Ch. 9 - Hypothesis Testing - Single Population

Ercan Karadas

New York University

ercan@nyu.edu

Summer, 2016

Concepts of Hypothesis Testing

- A hypothesis is a claim (assumption) about a population parameter. Examples:
  - The mean monthly cell phone bill of NYC is \$52, i.e. $\mu = 52$
  - Standard deviation of SAT scores of NYU students is a least 25, i.e. $\sigma > 25$.
- The Null Hypothesis, H_0 states the claim/assumption to be tested. It is always numeric.
  - In the examples above

$$H_0 : \mu = 52 \quad \text{and} \quad H_0 : \sigma > 25$$

  - It is always about a population parameter, not about a sample statistics

$$H_0 : \mu = 52, \quad \cancel{H_0 : \overline{X} = 52}$$

- The Alternative Hypothesis, H_1 is the opposite of the null hypothesis.

$$H_1 : \mu \neq 52 \quad \text{and} \quad H_1 : \sigma \leq 25$$

H_0

- Begin with the assumption that the null hypothesis is true $\rightarrow$ Similar to the notion of innocent until proven guilty
- Refers to the status quo
- Always contains "$=$", "$\leq$", or "$\geq$" sign
- May or may not be rejected

H_1

- Challenges the status quo
- Never contains the "$=$", "$\leq$", or "$\geq$" sign
- May or may not be supported
- Is generally the hypothesis that the researcher is trying to support!

Hypothesis Testing Process

Claim: **the population mean age is 50. (Null Hypothesis: H_0: μ = 50)**

Population

Now select a random sample

Is $\bar{x}$=20 likely if μ = 50?

Sample

Suppose the sample mean age is 20: $\bar{x}$ = 20

If not likely,

REJECT Null Hypothesis

The Main Idea

- Suppose you claim (or assume) that the population mean age is 50, i.e.

$$H_0 : \mu = 50$$

- From a sample of 15, you obtained $\overline{X} = 20$.
- Also suppose that population is normally distributed with the standard deviation 18.
- Let us summarize, what we know:

$$X \sim N(??, 18^2), \quad \overline{X} = 20, \quad \sigma = 18, \quad n = 15$$

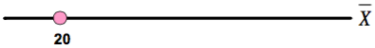
  where X is the age of an individual member of the population.
- Does the data support your initial claim? What do you think?

Some questions:

- Can you determine the distribution of $\overline{X}$?
- We know that

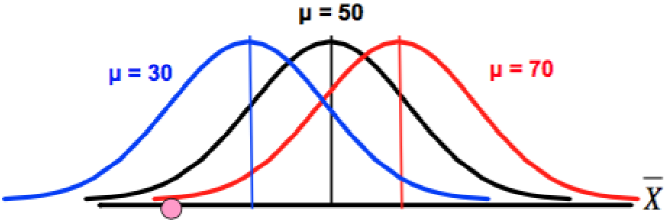
$$\overline{X} \sim N(\mu, 18^2/15)$$

- And since our data generated $\overline{X} = 20$, this should also be somewhere on the support of the distribution of $\overline{X}$, so we know at least

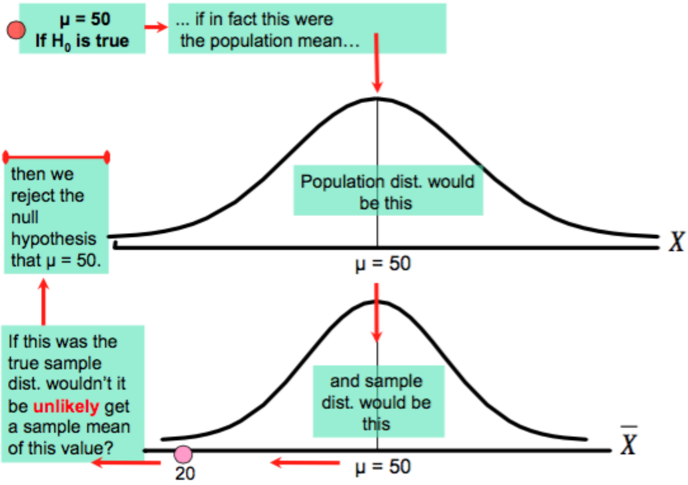


- But actually as long as we do not say something about μ we can not go any further!

