Lecture 5c: Hypothesis testing: One- and two-sample cases

Ernesto F. L. Amaral

September 28–October 03, 2023
Introduction to Sociological Data Analysis (SOCI 600)

www.ernestoamaral.com

Source: Healey, Joseph F. 2015. "Statistics: A Tool for Social Research." Stamford: Cengage Learning. 10th edition. Chapters 8 (pp. 185–215) and 9 (pp. 216–246).

TEXAS A&M UNIVERSITY

Outline

- Hypothesis testing
  - One-sample case
  - Two-sample case

One-sample case

- Explain the logic of hypothesis testing, including concepts of the null hypothesis, the sampling distribution, the alpha level, and the test statistic
- Explain what it means to “reject the null hypothesis” or “do not reject the null hypothesis”
- Identify and cite examples of situations in which one-sample tests of hypotheses are appropriate
- Test the significance of single-sample means and proportions using the five-step model, and correctly interpret the results
- Explain the difference between one- and two-tailed tests, and specify when each is appropriate
- Define and explain Type I and Type II errors, and relate each to the selection of an alpha level
- Use the Student’s *t* distribution to test the significance of a sample mean for a small sample

Significant differences

- Hypothesis testing is designed to detect significant differences
  - Differences that did not occur by random chance
  - Hypothesis testing is also called significance testing
- This chapter focuses on the “one sample” case
  - Compare a random sample against a population
  - Compare a sample statistic to a (hypothesized) population parameter to see if there is a statistically significant difference

Example 1: Question

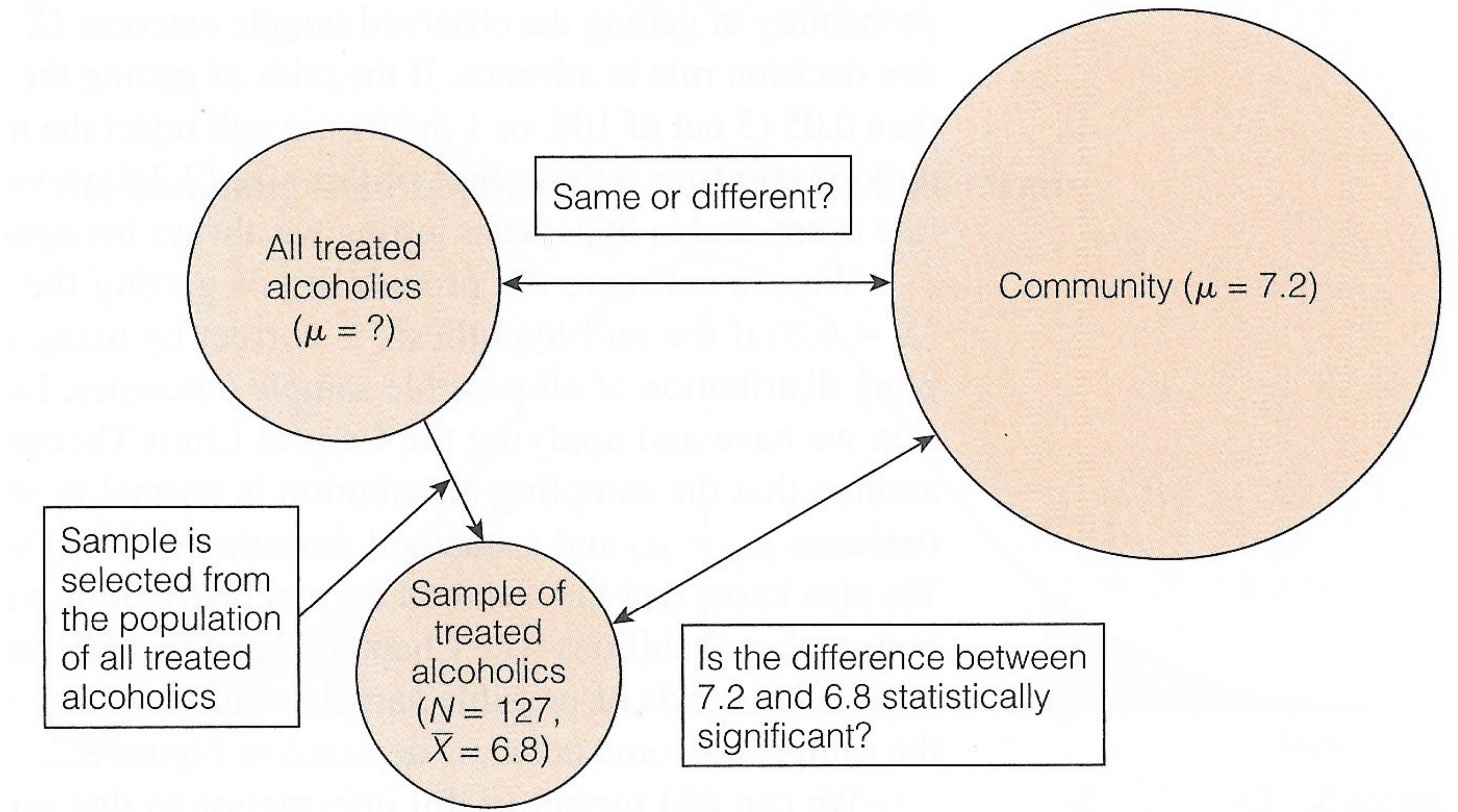
- Are people who have been treated for alcoholism more reliable workers than those in the community?
  - Does the group of all treated alcoholics have different absentee rates than the community as a whole?
  - Effectiveness of rehabilitation center for alcoholics
- Absentee rates for community and sample
  - Don't have resources to gather information of all people who have been treated by the program

Community	Sample of treated alcoholics
$\mu = 7.2\ days\ per\ year$	$\bar{X} = 6.8\ days\ per\ year$
$\sigma = 1.43$	$n = 127$

- What causes the difference between 7.2 and 6.8?
  - Real difference? Or difference due to random chance?

Source: Healey 2015, p.187.

A test of hypothesis for single-sample means

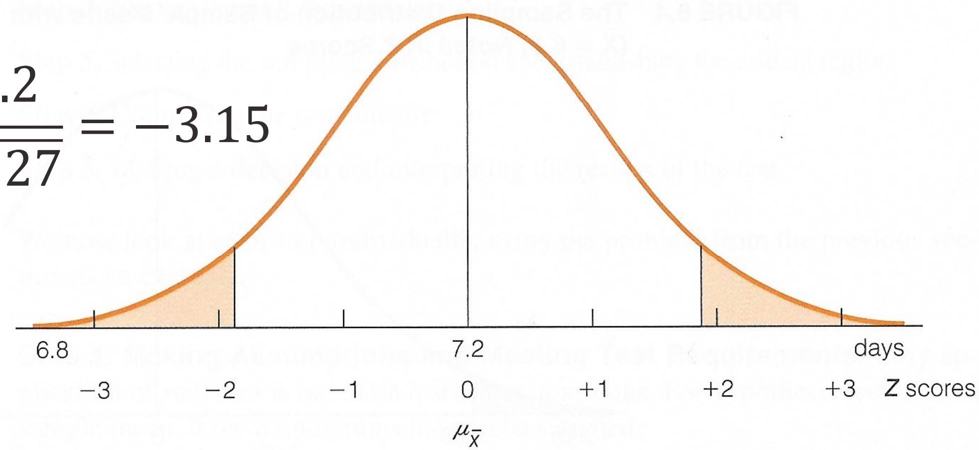


Source: Healey 2015, p.187.

Example 1: Result

- For a known/empirical distribution, we use: $Z = \frac{X_i - \bar{X}}{s}$
- However, we are concerned with the sampling distribution of all possible sample means

$$Z(obtained) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{6.8 - 7.2}{1.43/\sqrt{127}} = -3.15$$



- The sample outcome falls in the shaded area
  - *Z(obtained)* = –3.15
  - Reject H_0: μ = 7.2 days per year
  - The sample of 127 treated alcoholics comes from a population that is significantly different from the community on absenteeism

Source: Healey 2015, p.189.

The five-step model

1. Make assumptions and meet test requirements
2. Define the null hypothesis (H_0)
3. Select the sampling distribution and establish the critical region
4. Compute the test statistic
5. Make a decision and interpret the test results

Example 2: Question

- The education department at a university has been accused of "grade inflation"
  - Thus, education majors have much higher GPAs than students in general
- GPAs of all education majors should be compared with the GPAs of all students
  - There are 1000s of education majors, far too many to interview
  - How can the dispute be investigated without interviewing all education majors?

