

Lecture 8: Hypothesis testing: One-sample case

Ernesto F. L. Amaral

October 03–05, 2023

Advanced Methods of Social Research (SOCL 420)

www.ernestoamaral.com

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



TEXAS A&M
UNIVERSITY.

Outline

- 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

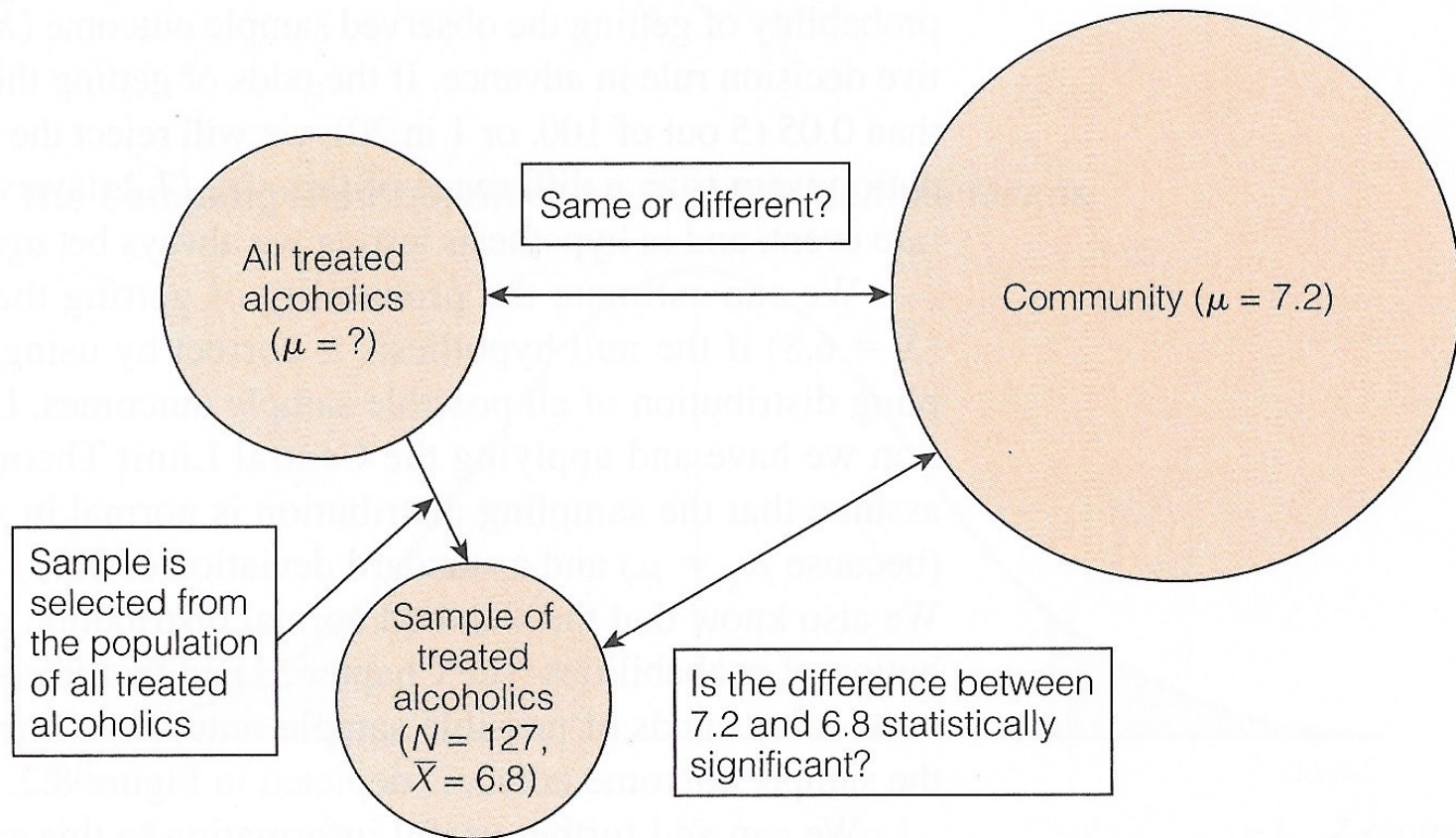
- Are people who have been treated for alcoholism more reliable workers than those in the community?
 - Does the group of all treated alcoholics have different absentee rates than the community as a whole?
 - Effectiveness of rehabilitation center for alcoholics
- Absentee rates for community and sample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