- However, under different assumptions on μ we know exactly what the distribution of $\overline{X}$ would look like $\rightarrow$



Hypothesis Testing Cycle



Level of Significance, α

- Level of Significance, denoted by α defines the unlikely values of the sample static if the null hypothesis is true
- In another words, defines the rejection region of the sampling distribution
- Typical values are 0.001, 0.05, or 0.10
- It is selected by the researcher at the beginning
- Provides the critical value(s) of the test

Level of Significance and the Rejection Region

Level of significance = α

Represents critical value

Rejection region is shaded

H_0: $\mu = 3$
H_1: $\mu \neq 3$

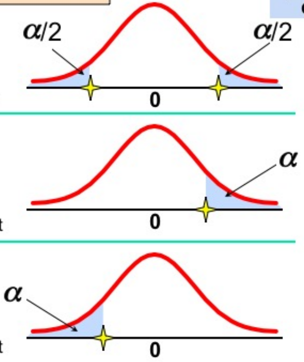
Two-tail test

H_0: $\mu \leq 3$
H_1: $\mu > 3$

Upper-tail test

H_0: $\mu \geq 3$
H_1: $\mu < 3$

Lower-tail test



Errors in Decision Making

Type I Error

- is rejecting a true null hypothesis
- Considered a serious type of error
- The probability of Type I Error is α
- It is equal to the level of significance that was set by researcher in advance

Type II Error

- is failing to reject a false hypothesis
- The probability of Type II Error is β
- $(1-\beta)$ is called the power of the test.
- We will say more about how we can calculate β and how these two type of errors are related later ...

Outcomes and Probabilities

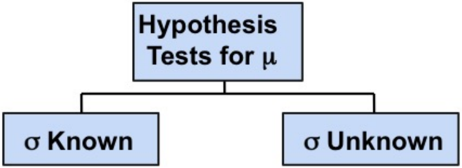
Possible Hypothesis Test Outcomes

	Actual Situation	
Decision	H_0 True	H_0 False
Fail to Reject H_0	**Correct Decision** **$(1 - \alpha)$**	**Type II Error** **(β)**
Reject H_0	**Type I Error** **(α)**	**Correct Decision** **$(1 - \beta)$**

Key:
Outcome
(Probability)

$(1 - \beta)$ is called the power of the test

Hypothesis Tests for the Mean



Tests of the Mean of a Normal Distribution (σ Known)

- Convert sample result ($\bar{x}$) to a z value

Hypothesis Tests for μ

σ Known | **σ Unknown**

Consider the test

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

(Assume the population is normal)

The decision rule is:

$$\text{Reject } H_0 \text{ if } z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_\alpha$$

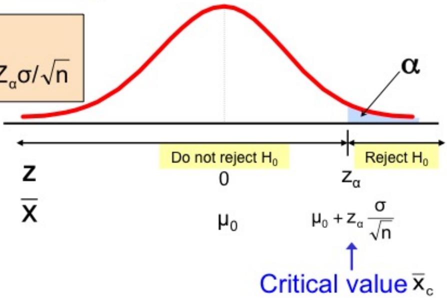
Decision Rule

Reject H_0 if $z = \dfrac{\bar{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} > z_\alpha$

H_0: $\mu = \mu_0$
H_1: $\mu > \mu_0$

Alternate rule:

Reject H_0 if $\bar{x} > \mu_0 + Z_\alpha \sigma/\sqrt{n}$



p-Value

- p-value: Probability of obtaining a test statistic more extreme (≤ or ≥) than the observed sample value given H_0 is true
  - Also called observed level of significance
  - Smallest value of α for which H_0 can be rejected

p-Value Approach to Testing

- Convert sample result (e.g., $\bar{x}$) to test statistic (e.g., z statistic)
- Obtain the p-value
  - For an upper tail test:

$$\text{p-value} = P\left(z > \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}, \text{ given that } H_0 \text{ is true}\right)$$
$$= P\left(z > \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \mid \mu = \mu_0\right)$$

- Decision rule: compare the p-value to α
  - If p-value $< \alpha$, reject H_0
  - If p-value $\geq \alpha$, do not reject H_0