Example 2: Numbers

- The average GPA for all students is 2.70 (μ)
  - This value is a parameter
- Random sample of education majors
  - Mean = $\bar{X}$ = 3.00
  - Standard deviation = s = 0.70
  - Sample size = n = 117
- There is a difference between parameter (μ=2.70) and statistic ($\bar{X}$=3.00)
  - It seems that education majors do have higher GPAs

Example 2: Explanations

- We are working with a random sample
  - Not all education majors

- Two explanations for the difference

1. The sample mean ($\bar{X}$=3.00) is the same as the population mean (μ=2.70)
   - The observed difference may have been caused by random chance

2. The difference is real (statistically significant)
   - Education majors are different from all students

Step 1: Assumptions,requirements

- Make assumptions
  - Random sampling
  - Hypothesis testing assumes samples were selected according to EPSEM
- Meet test requirements
  - The sample of 117 was randomly selected from all education majors
  - Level of measurement is interval-ratio
    - GPA is an interval-ratio level variable, so the mean is an appropriate statistic
  - Sampling distribution is normal in shape
    - This is a large sample ($n \geq 100$)

Step 2: Null hypothesis

- Null hypothesis, H_0: $\mu = 2.7$
  - H_0 always states there is no significant difference
  - The sample of 117 comes from a population that has a GPA of 2.7
  - The difference between 2.7 and 3.0 is trivial and caused by random chance
- Alternative hypothesis, H_1: $\mu \neq 2.7$
  - H_1 always contradicts H_0
  - The sample of 117 comes from a population that does not have a GPA of 2.7
  - There is an actual difference between education majors ($\bar{X}$=3.0) and other students (μ=2.7)

Step 3: Distribution, critical region

- Sampling distribution: standard normal shape
  - Alpha (α) = 0.05
  - Use the 0.05 value as a guideline to identify differences that would be rare if H_0 is true
  - Any difference with a probability less than α is rare and will cause us to reject the H_0
- Use the *Z* score to determine the probability of getting the observed difference
  - If the probability is less than 0.05, the obtained *Z* score will be beyond the critical *Z* score of ±1.96
  - This is the critical *Z* score associated with a two-tailed test and α=0.05

Step 4: Test statistic

- For a known/empirical distribution, we would use

$$Z = \frac{X_i - \bar{X}}{s}$$

- However, we are concerned with the sampling distribution of all sample means

- We only have the standard deviation for the sample (s), not for the population (σ)

$$Z(obtained) = \frac{\bar{X} - \mu}{s/\sqrt{n-1}} = \frac{3.0 - 2.7}{0.7/\sqrt{117-1}} = 4.62$$

Step 5: Decision, interpret

- *Z(obtained)* = 4.62
  - This is beyond *Z(critical)* = ±1.96
  - The obtained *Z* score fell in the critical region, so we ***reject*** the H_0
  - If H_0 was true, a sample GPA of 3.0 would be unlikely
  - Therefore, the H_0 is false and must be rejected
- Education majors have a GPA that is significantly higher than general student body
  - The difference between the parameter (μ=2.7) and the statistic ($\bar{X}$=3.0) was large and unlikely to have occurred by random chance (p<0.05)

Five-step model summary

Situation	Decision	Interpretation
The test statistic is in the critical region	Reject the null hypothesis (H_0)	The difference is statistically significant
The test statistic is not in the critical region	Do not reject the null hypothesis (H_0)	The difference is not statistically significant

- Model is fairly rigid, but there are two crucial choices
  - One-tailed or two-tailed test
  - Alpha (α) level

Source: Healey 2015, p.194.

One or two-tailed test

- Null hypothesis always has the equal sign

$$H_0: \mu = 2.7$$

- Two-tailed test states that population mean is not equal to the value stated in null hypothesis

$$H_1: \mu \neq 2.7$$

- One-tailed test estimates differences in a specific direction (based on theory)

$$H_1: \mu > 2.7$$

$$H_1: \mu < 2.7$$

One or two-tailed test

One- vs. Two-Tailed Tests, $\alpha = 0.05$

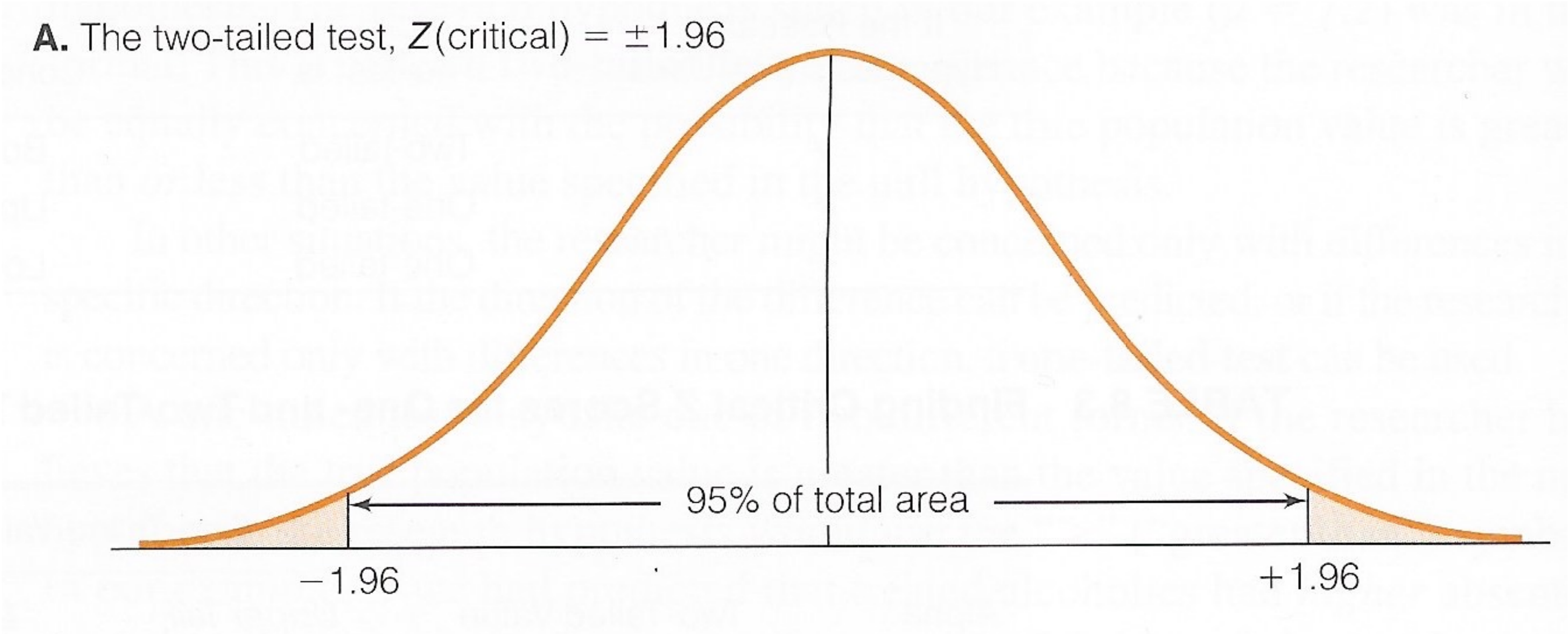
If the Research Hypothesis (H_1) Uses	The Test Is	Concern Is on	*Z*(critical) Is
$\neq$	Two-tailed	Both tails	±1.96
$>$	One-tailed	Upper tail	+1.65
$<$	One-tailed	Lower tail	−1.65

Finding Critical *Z* Scores for One- and Two-Tailed Tests

		One-Tailed Value	
Alpha	Two-Tailed Value	*Upper Tail*	*Lower Tail*
0.10	±1.65	+1.29	−1.29
0.05	±1.96	+1.65	−1.65
0.01	±2.58	+2.33	−2.33
0.001	±3.32	+3.10	−3.10
0.0001	±3.90	+3.70	−3.70

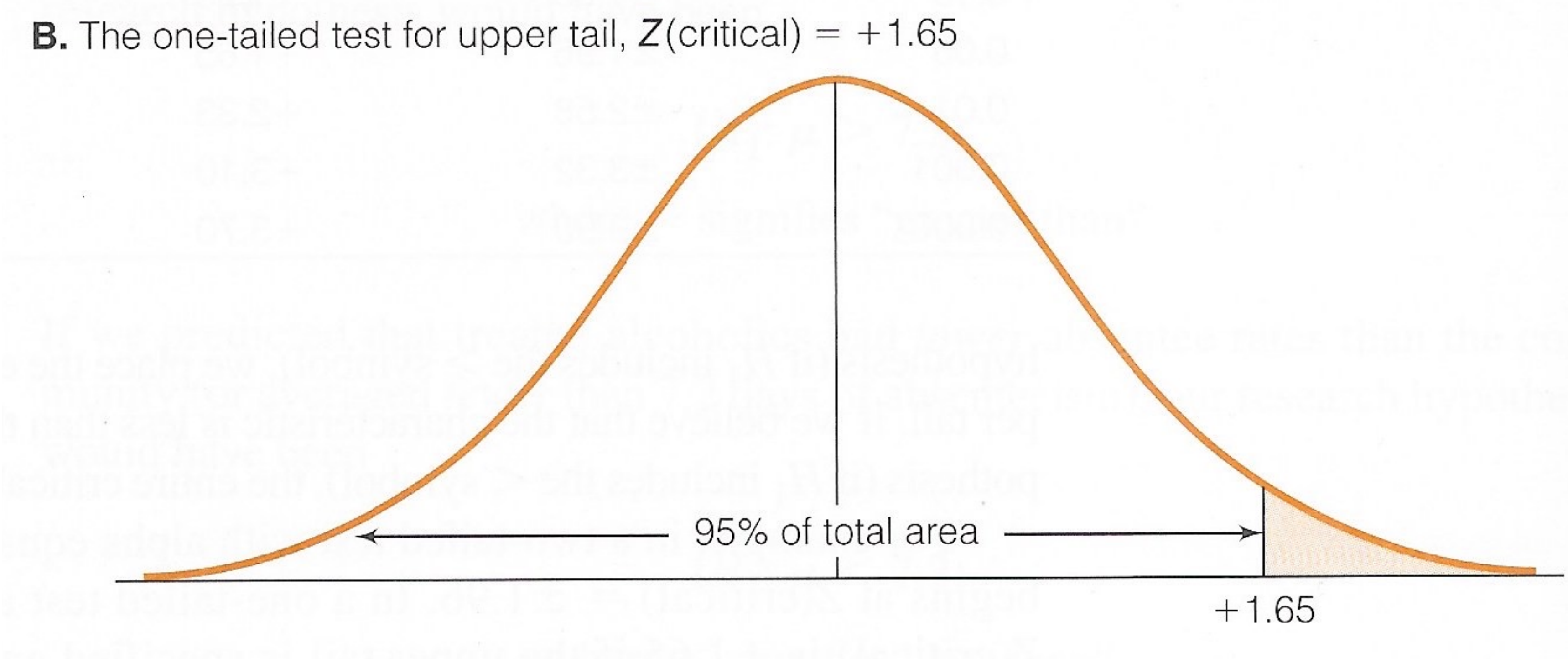
Source: Healey 2015, p.197.

