured for testing purposes, and the emission level for each is determined. The mean and standard deviation for the tests are:

$\bar{x} = 17.17$ s=2.98

Do the data supply enough evidence to allow the manufacturer to conclude that this type of engine meets the pollution standard? Assume the manufacturer is willing to risk a Type I error with probability α=.01.

55

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Small Sample t-test Example 2 (2)

1. Establish hypotheses
2. Set the decision rule for the test:

pick α

find $t_\alpha$

3. Find test statistic $t = \frac{\bar{x} - m}{\frac{s}{\sqrt{n}}} = \frac{17.17 - 20}{\frac{2.98}{\sqrt{10}}} = -3.00$
4. Compare test statistic to critical value.

56

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Large Sample Test for the Population Proportion

When the sample size is large (np and nq are greater than 5)

- Assume $\hat{p}$ is distributed normally with mean p and standard deviation $\sqrt{\frac{pq}{n}}$ where q=1-p
- Test statistic: $z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$
- 2- or 1-tailed tests

57

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Large Sample Tests for Proportion Example (1)

In screening women for breast cancer, doctors use a method that fails to detect cancer in 20% of the women who actually have the disease. Suppose a new method has been developed that researchers hope will detect cancer more accurately. This new method was used to screen a random sample of 140 women known to have breast cancer. Of these, the new method failed to detect cancer in 12 women.

Does this sample provide evidence that the failure rate of the new method differs from the one currently in use?

58

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Large Sample Tests for Proportion Example (2)

1. Establish hypotheses
2. Set the decision rule for the test:

pick $\alpha$

find $t_\alpha$

3. Find test statistic $z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{.086 - .2}{\sqrt{(.2)(.8)/140}} = \frac{-.114}{-.034} = -3.36$
4. Compare test statistic to critical value.

59

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## It helps to remember…

"Hypothesis testing is a ritualized exercise in devil's advocacy."

---Robert P. Abelson

60

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Scallops, Sampling, and the Law

- Read Case

a. Can a reliable estimate of the mean weight of all the scallops be obtained from a sample size of 18?

b. Do you see any flaws in the rule to confiscate a scallop catch if the sample mean weight is less than 1/36 of a pound?

c. Develop your own procedure for determining whether a ship is in violation of the weight restriction.

d. Apply your procedure to the data provided.

61

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

## Today

1. Tools for Samples (recap)
2. Confidence Intervals and Sample Size: Scallops, Sampling, and the Law
3. Hypothesis Testing
   1. Distinguish Types of Hypotheses
   2. Describe Hypothesis Testing Process
   3. Solve Hypothesis Testing Problems Based on a Single Sample
   4. Explain Power of a Test
   5. Explain p-Value Concept
   6. Summarize the Elements of a Hypothesis Test
   7. Things to Think about When Hypothesis Testing
   8. Special Cases
      - Small Sample Test
      - Large Sample Test of Proportions

62

## End of Chapter

Any blank slides that follow are blank intentionally.