Example

- A phone industry manager thinks that customer monthly cell phone bill have increased, and now average over \$52 per month. The company wishes to test this claim. (Assume $\sigma = 10$ is known).
- Suppose a sample is taken with following results:

$$\overline{X} = 53.1 \quad \text{and} \quad n = 64$$

- Form the hypothesis test:

$H_0 : \mu \leq 52$ the average is **not** over \$52 per month

$H_1 : \mu > 52$ the average is greater than \$52 per month

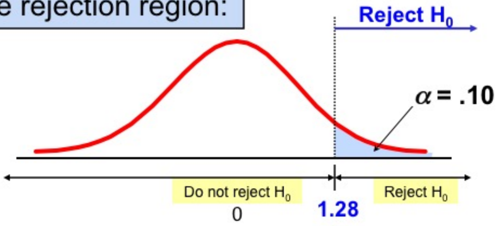
(i.e., sufficient evidence to support the claim)

Example: Find Rejection Region

(continued)

- Suppose that α = .10 is chosen for this test

Find the rejection region:



$$\text{Reject } H_0 \text{ if } \; z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > 1.28$$

Example: Sample Results

(continued)

Obtain sample and compute the test statistic

Suppose a sample is taken with the following results: $n = 64$, $\bar{x} = 53.1$ ($\sigma = 10$ was assumed known)

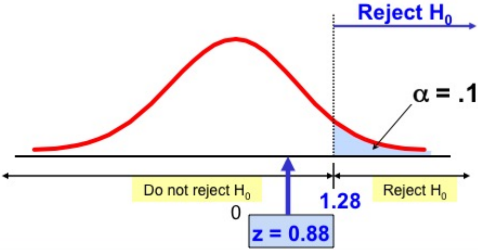
- Using the sample results,

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{53.1 - 52}{\frac{10}{\sqrt{64}}} = 0.88$$

Example: Decision

(continued)

Reach a decision and interpret the result:



Do not reject H_0 since z = 0.88 < 1.28

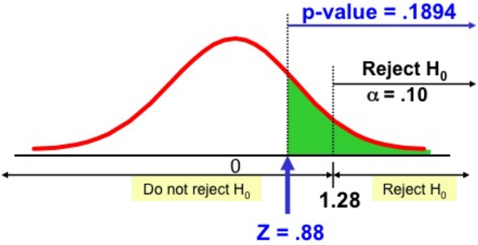
i.e.: there is not sufficient evidence that the mean bill is over $52

Example: p-Value Solution

(continued)

Calculate the p-value and compare to α

(assuming that μ = 52.0)



$$P(\bar{x} \geq 53.1 \mid \mu = 52.0)$$

$$= P\left(z \geq \frac{53.1 - 52.0}{10/\sqrt{64}}\right)$$

$$= P(z \geq 0.88) = 1 - .8106$$

$$= .1894$$

Do not reject H_0 since p-value = .1894 > α = .10

One-Tail Tests

- In many cases, the alternative hypothesis focuses on one particular direction

$H_0: \mu \leq 3$
$H_1: \mu > 3$

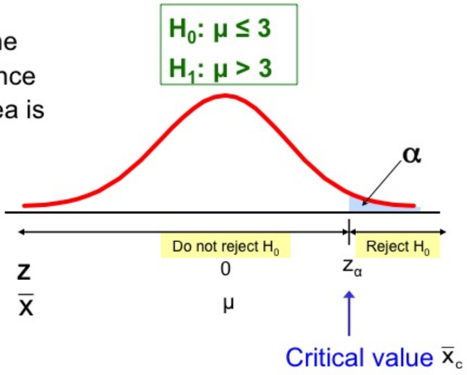
This is an upper-tail test since the alternative hypothesis is focused on the upper tail above the mean of 3

$H_0: \mu \geq 3$
$H_1: \mu < 3$

This is a lower-tail test since the alternative hypothesis is focused on the lower tail below the mean of 3