Two-tailed test: α=0.05



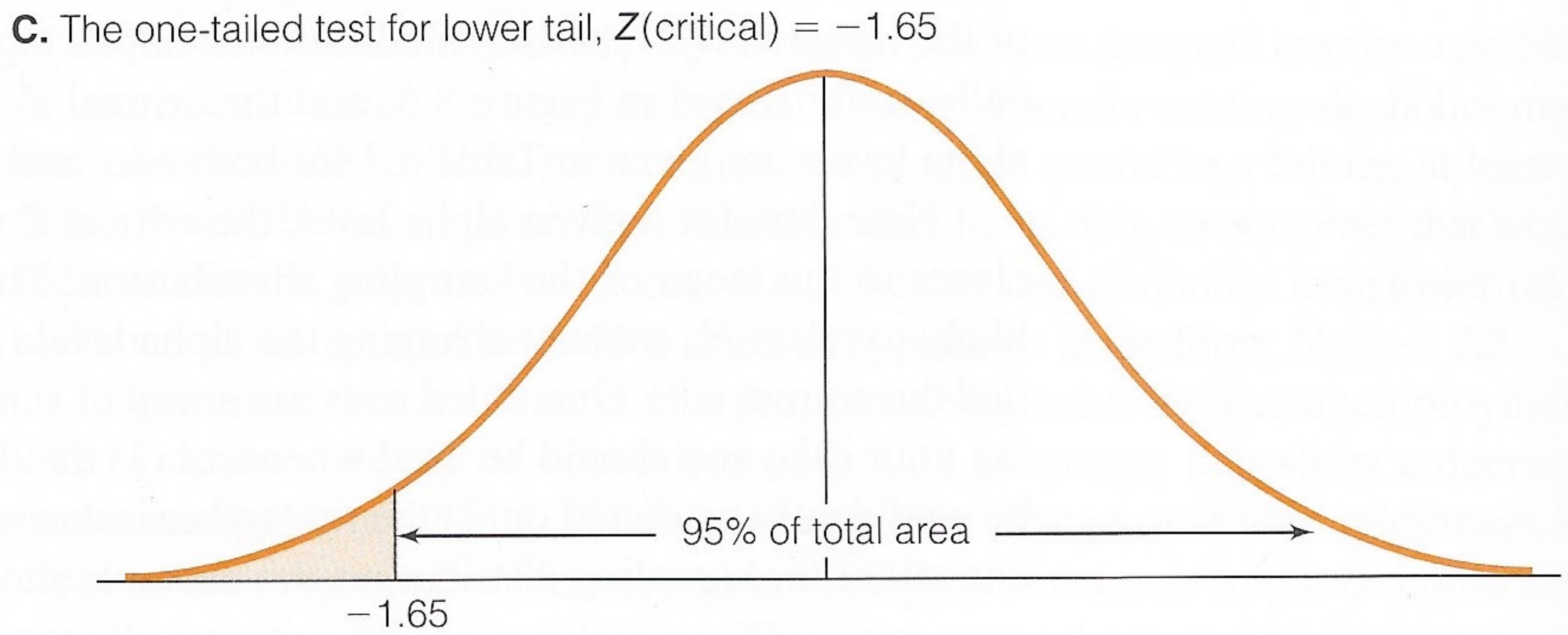
Source: Healey 2015, p.198.

One-tailed test (upper): α=0.05



Source: Healey 2015, p.198.

One-tailed test (lower): α=0.05



Source: Healey 2015, p.198.

Selecting an alpha level

- By assigning an alpha level, one defines an "unlikely" sample outcome
- Alpha level is the probability that the decision to reject the null hypothesis is incorrect
- Examine this table for critical regions

The Relationship Between Alpha and *Z*(Critical) for a Two-Tailed Test

If Alpha =	The Two-Tailed Critical Region Will Begin at *Z*(Critical) =
0.100	±1.65
0.050	±1.96
0.010	±2.58
0.001	±3.32

Source: Healey 2015, p.201.

Type I and Type II errors

- Type I error (alpha error)
  - Rejecting a true null hypothesis
- Type II error (beta error)
  - Not rejecting a false null hypothesis
- Examine table below for relationships between decision making and errors

Decision Making and the Five-Step Model

	If Our Decision Is to	And H_0 Is Actually	The Result Is
a	Reject H_0	False	OK
b	Fail to reject H_0	True	OK
c	Reject H_0	True	Type I or alpha (α) error
d	Fail to reject H_0	False	Type II or beta (β) error

Source: Healey 2015, p.201.

Decisions about hypotheses

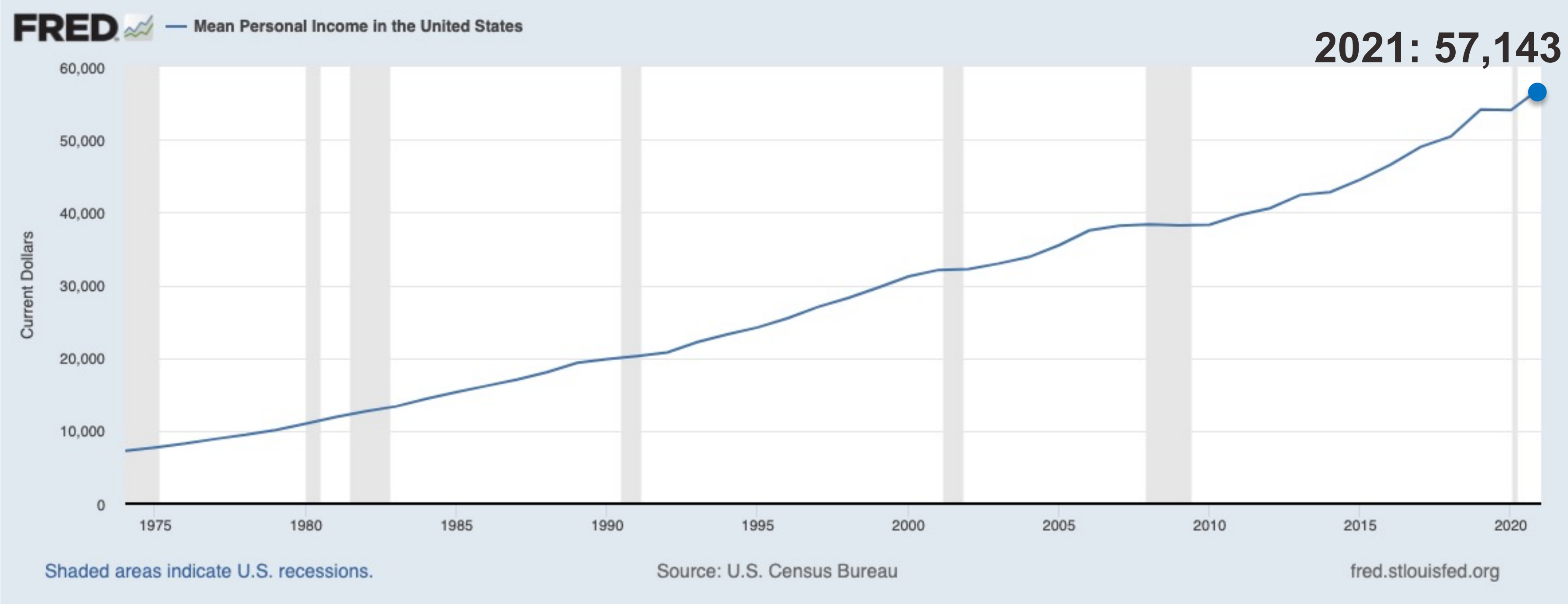
Hypotheses	***p* < *α***	***p* > *α***
Null hypothesis (H_0)	Reject	Do not reject
Alternative hypothesis (H_1)	Accept	Do not accept

- ***p*-value** is the probability of not rejecting the null hypothesis
- If a statistical software gives only the two-tailed *p*-value, divide it by 2 to obtain the one-tailed *p*-value

Significance level (*α*)	**Confidence level**
0.10 (10%)	90%
0.05 (5%)	95%
0.01 (1%)	99%
0.001 (0.1%)	99.9%

Example 3: Income, 2021

- Is the mean personal income of Veterans (GSS) lower than mean income of population 15+ (Census Bureau)?
- We know the income for the **population 15+**



Source: U.S. Census Bureau, Mean Personal Income in the United States [MAPAINUSA646N], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MAPAINUSA646N, October 24, 2022. Shaded areas indicate U.S. recessions.