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

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



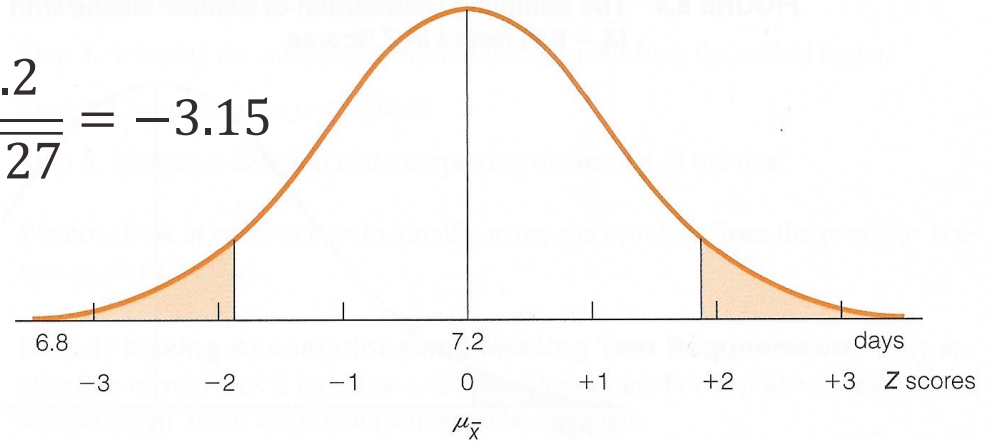
A test of hypothesis for single-sample means



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

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



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

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 ($\mu=2.70$) 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 ($\mu=2.70$)
 - 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 ($\mu=2.7$)



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 $\alpha=0.05$



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)

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



Step 5: Decision, interpret

- $Z(\text{obtained}) = 4.62$
 - This is beyond $Z(\text{critical}) = \pm 1.96$
 - 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 ($\mu=2.7$) 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



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

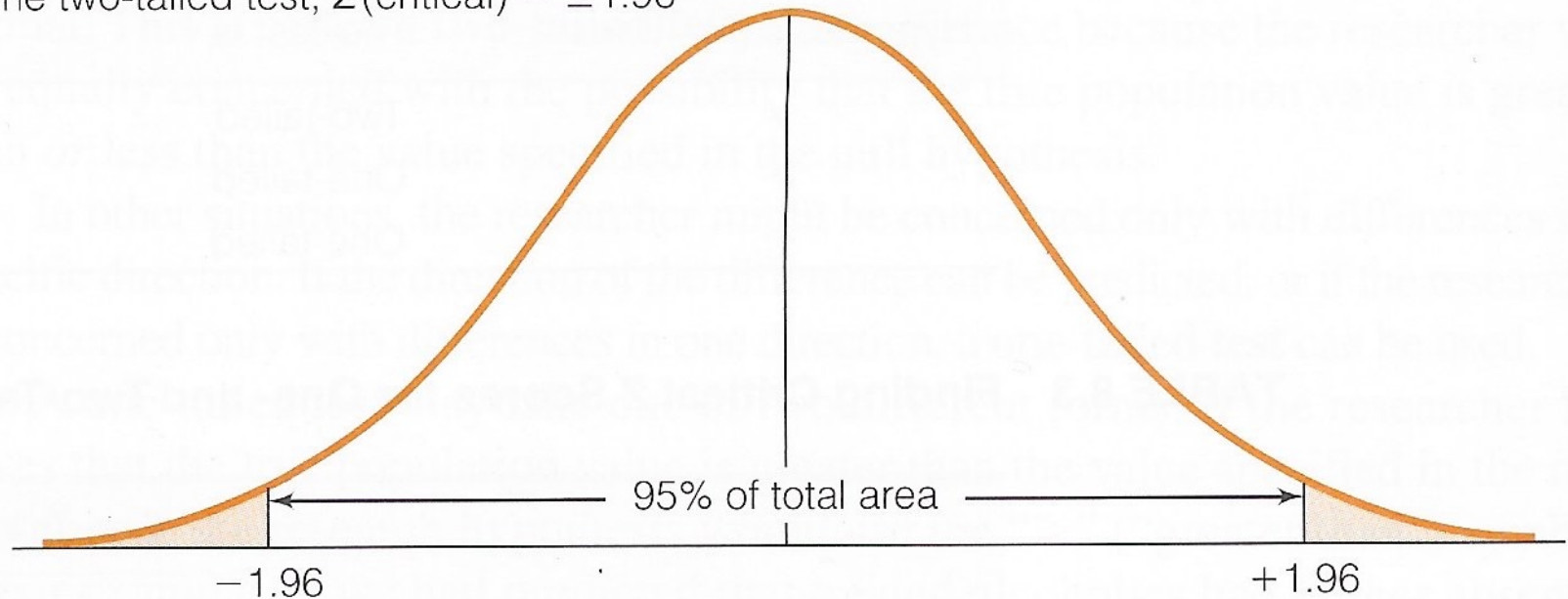
Finding Critical Z Scores for One- and Two-Tailed Tests