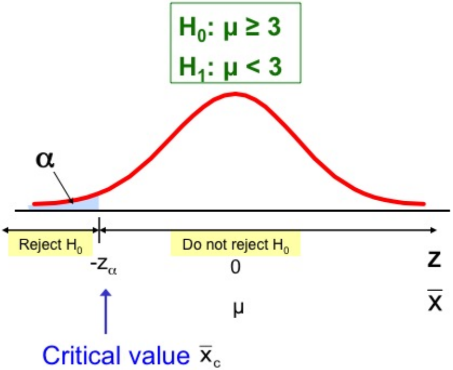
Upper-Tail Tests

- There is only one critical value, since the rejection area is in only one tail



Lower-Tail Tests

- There is only one critical value, since the rejection area is in only one tail

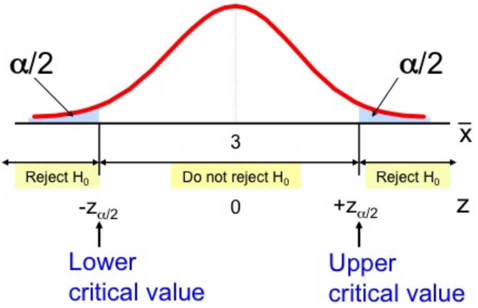


Two-Tail Tests

- In some settings, the alternative hypothesis does not specify a unique direction

H_0: $\mu = 3$
H_1: $\mu \neq 3$

- There are two critical values, defining the two regions of rejection



Hypothesis Testing Example

Test the claim that the true mean # of TV sets in US homes is equal to 3. (Assume $\sigma = 0.8$)

- State the appropriate null and alternative hypotheses
  - H_0: $\mu = 3$, H_1: $\mu \neq 3$ (This is a two tailed test)
- Specify the desired level of significance
  - Suppose that $\alpha = .05$ is chosen for this test
- Choose a sample size
  - Suppose a sample of size n = 100 is selected

Hypothesis Testing Example

(continued)

- Determine the appropriate technique
  - σ is known so this is a z test
- Set up the critical values
  - For α = .05 the critical z values are ±1.96
- Collect the data and compute the test statistic
  - Suppose the sample results are

    n = 100, $\bar{x}$ = 2.84 (σ = 0.8 is assumed known)

So the test statistic is:

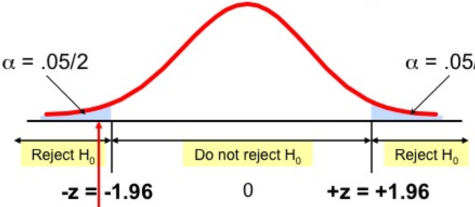
$$\boxed{z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{2.84 - 3}{\frac{0.8}{\sqrt{100}}} = \frac{-.16}{.08} = -2.0}$$

Hypothesis Testing Example

(continued)

- Is the test statistic in the rejection region?

Reject H_0 if $z < -1.96$ or $z > 1.96$; otherwise do not reject H_0

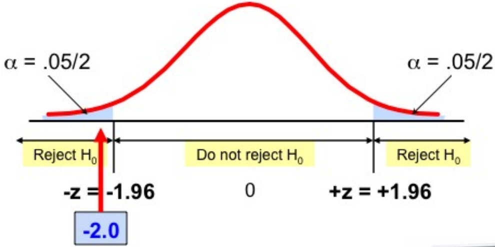


Here, $z = -2.0 < -1.96$, so the test statistic is in the rejection region

Hypothesis Testing Example

(continued)

- Reach a decision and interpret the result



Since $z = -2.0 < -1.96$, we reject the null hypothesis and conclude that there is sufficient evidence that the mean number of TVs in US homes is not equal to 3

Example: p-Value

- **Example:** How likely is it to see a sample mean of 2.84 (or something further from the mean, in either direction) if the true mean is $\mu = 3.0$?

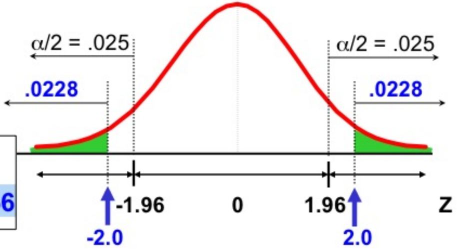
$\overline{x} = 2.84$ is translated to a z score of $z = -2.0$

$P(z < -2.0) = .0228$

$P(z > 2.0) = .0228$

p-value

$= .0228 + .0228 = .0456$



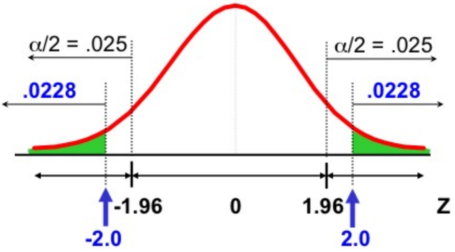
Example: p-Value

(continued)