Example 3: Census & GSS

- We know the income for the **2021 GSS sample of Veterans**

```
. mean conrinc if veteran==1

Mean estimation                              Number of obs = 229
```

	Mean	Std. err.	[95% conf.	interval]
conrinc	49562.49	2932.717	43783.8	55341.19

- What causes the difference between $57,143.00 (pop.15+, Census) and $49,562.49 (Veterans, GSS)?
- Real difference? Or difference due to random chance?

Example 3: Result

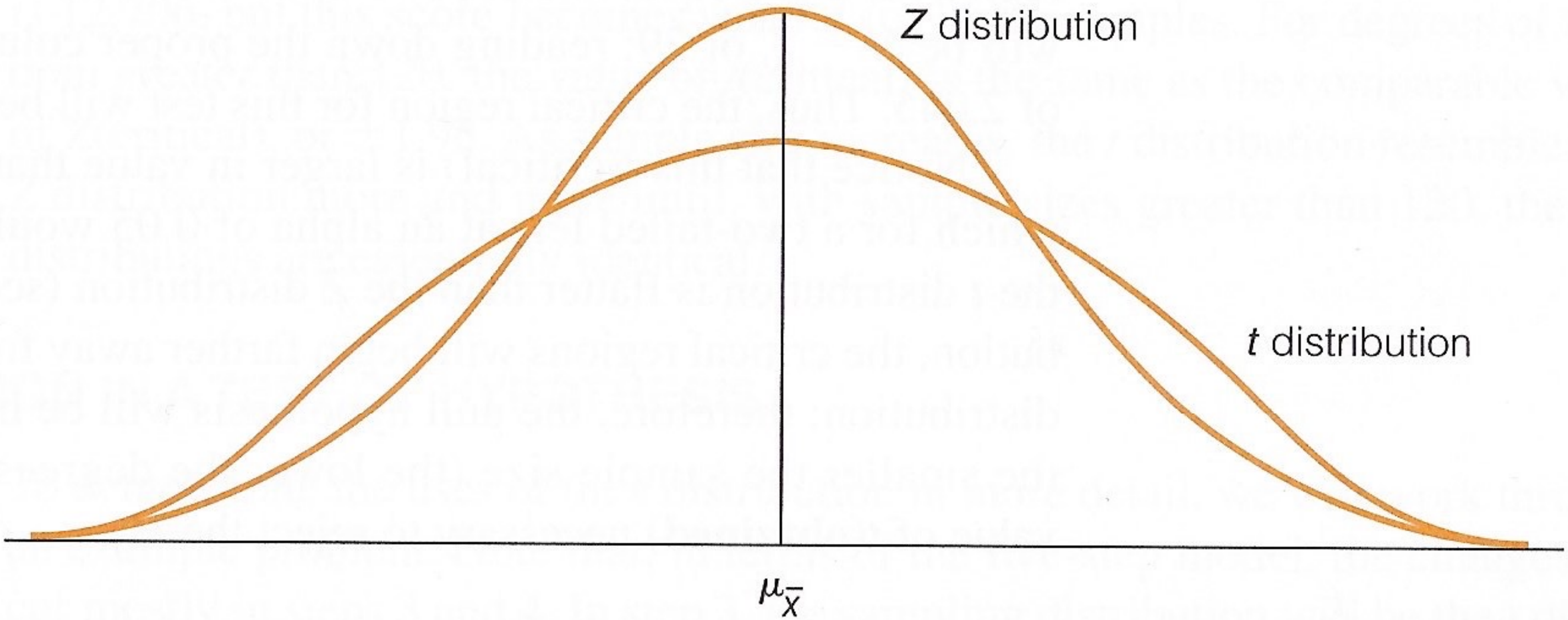
- Veteran population has mean income that is significantly lower than mean income of the population 15+
  - The difference between the parameter $57,143.00 and the statistic $49,562.49 was large and unlikely to have occurred by random chance (*p*-value<0.05)

```
. ztest conrinc=57143 if veteran==1

One-sample z test

Variable |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+---------------------------------------------------------------------
 conrinc |     229    49562.49    .0660819           1    49562.36    49562.62
---------+---------------------------------------------------------------------
    mean = mean(conrinc)                                          z = -1.1e+05
H0: mean = 57143

   Ha: mean < 57143           Ha: mean != 57143            Ha: mean > 57143
 Pr(Z < z) = 0.0000       Pr(|Z| > |z|) = 0.0000        Pr(Z > z) = 1.0000
```

The Student's *t* distribution

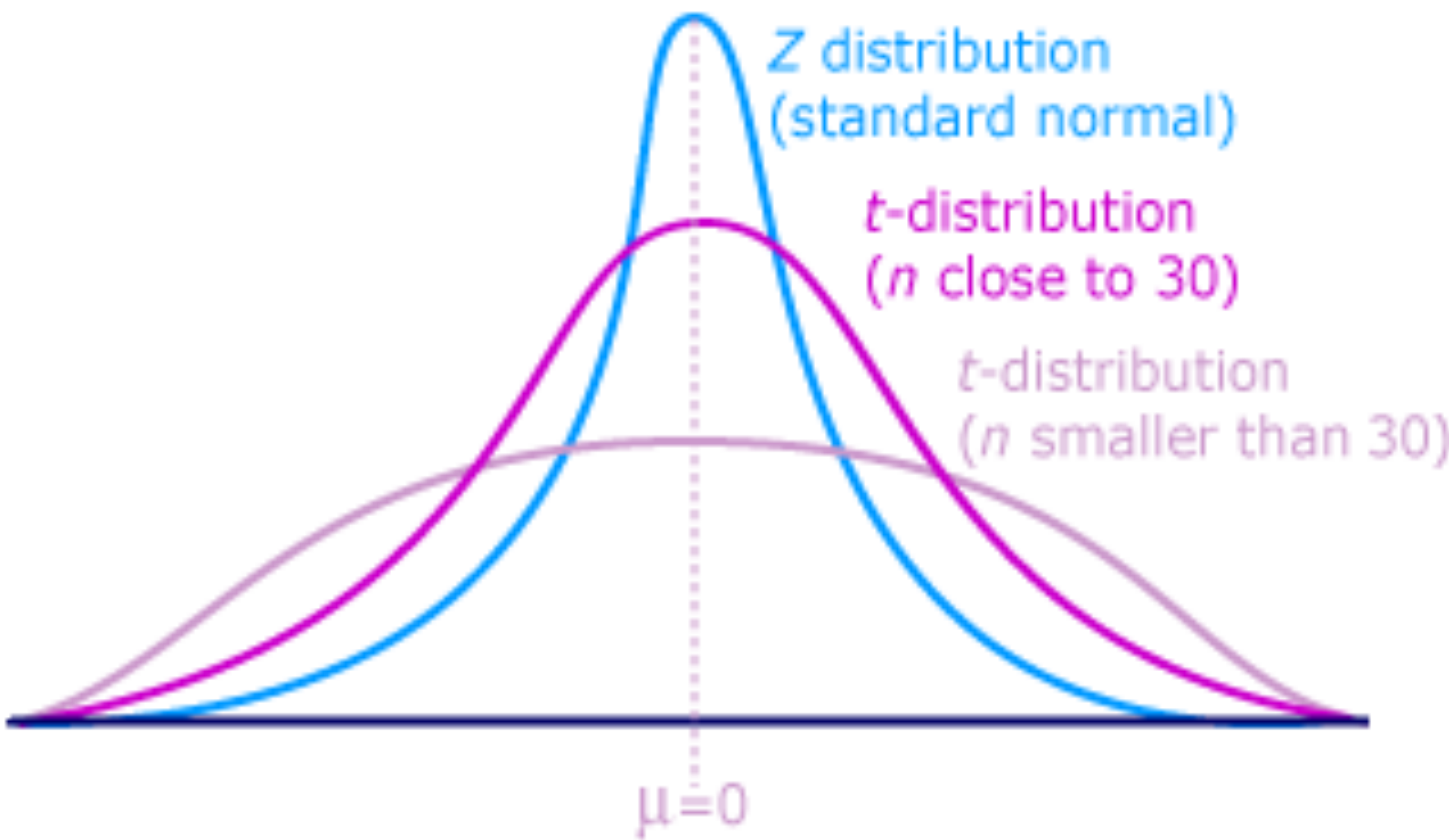
- How can we test a hypothesis when the population standard deviation (σ) is unknown, as is usually the case?
- For large samples ($n \geq 100$), we can use the sample standard deviation (s) as an estimator of the population standard deviation (σ)
  - Use standard normal distribution (Z)
- For small samples, s is too biased to estimate σ
  - Do not use standard normal distribution
  - Use Student's *t* distribution

t and *Z* distributions



Source: Healey 2015, p.203.

t and *Z* distributions



Source: https://joejeong33.wordpress.com/2013/06/03/t-distributionin-the-normal-distribution-there-are-enough/.

Choosing the distribution

- Choosing a sampling distribution when testing single-sample means for significance

If population standard deviation (σ) is	Sampling distribution is the
Known	*Z* distribution
Unknown and sample size (*n*) is large	*Z* distribution
Unknown and sample size (*n*) is small	*t* distribution

Source: Healey 2015, p.207.

Example 4: With *t*-test

- This is the same as example 3, but with *t*-test
  - GSS has a large sample. This is just an illustration
- Veteran population has mean income that is significantly lower than mean income of the population 15+ (p-value<0.05)

```
. ttest conrinc=57143 if veteran==1

One-sample t test
```

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
conrinc	229	49562.49	2932.717	44380.07	43783.8	55341.19

```
    mean = mean(conrinc)                                          t =  -2.5848
H0: mean = 57143                             Degrees of freedom =      228

   Ha: mean < 57143           Ha: mean != 57143           Ha: mean > 57143
 Pr(T < t) = 0.0052       Pr(|T| > |t|) = 0.0104       Pr(T > t) = 0.9948
```

Five-step model for proportions

- When analyzing variables that are not measured at the interval-ratio level
  - A mean is inappropriate
  - We can test a hypothesis on a one sample proportion
- The five step model remains primarily the same, with the following changes
  - The assumptions are: random sampling, nominal level of measurement, and normal sampling distribution
  - The formula for *Z* is

$$Z = \frac{P_s - P_u}{\sqrt{P_u(1 - P_u)/n}}$$

Example 5: Proportions

- A random sample of 122 households in a low-income neighborhood revealed that 53 of the households were headed by women
  - P_s = 53 / 122 = 0.43
- In the city as a whole, the proportion of women-headed households (P_u) is 0.39
- Are households in lower-income neighborhoods significantly different from the city as a whole?
- Conduct a 90% hypothesis test (α = 0.10)

Step 1: Assumptions,requirements

- Make assumptions
  - Random sampling
  - Hypothesis testing assumes samples were selected according to EPSEM
- Meet test requirements
  - The sample of 122 was randomly selected from all lower-income neighborhoods
  - Level of measurement is nominal
    - Women-headed households is measured as a proportion
  - Sampling distribution is normal in shape
    - This is a large sample ($n \geq 100$)

Step 2: Null hypothesis

- Null hypothesis, H_0: $P_u = 0.39$
  - The sample of 122 comes from a population where 39% of households are headed by women
  - The difference between 0.43 and 0.39 is trivial and caused by random chance
- Alternative hypothesis, H_1: $P_u \neq 0.39$
  - The sample of 122 comes from a population where the percent of women-headed households is not 39
  - The difference between 0.43 and 0.39 reflects an actual difference between lower-income neighborhoods and all neighborhoods

Step 3: Distribution, critical region

- Sampling distribution
  - Standard normal distribution (*Z*)

- Alpha (α) = 0.10 (two-tailed)

- Critical region begins at *Z(critical)* = ±1.65
  - This is the critical *Z* score associated with a two-tailed test and alpha equal to 0.10
  - If the obtained *Z* score falls in the critical region, we reject H_0

Step 4: Test statistic

- Proportion of households headed by women

City	Sample in a low-income neighborhood
$P_u = 0.39$	$P_s = 0.43$
	$n = 122$

- The formula for Z is

$$Z = \frac{P_s - P_u}{\sqrt{P_u(1 - P_u)/n}} = \frac{0.43 - 0.39}{\sqrt{0.39(1 - 0.39)/122}} = 0.91$$

Step 5: Decision, interpret

- *Z(obtained)* = 0.91
  - *Z(obtained)* did not fall in the critical region delimited by *Z(critical)* = ±1.65, so we ***do not reject*** the H_0
  - This means that if H_0 was true, a sample outcome of 0.43 would be likely
  - Therefore, the H_0 is not false and cannot be rejected
- The population of women-headed households in lower-income neighborhoods is not significantly different from the city as a whole
  - The difference between the parameter (P_u=0.39) and the statistic (P_s=0.43) was small and likely to have occurred by random chance (p>0.10)

Example 6: Sex, 2021

- Is the female proportion of the adult population (18+) in the U.S. higher than among the total population?
- We know the percentage of women for the **population**

PEOPLE	
Population	
Population Estimates, July 1 2021, (V2021)	**331,893,745**
Population estimates base, April 1, 2020, (V2021)	331,449,281
Population, percent change - April 1, 2020 (estimates base) to July 1, 2021, (V2021)	0.1%
Population, Census, April 1, 2020	331,449,281
Population, Census, April 1, 2010	308,745,538
Age and Sex	
Persons under 5 years, percent	5.7%
Persons under 18 years, percent	22.2%
Persons 65 years and over, percent	16.8%
Female persons, percent	50.5%

Source: U.S. Census Bureau (https://www.census.gov/quickfacts/fact/table/US/PST045221).

Example 6: Census & GSS

- The percentage of women in the **2021 GSS sample 18+**

```
. tab female
```

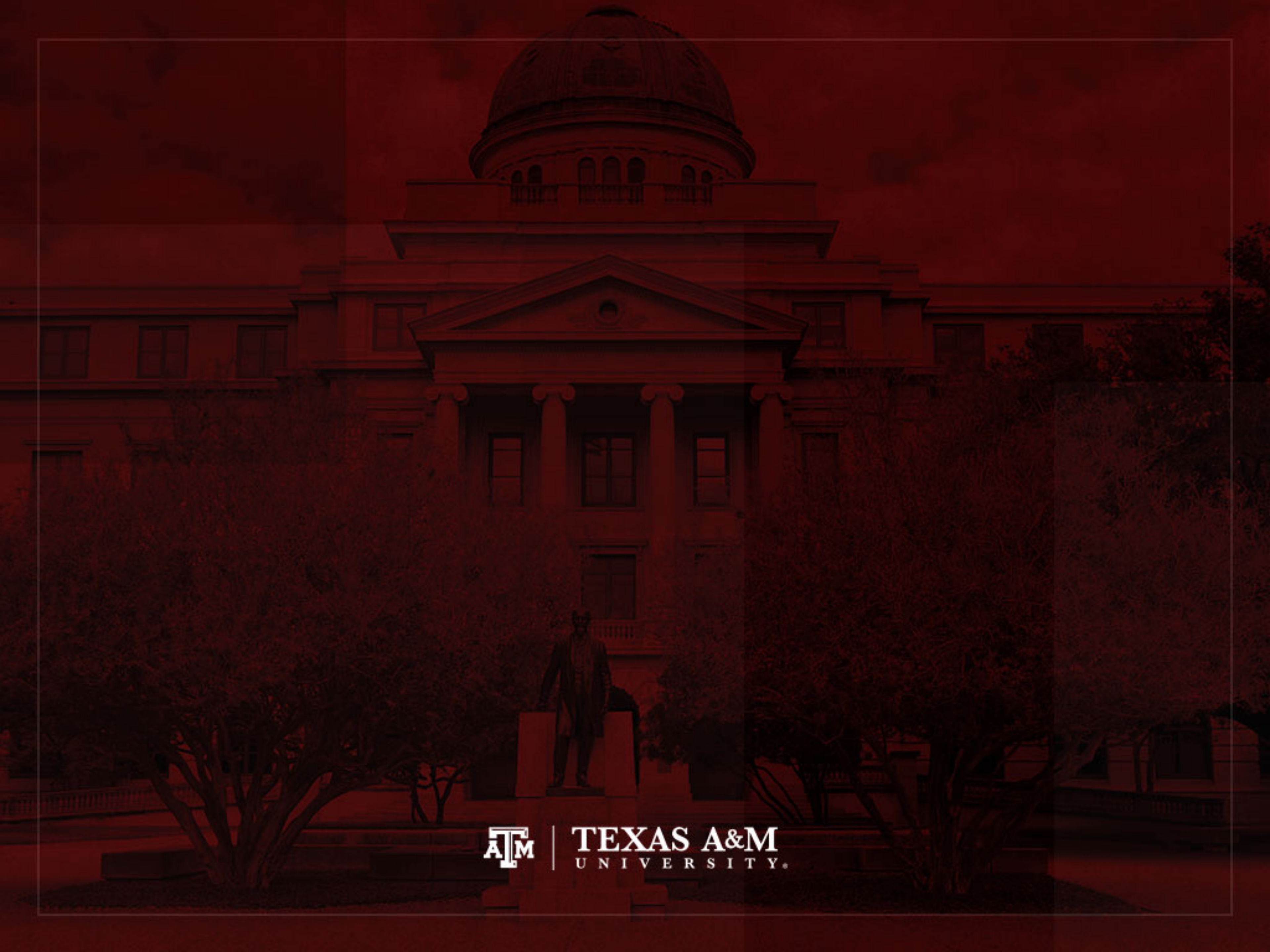
female	Freq.	Percent	Cum.
0	1,736	44.06	44.06
1	2,204	55.94	100.00
Total	3,940	100.00	

- What causes the difference between 50.5% (population, Census) and 55.94% (sample 18+, GSS)?
- Real difference? Or difference due to random chance?