- Compare the p-value with α
  - If p-value $<$ α, reject H_0
  - If p-value $\geq$ α, do not reject H_0

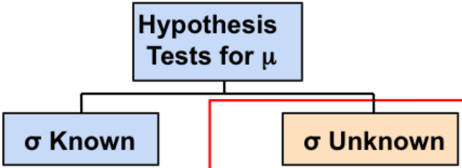
Here: p-value = .0456
α = .05

Since .0456 < .05, we reject the null hypothesis



Tests of the Mean of a Normal Population (σ Unknown)

- Convert sample result ($\bar{x}$) to a t test statistic



Consider the test

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

(Assume the population is normal)

The decision rule is:

$$\text{Reject } H_0 \text{ if } \quad t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{n-1,\alpha}$$

Tests of the Mean of a Normal Population (σ Unknown)

(continued)

- For a two-tailed test:

Consider the test

$$\boxed{\begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array}}$$

(Assume the population is normal, and the population variance is unknown)

The decision rule is:

Reject H_0 if $\boxed{t = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} < -t_{n-1, \alpha/2}}$ or if $\boxed{t = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{n-1, \alpha/2}}$

Example: Two-Tail Test (σ Unknown)

The average cost of a hotel room in Chicago is said to be $168 per night. A random sample of 25 hotels resulted in $\bar{x}$ = $172.50 and s = $15.40. Test at the α = 0.05 level.

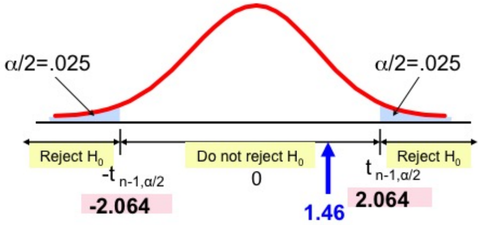
(Assume the population distribution is normal)

H_0: $\mu = 168$
H_1: $\mu \neq 168$

Example Solution: Two-Tail Test

H_0: $\mu = 168$

H_1: $\mu \neq 168$



- $\alpha = 0.05$
- $n = 25$
- σ is unknown, so use a t statistic
- Critical Value:

$t_{24\,,\,.025} = \pm\, 2.064$

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{172.50 - 168}{\frac{15.40}{\sqrt{25}}} = 1.46$$

Do not reject H_0: not sufficient evidence that true mean cost is different than $168

Assessing the Power of a Test

- Recall the possible hypothesis test outcomes:

Key: **Outcome** **(Probability)**

	Actual Situation	
Decision	H_0 True	H_0 False
Do Not Reject H_0	**Correct Decision** ($1 - \alpha$)	**Type II Error** (β)
Reject H_0	**Type I Error** (α)	**Correct Decision** ($1 - \beta$)

- β denotes the probability of Type II Error
- $1 - \beta$ is defined as the power of the test

Power = $1 - \beta$ = the probability that a false null hypothesis is rejected

Type II Error

Assume the population is normal and the population variance is known. Consider the test

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

The decision rule is:

$$\text{Reject } H_0 \text{ if } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \quad \text{or} \quad \text{Reject } H_0 \text{ if } \bar{x} > \bar{x}_c = \mu_0 + z_\alpha \sigma/\sqrt{n}$$

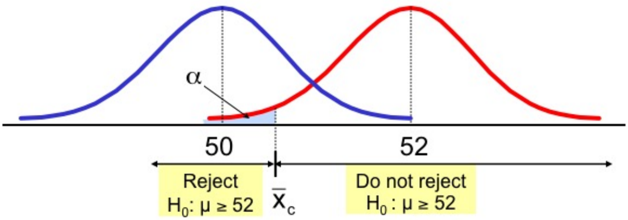
If the null hypothesis is false and the true mean is μ^*, then the probability of type II error is

$$\beta = P(\bar{x} < \bar{x}_c \mid \mu = \mu^*) = P\left(z < \frac{\bar{x}_c - \mu^*}{\sigma/\sqrt{n}}\right)$$