Example 6: Result

- Population 18+ has a statistically significant higher proportion of women than overall population
  - The difference between the parameter 50.5% and the statistic 55.94% was large and unlikely to have occurred by random chance (*p*-value<0.05)

```
. prtest female=.505

One-sample test of proportion                 Number of obs      =      3940

------------------------------------------------------------------------------
    Variable |       Mean   Std. err.                     [95% conf. interval]
-------------+----------------------------------------------------------------
      female |   .5593909   .0079093                      .543889    .5748927
------------------------------------------------------------------------------
    p = proportion(female)                                         z =   6.8285
H0: p = 0.505

    Ha: p < 0.505                Ha: p != 0.505                 Ha: p > 0.505
 Pr(Z < z) = 1.0000         Pr(|Z| > |z|) = 0.0000          Pr(Z > z) = 0.0000
```


TEXAS A&M
UNIVERSITY

Two-sample case

- Identify and cite examples of situations in which the two-sample test of hypothesis is appropriate
- Explain the logic of hypothesis testing, as applied to the two-sample case
- Explain what an independent random sample is
- Perform a test of hypothesis for two sample means or two sample proportions, following the five-step model and correctly interpret the results
- List and explain each of the factors (especially sample size) that affect the probability of rejecting the null hypothesis
- Explain the differences between statistical significance and importance

Basic logic

- We analyze a difference between two sample statistics
  - We compare means or proportions of two samples from specific sub-groups of the population

- This is the question under consideration
  - "Is the difference between the samples large enough to allow us to conclude (with a known probability of error) that the populations represented by the samples are different?"

Source: Healey 2015, p.217.

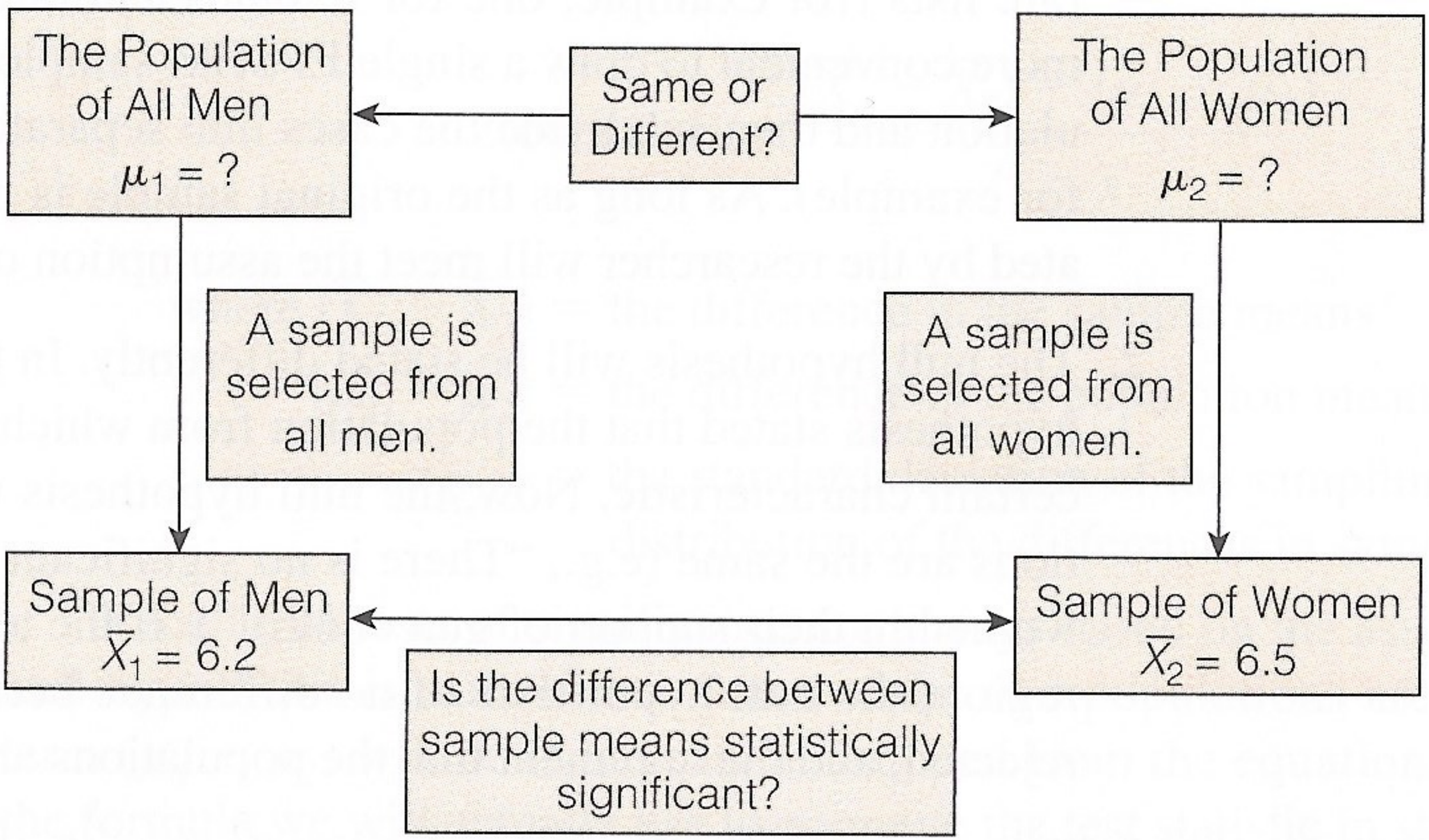
Null hypothesis

- The H_0 indicates that the populations are the same
  - Assuming that the H_0 is true, there is no difference between the parameters of the two populations
- On the other hand, we reject the H_0 and say there is a difference between the populations
  - If the difference between the sample statistics is large enough
  - Or if the size of the estimated difference is unlikely

H_0, *α*, *Z* score, *p*-value

- The H_0 is a statement of "no difference"
- The 0.05 level (*α*) will continue to be our indicator of a significant difference
- We change the sample statistics to a *Z* score
  - Place the *Z(obtained)* on the sampling distribution
- Estimate probability (*p*-value) above *Z(obtained)*
  - *p*-value is the probability of not rejecting the null hypothesis
  - Compare the *p*-value to the *α*
  - If $p<\alpha$, we reject H_0
  - If $p>\alpha$, we do not reject H_0

Test of hypothesis for two sample means



Source: Healey 2015, p.217.

The five-step model

1. Make assumptions and meet test requirements
2. Define the null hypothesis (H_0)
3. Select the sampling distribution and establish the critical region
4. Compute the test statistic
5. Make a decision and interpret the test results

Changes from one-sample case

- Step 1
  - In addition to samples selected according to EPSEM principles
  - Samples must be selected independently of each other: independent random sampling
- Step 2
  - Null hypothesis statement will state that the two populations are not different
- Step 3
  - Sampling distribution refers to difference between the sample statistics

Two-sample test of means
(large samples)

- Do men and women significantly differ on their support of gun control?
- For men (sample 1)
  - Mean = 6.2
  - Standard deviation = 1.3
  - Sample size = 324
- For women (sample 2)
  - Mean = 6.5
  - Standard deviation = 1.4
  - Sample size = 317

Step 1: Assumptions,requirements

- Independent random sampling
  - The samples must be independent of each other
- Level of measurement is interval-ratio
  - Support of gun control is assessed with an interval-ratio level scale, so the mean is an appropriate statistic
- Sampling distribution is normal in shape
  - Total $n \geq 100$ ($n_1 + n_2 = 324 + 317 = 641$)
  - Thus, the Central Limit Theorem applies and we can assume a standard normal distribution (Z)

Step 2: Null hypothesis

- Null hypothesis, H_0: $\mu_1 = \mu_2$
  - The null hypothesis asserts there is no difference between the populations

- Alternative hypothesis, H_1: $\mu_1 \neq \mu_2$
  - The research hypothesis contradicts the H_0 and asserts there is a difference between the populations

Step 3: Distribution, critical region

- Sampling distribution
  - Standard normal distribution (*Z*)

- Significance level
  - Alpha (*α*) = 0.05 (two-tailed)
  - The decision to reject the null hypothesis has only a 0.05 probability of being incorrect

- *Z(critical)* = ±1.96
  - If the probability (*p*-value) is less than 0.05
  - *Z(obtained)* will be beyond *Z(critical)*

Step 4: Test statistic

- Sample outcomes for support of gun control

Sample 1 (men)	Sample 2 (women)
$\bar{X}_1 = 6.2$	$\bar{X}_2 = 6.5$
$s_1 = 1.3$	$s_2 = 1.4$
$n_1 = 324$	$n_2 = 317$

- Pooled estimate of the standard error

$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}} = \sqrt{\frac{(1.3)^2}{324 - 1} + \frac{(1.4)^2}{317 - 1}} = 0.107$$

- Obtained *Z* score

$$Z(obtained) = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}-\bar{X}}} = \frac{6.2 - 6.5}{0.107} = -2.80$$

Step 5: Decision, interpret

- *Z(obtained)* = –2.80
  - This is beyond *Z(critical)* = ±1.96
  - The obtained *Z* score falls in the critical region, so we ***reject*** the H_0
  - Therefore, the H_0 is false and must be rejected

- The difference between men's and women's support of gun control is statistically significant
  - The difference between the sample means is so large that we can conclude (at $\alpha = 0.05$) that a difference exists between the populations represented by the samples

Two-sample test of means
(small samples)

- Do families that reside in the center-city have more children than families that reside in the suburbs?
- For suburbs (sample 1)
  - Mean = 2.37
  - Standard deviation = 0.63
  - Sample size = 42
- For center-city (sample 2)
  - Mean = 2.78
  - Standard deviation = 0.95
  - Sample size = 37