Type II Error Example

- Type II error is the probability of failing to reject a false H_0

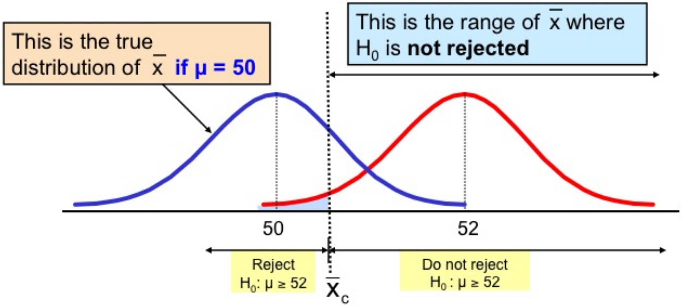
Suppose we fail to reject **H_0: $\mu \geq 52$** when in fact the true mean is **$\mu^* = 50$**



Type II Error Example

(continued)

- Suppose we do not reject H_0: $\mu \geq 52$ when in fact the true mean is $\mu^* = 50$

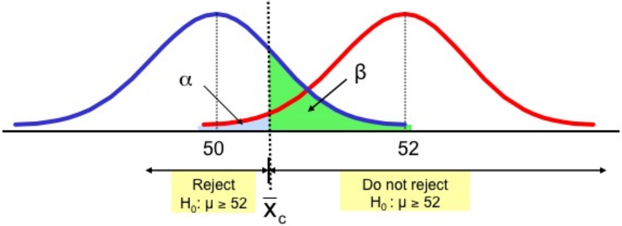


Type II Error Example

(continued)

- Suppose we do not reject H_0: $\mu \geq 52$ when in fact the true mean is $\mu^* = 50$

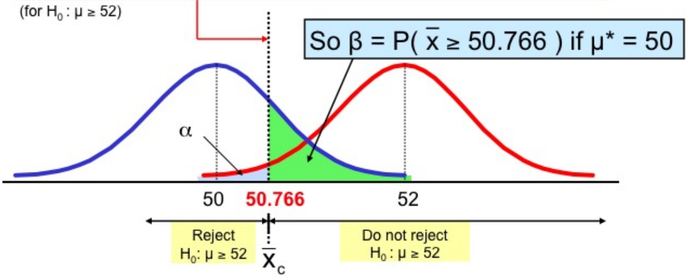
Here, $\beta = P(\bar{x} \geq \bar{x}_c)$ if $\mu^* = 50$



Calculating β

- Suppose n = 64 , σ = 6 , and α = .05

$$\bar{x}_c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = 52 - 1.645 \frac{6}{\sqrt{64}} = \boxed{50.766}$$

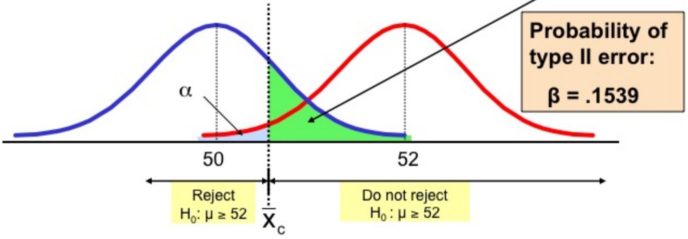


Calculating β

(continued)

- Suppose n = 64 , σ = 6 , and α = .05

$$P(\bar{x} \geq 50.766 \mid \mu^* = 50) = P\left(z \geq \frac{50.766 - 50}{6/\sqrt{64}}\right) = P(z \geq 1.02) = .5 - .3461 = \boxed{.1539}$$



Power of the Test Example

If the true mean is $\mu^* = 50$,

- The probability of Type II Error = $\beta = 0.1539$
- The power of the test = $1 - \beta = 1 - 0.1539 = 0.8461$

Key:
Outcome
(Probability)

	Actual Situation	
Decision	H_0 True	H_0 False
Do Not Reject H_0	**Correct Decision** **$1 - \alpha = 0.95$**	**Type II Error** **$\beta = 0.1539$**
Reject H_0	**Type I Error** **$\alpha = 0.05$**	**Correct Decision** **$1 - \beta = 0.8461$**

(The value of β and the power will be different for each μ^*)

Type I & II Error Relationship

- Type I and Type II errors can not happen at the same time
  - Type I error can only occur if H_0 is true
  - Type II error can only occur if H_0 is false

If Type I error probability (α) $\uparrow$, then
Type II error probability (β) $\downarrow$