Step 1: Assumptions,requirements

- Independent random sampling
  - The samples must be independent of each other
- Level of measurement is interval-ratio
  - Number of children can be treated as interval-ratio
- Population variances are equal
  - As long as the two samples are approximately the same size, we can make this assumption
- Sampling distribution is normal in shape
  - Because we have two small samples (n < 100), we have to add the previous assumption in order to meet this assumption

Step 2: Null hypothesis

- Null hypothesis, H_0: $\mu_1 = \mu_2$
  - The null hypothesis asserts there is no difference between the populations

- Alternative hypothesis, H_1: $\mu_1 < \mu_2$
  - The research hypothesis contradicts the H_0 and asserts there is a difference between the populations

Step 3: Distribution, critical region

- Sampling distribution
  - Student's *t* distribution
- Significance level
  - Alpha (α) = 0.05 (one-tailed)
- Degrees of freedom
  - $n_1 + n_2 - 2 = 42 + 37 - 2 = 77$
- Critical *t*
  - $t(\text{critical}) = -1.671$

Step 4: Test statistic

- Sample outcomes for number of children

Sample 1 (suburban)	Sample 2 (center-city)
$\bar{X}_1 = 2.37$	$\bar{X}_2 = 2.78$
$s_1 = 0.63$	$s_2 = 0.95$
$n_1 = 42$	$n_2 = 37$

- Pooled estimate of the standard error

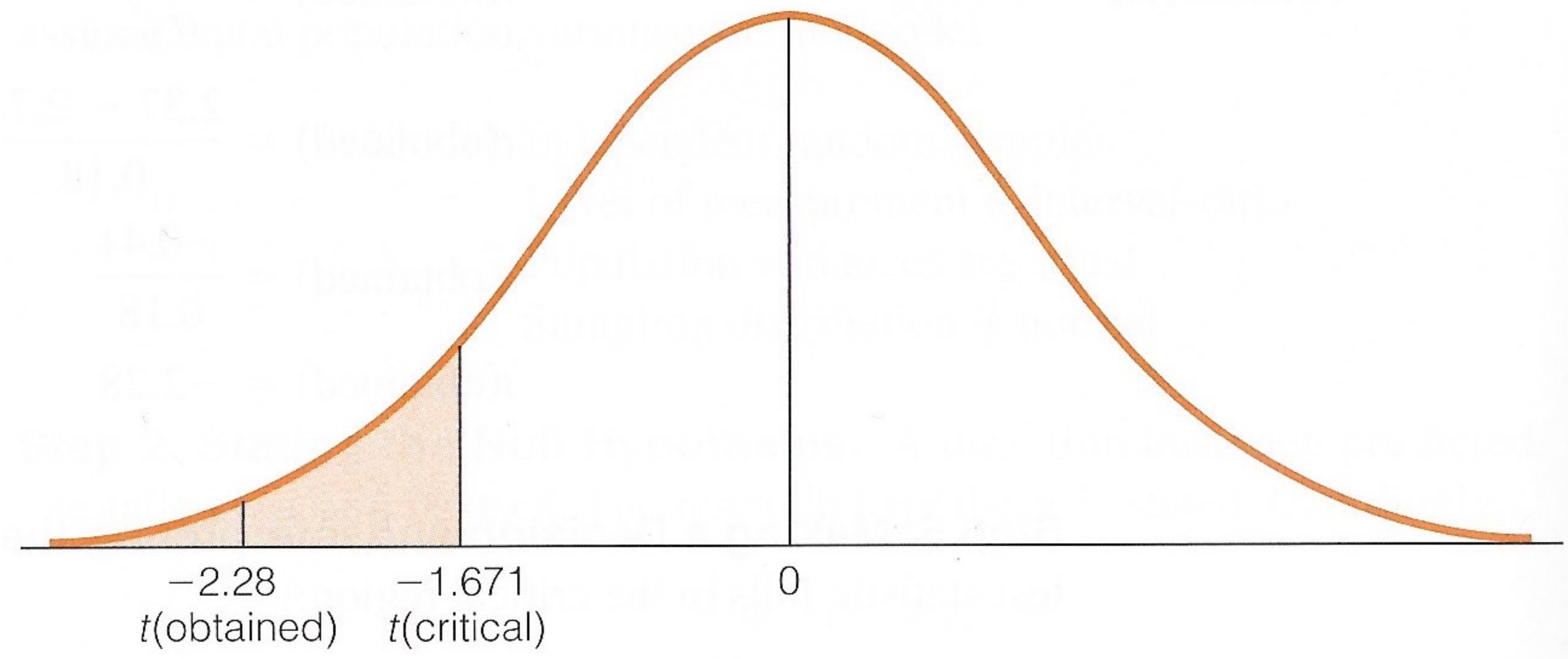
$$\sigma_{\bar{X}-\bar{X}} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = \sqrt{\frac{(42)(0.63)^2 + (37)(0.95)^2}{42 + 37 - 2}} \sqrt{\frac{42 + 37}{(42)(37)}} = 0.18$$

- Obtained t

$$t(obtained) = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}-\bar{X}}} = \frac{2.37 - 2.78}{0.18} = -2.28$$

t(obtained) & *t*(critical)

- Sampling distribution with critical region and test statistic displayed



Source: Healey 2015, p.226.

Step 5: Decision, interpret

- t(obtained) = –2.28
  - This is beyond t(critical) = –1.671
  - The obtained test statistic falls in the critical region, so we ***reject*** the H_0

- The difference between the number of children in center-city families and the suburban families is statistically significant
  - The difference between the sample means is so large that we can conclude (at α = 0.05) that a difference exists between the populations represented by the samples

Example from GSS: *t*-test

- We know the average income by sex from the 2016 GSS

```
. table sex, c(mean conrinc)
```

respondents sex	mean(conrinc)
male	**41583.52814**
female	**28353.34628**

- What causes the difference between male income of $41,583.53 and female income of $28,353.35?
- Real difference? Or difference due to random chance?

Example from GSS: Result

- Men have an average income that is significantly higher than the female average income
  - The difference between male income ($41,583.53) and female income ($28,353.35) was large and unlikely to have occurred by random chance (p<0.05) in 2016

```
. ttest conrinc, by(sex)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |     798    41583.53    1433.963    40507.87    38768.74    44398.32
  female |     834    28353.35    1049.496    30308.45    26293.38    30413.31
---------+--------------------------------------------------------------------
combined |   1,632    34822.52    897.5571    36259.53    33062.03       36583
---------+--------------------------------------------------------------------
    diff |            13230.18    1765.955                9766.402    16693.96
------------------------------------------------------------------------------
    diff = mean(male) - mean(female)                              t =   7.4918
Ho: diff = 0                                     degrees of freedom =     1630

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Edited table

Table 1. Two-sample *t*-test of individual average income of the U.S. adult population by sex, 2004, 2010, and 2016

Sex	2004	2010	2016
Male	45,741.48	37,864.34	41,583.53
	(1,343.92)	(1,359.39)	(1,433.96)
Female	29,264.54	26,141.60	28,353.35
	(972.15)	(972.97)	(1,049.50)
Difference	16,476.94***	11,722.74***	13,230.18***
	(1,665.71)	(1,643.94)	(1,765.96)
Sample size	1,688	1,202	1,632

Note: Standard errors are reported in parentheses. *Significant at p<0.10; **Significant at p<0.05; ***Significant at p<0.01.
Source: 2004, 2010, 2016 General Social Surveys.

Two-sample test of proportions (large samples)

- Do Black and White senior citizens differ in their number of memberships in clubs and organizations?
  - Using the proportion of each group classified as having a “high” level of membership
- For Black senior citizens (sample 1)
  - Proportion = 0.34
  - Sample size = 83
- For White senior citizens (sample 2)
  - Proportion = 0.25
  - Sample size = 103

Step 1: Assumptions,requirements

- Independent random sampling
  - The samples must be independent of each other
- Level of measurement is nominal
  - We have measured the proportion of each group classified as having a “high” level of membership
- Population variances are equal
  - As long as the two samples are approximately the same size, we can make this assumption
- Sampling distribution is normal in shape
  - Total $n \geq 100$ ($n_1 + n_2 = 83 + 103 = 186$)
  - Thus, the Central Limit Theorem applies and we can assume a standard normal distribution

Step 2: Null hypothesis

- Null hypothesis, H_0: $P_{u1} = P_{u2}$
  - The null hypothesis asserts there is no difference between the populations
- Alternative hypothesis, H_1: $P_{u1} \neq P_{u2}$
  - The research hypothesis contradicts the H_0 and asserts there is a difference between the populations

Step 3: Distribution, critical region

- Sampling distribution
  - Standard normal distribution (*Z*)

- Significance level
  - Alpha (α) = 0.05 (two-tailed)
  - The decision to reject the null hypothesis has only a 0.05 probability of being incorrect

- *Z(critical)* = ±1.96
  - If the probability (*p*-value) is less than 0.05
  - *Z(obtained)* will be beyond *Z(critical)*

Step 4: Test statistic

- Sample outcomes for club memberships

Sample 1 (Black senior citizens)	Sample 2 (White senior citizens)
$P_{s1} = 0.34$	$P_{s2} = 0.25$
$n_1 = 83$	$n_2 = 103$

- Population proportion

$$P_u = \frac{n_1 P_{s1} + n_2 P_{s2}}{n_1 + n_2} = \frac{(83)(0.34) + (103)(0.25)}{83 + 103} = 0.29$$

- Pooled estimate of the standard error

$$\sigma_{p-p} = \sqrt{P_u(1 - P_u)}\sqrt{\frac{n_1 + n_2}{n_1 n_2}} = \sqrt{(0.29)(0.71)}\sqrt{\frac{83 + 103}{(83)(103)}} = 0.07$$

- Obtained Z score

$$Z(obtained) = \frac{P_{s1} - P_{s2}}{\sigma_{p-p}} = \frac{0.34 - 0.25}{0.07} = 1.29$$

Step 5: Decision, interpret

- *Z(obtained)* = 1.29
  - This is below the *Z(critical)* = 1.96
  - The obtained test statistic does not fall in the critical region, so we ***do not reject*** the H_0

- The difference between the memberships of Black and White senior citizens is not significant
  - The difference between the sample means is small enough that we can conclude (at α = 0.05) that no difference exists between the populations represented by the samples

Example from GSS: proportion

- We know the proportion of pro-immigrants by political party from the 2016 GSS