Factors Affecting Type II Error

- All else equal,
  - $\beta \uparrow$ when the difference between hypothesized parameter and its true value $\downarrow$
  - $\beta \uparrow$ when $\alpha \downarrow$
  - $\beta \uparrow$ when $\sigma \uparrow$
  - $\beta \uparrow$ when $n \downarrow$

Power of the Test

- The power of a test is the probability of rejecting a null hypothesis that is false
- i.e., Power = P(Reject H_0 | H_1 is true)
  - Power of the test increases as the sample size increases

Tests of the Variance of a Normal Distribution

- Goal: Test hypotheses about the population variance, σ^2 (e.g., H_0: $\sigma^2 = \sigma_0^2$)
  - If the population is normally distributed,

$$\boxed{\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}}$$

has a chi-square distribution with (n – 1) degrees of freedom

Tests of the Variance of a Normal Distribution

(continued)

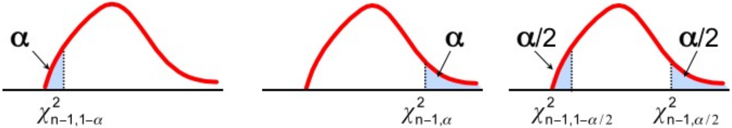
The test statistic for hypothesis tests about one population variance is

$$\boxed{\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}}$$

Decision Rules: Variance

Population variance

Lower-tail test:	Upper-tail test:	Two-tail test:
$H_0: \sigma^2 \geq \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$



Reject H_0 if $\chi^2_{n-1} < \chi^2_{n-1,1-\alpha}$

Reject H_0 if $\chi^2_{n-1} > \chi^2_{n-1,\alpha}$

Reject H_0 if $\chi^2_{n-1} > \chi^2_{n-1,\alpha/2}$
or $\chi^2_{n-1} < \chi^2_{n-1,1-\alpha/2}$

Review Exercise

A simple random sample of 25 filtered cigarettes is obtained, and the tar content of each cigarette is measured. The sample has a mean of 13.2 mg and a standard deviation of 3.7 mg. We are interested in testing the claim that the mean tar content of filtered cigarettes is less than 21.1 mg, which is the mean for unfiltered cigarettes.

1. Formulate the appropriate null and alternative hypotheses.
2. Calculate the appropriate test statistic and conclude the test at $\alpha = 0.05$?
3. Find the p-value?
4. Find the power of the test given that true population mean tar content is $\mu^* = 17.98$.
5. Test the hypothesis that population variance is equal to 8 by constructing a 90% confidence interval.

Review Exercise

A simple random sample of 25 filtered cigarettes is obtained, and the tar content of each cigarette is measured. The sample has a mean of 13.2 mg and a standard deviation of 3.7 mg. We are interested in testing the claim that the mean tar content of filtered cigarettes is less than 21.1 mg, which is the mean for unfiltered cigarettes.

1. Formulate the appropriate null and alternative hypotheses.
   $H_0 : \mu \geq 21.1$
   $H_1 : \mu < 21.1$

Review Exercise

$$n = 25, \quad \overline{X} = 13.2, \quad s = 3.7,$$

2. Calculate the appropriate test statistic and conclude the test at $\alpha = 0.05$?

   $H_0 : \mu \geq 21.1, \qquad H_1 : \mu < 21.1$

   Test statistic: $t_{24} = \frac{13.2-21.1}{3.7/5} = -10.68$

   Critical table value: $t_{24,0.05} = 1.711$

   Decision: $-10.68 < -1.711 \Longrightarrow$ reject H_0.

   $\rightarrow$ Filters are effective in reducing the amount of tar.