```
. table democrat, c(mean proimmig)
```

Political party	mean(proimmig)
Republicans	.117096
Democrats	.4559471

- What causes the difference between the percentage of Republicans who a pro-immigration (11.7%) and the percentage of Democrats who are pro-immigration (45.6%)?
  - Real difference? Or difference due to random chance?

Example from GSS: Result

- Republicans are less pro-immigration than Democrats
  - The difference between the percentage of Republicans who are pro-immigration (11.7%) and the percentage of Democrats who are pro-immigration (45.6%) was large and unlikely to have occurred by random chance (*p*<0.05) in 2016

```
. prtest proimmig, by(democrat)

Two-sample test of proportions          Republicans: Number of obs =       427
                                          Democrats: Number of obs =       454

------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Republicans |    .117096   .0155602                        .0865987    .1475934
   Democrats |   .4559471   .0233749                        .4101332    .5017611
-------------+----------------------------------------------------------------
        diff |  -.3388511   .0280803                       -.3938875   -.2838147
             |  under Ho:   .0306428   -11.06   0.000
------------------------------------------------------------------------------
        diff = prop(Republicans) - prop(Democrats)                z = -11.0581
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.0000         Pr(|Z| > |z|) = 0.0000          Pr(Z > z) = 1.0000
```

Edited table

Table 2. Test of proportions of pro-immigrants among the U.S. adult population by political party, 2004, 2010, and 2016

Political Party	2004	2010	2016
Republican	0.0911	0.1429	0.1171
	(0.0124)	(0.0193)	(0.0156)
Democratic	0.2164	0.2761	0.4559
	(0.0178)	(0.0223)	(0.0234)
Difference	–0.1253***	–0.1333***	–0.3389***
	(0.0217)	(0.0295)	(0.0281)
Sample size	1,074	731	881

Note: Standard errors are reported in parentheses. *Significant at $p<0.10$; **Significant at $p<0.05$; ***Significant at $p<0.01$.
Source: 2004, 2010, 2016 General Social Surveys.

Statistical significance vs. importance (magnitude)

- As long as we work with random samples, we must conduct a test of significance

- Statistical significance is not the same thing as importance
  - Importance is also known as magnitude of the effect

- Differences that are otherwise trivial or uninteresting may be significant

Influence of sample size

- When working with large samples, even small differences may be statistically significant
- The larger the sample size (n)
  - The greater the value of the test statistic
  - The more likely it will fall in the critical region and be declared statistically significant
- In general, when working with random samples, statistical significance is a necessary but not a sufficient condition for importance

Sample size & test statistic

Test Statistics for Single-Sample Means Computed from Samples of Various Sizes ($\bar{X} = 80$, $\mu = 79$, $s = 5$ throughout)

Sample Size (*N*)	Computing the Test Statistic	Test Statistic, *Z*(Obtained)
50	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{49}} = \frac{1}{0.71} =$	1.41
100	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{99}} = \frac{1}{0.50} =$	2.00
500	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{499}} = \frac{1}{0.22} =$	4.55
1000	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{999}} = \frac{1}{0.16} =$	6.25
10,000	$Z(\text{obtained}) = \frac{\bar{X} - \mu}{\sigma/\sqrt{N-1}} = \frac{80 - 79}{5/\sqrt{9999}} = \frac{1}{0.05} =$	20.00

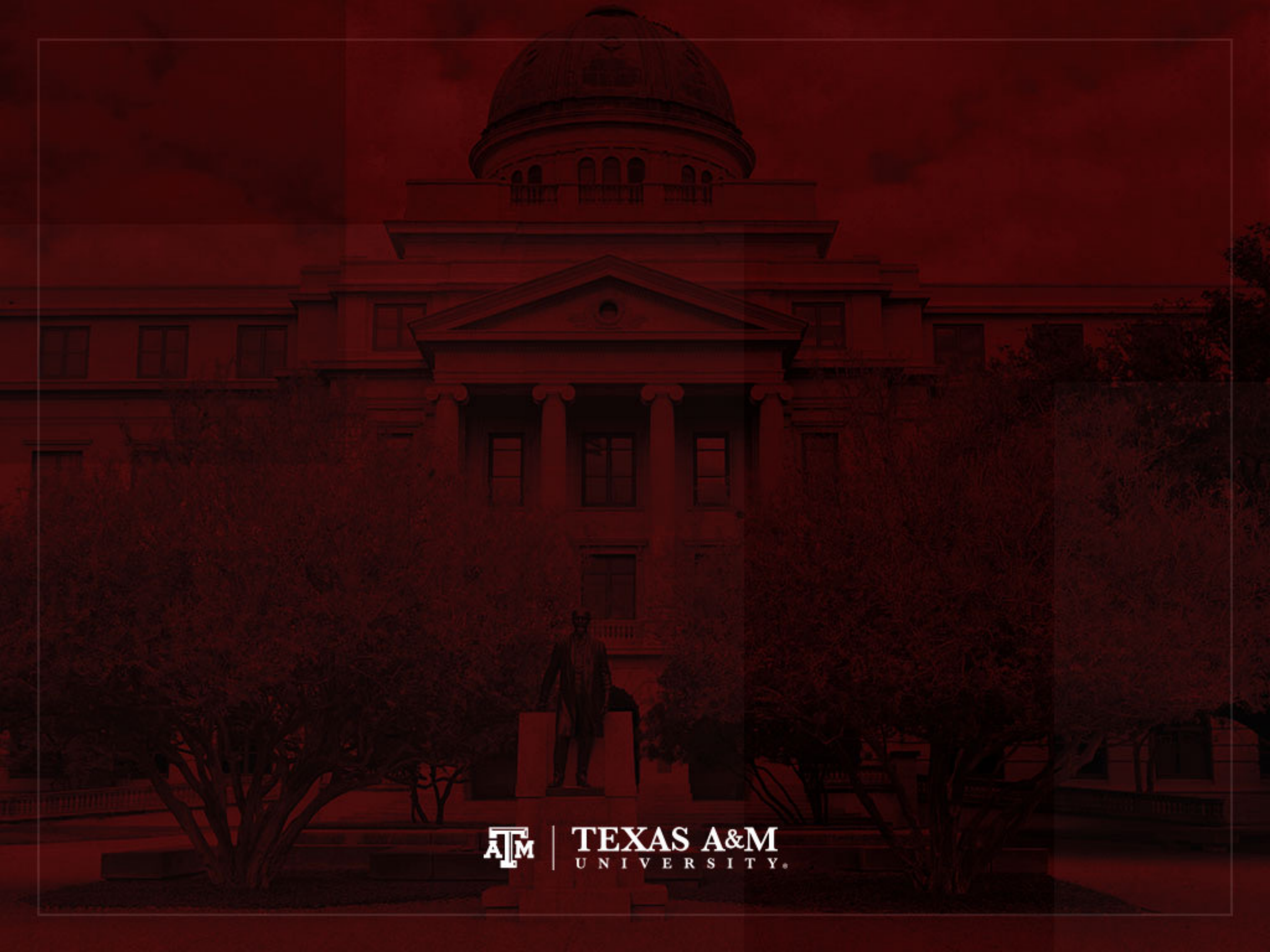
Source: Healey 2015, p.234.

Outcomes of hypothesis testing

- Result of a specific analysis could be
  - Statistically significant and
    - Important (large magnitude)
  - Statistically significant, but
    - Unimportant (small magnitude)
  - Not statistically significant, but
    - Important (large magnitude)
  - Not statistically significant and
    - Unimportant (small magnitude)

Factors influencing the decision

1. The size of the observed difference
   - For larger differences, we are more likely to reject H_0
2. The value of alpha
   - Usually the decision to reject the null hypothesis has only a 0.05 probability of being incorrect
   - The higher the alpha
     - The more likely we are to reject the H_0
     - But we would have a higher chance of being incorrect
3. The use of one- vs. two-tailed tests
   - We are more likely to reject H_0 with a one-tailed test
4. The size of the sample (*n*)
   - For larger samples, we are more likely to reject H_0


TEXAS A&M
UNIVERSITY