Review Exercise

$$n = 25, \quad \overline{X} = 13.2, \quad s = 3.7,$$

3. Find the p-value?

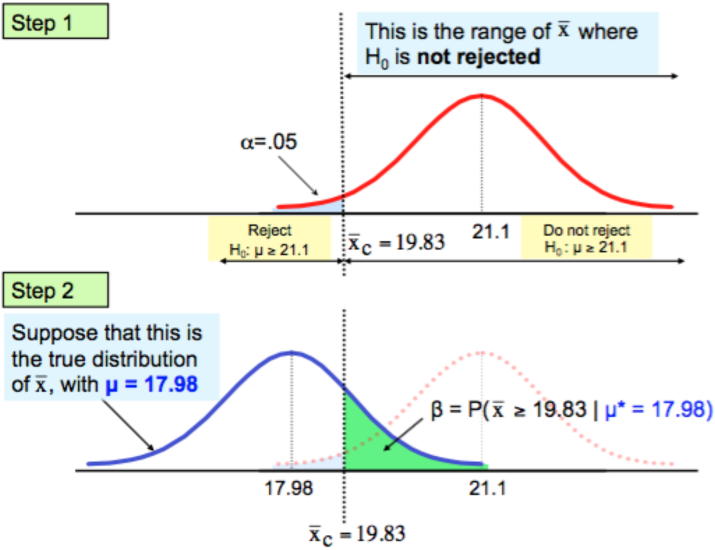
- We can not find an exact p-value from the t table
- But note that $t_{24,0.001} = 3.467$ so even with $\alpha = 0.001$ we would reject H_0
- Therefore, $p - value < 0.001$.

Review Exercise

$$n = 25, \quad \overline{X} = 13.2, \quad s = 3.7,$$

4. Find the power of the test given that true population mean tar content is $\mu^* = 17.98$?

   - Look at the figure below first ...
   - **Step 1:** For what value of $\overline{X}$, H_0 would not be rejected?
     $t_{n-1} = \frac{\overline{X}_c - \mu}{s/\sqrt{n}} > -t_{n-1,0.05} \Longrightarrow$ Fail to Reject H_0
     $\Longrightarrow \frac{\overline{X}_c - 21.1}{3.7/\sqrt{25}} = -1.711$
     $\Longrightarrow \overline{X}_c = 19.83$ (i.e. for any $\overline{X} > 19.83$, H_0 can't be rejected)
   - **Step 2:** What is the probability of observing $\overline{X} > 19.83$ if the true mean were $\mu^* = 17.89$?
     $\beta = P(\overline{X} > \overline{X}_c \mid \mu = \mu^*) =?$
     $\Longrightarrow \beta = P(t_{24} > \frac{19.83 - 17.98}{3.7/\sqrt{25}}) = P(t_{24} > 2.5) \approx 0.01.$
     $\Longrightarrow$ The power of the test is $1 - \beta = 99\%$


Step 1
This is the range of x̄ where
H0 is not rejected
α=.05
Reject
H0: μ ≥ 21.1
x̄c = 19.83
21.1
Do not reject
H0 : μ ≥ 21.1
Step 2
Suppose that this is
the true distribution
of x̄, with μ = 17.98
β = P(x̄ ≥ 19.83 | μ* = 17.98)
17.98
21.1
x̄c = 19.83

5. Test the hypothesis that population variance is equal to 8 by constructing a 90% confidence interval.

   - $H_0 : \sigma^2 = 8$
     $H_1 : \sigma^2 \neq 8$
     $\Longrightarrow$ 90% C.I. for σ^2: $\frac{(24)(3.7)^2}{\chi^2_{24,0.05}} < \sigma^2 < \frac{(24)(3.7)^2}{\chi^2_{24,0.95}}$
     $\Longrightarrow$ 90% C.I. for σ^2: $9 < \sigma^2 < 23.7$
     it doesn't contain 8, so we reject H_0 that $\sigma^2 = 8$.
   - As a last exercise, suppose we would like to test (take $\alpha = 0.1$)
     - the population variance is at least 8
       $\Longrightarrow$ no need to test this as s^2 is already larger than this (i.e. this claim is not falsifiable)
     - the population variance is at most 8
       $H_0 : \sigma^2 \leq 8$
       $H_1 : \sigma^2 > 8$
       Upper-tail test: reject if $\chi^2_{n-1} > \chi^2_{n-1,0.1}$
       $\chi^2_{24} = \frac{(n-1)s^2}{\sigma^2} = \frac{(24)(3.7^2)}{8} = \frac{(24)(3.7^2)}{8} = 41.07$,
       $\chi^2_{24,0.1} = 33.19$
       $\Longrightarrow \chi^2_{24} = 41.07 > \chi^2_{24,0.1} = 33.19$
       $\Longrightarrow$ Reject H_0
       $\Longrightarrow$ with 90% confidence we can say that σ^2 is at least 8.