Alpha	Two-Tailed Value	One-Tailed Value	
		<i>Upper Tail</i>	<i>Lower Tail</i>
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



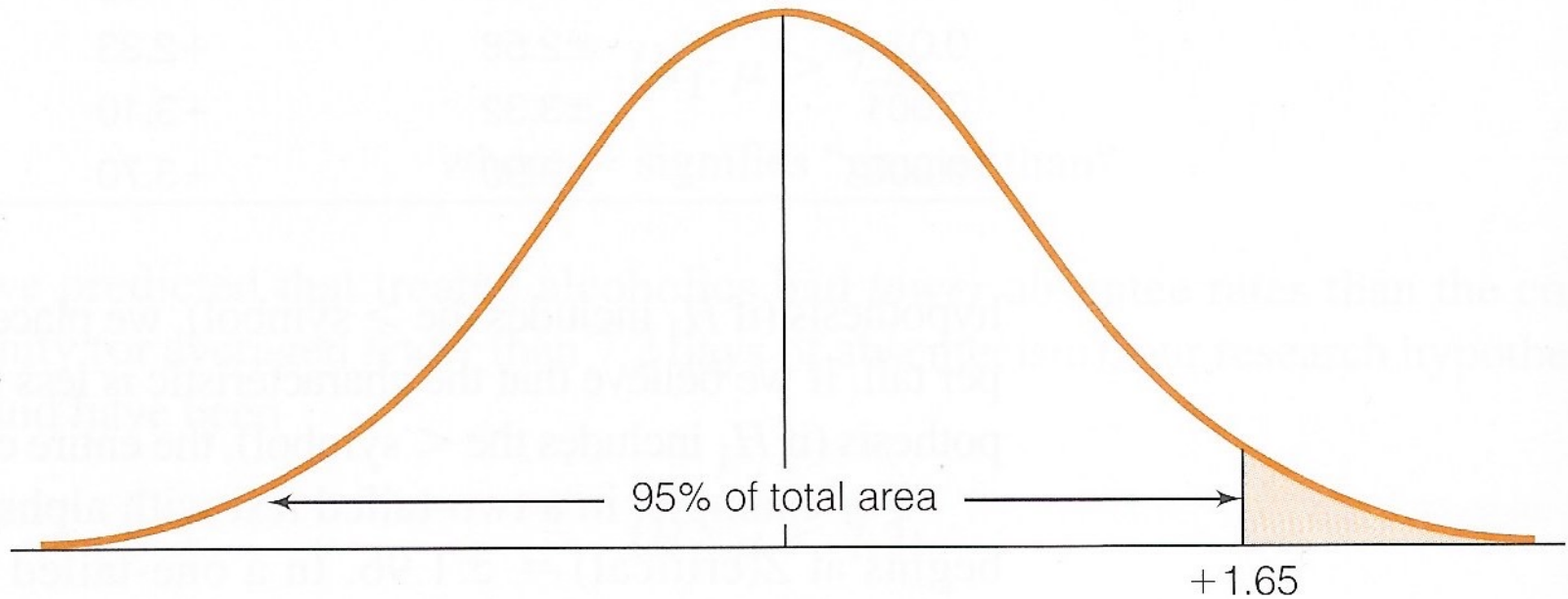
Two-tailed test: $\alpha=0.05$

A. The two-tailed test, $Z(\text{critical}) = \pm 1.96$



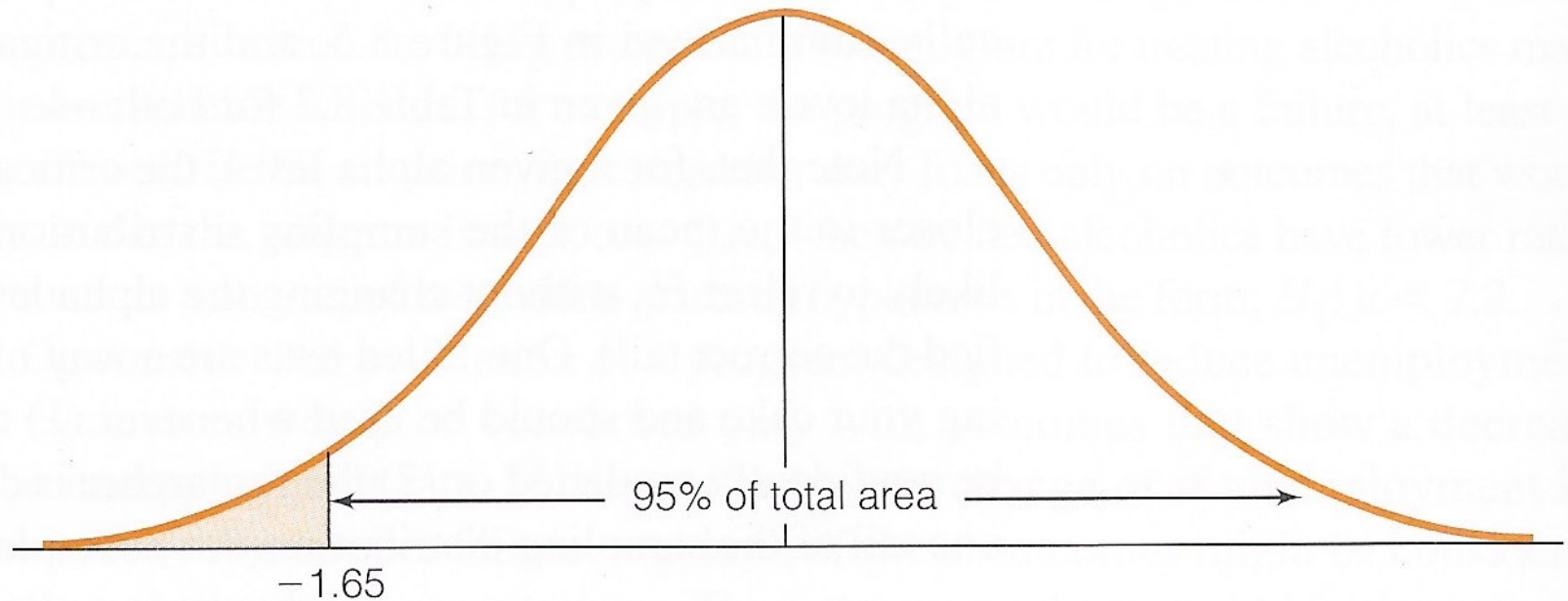
One-tailed test (upper): $\alpha=0.05$

B. The one-tailed test for upper tail, $Z(\text{critical}) = +1.65$



One-tailed test (lower): $\alpha=0.05$

C. The one-tailed test for lower tail, $Z(\text{critical}) = -1.65$



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



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



Decisions about hypotheses

Hypotheses	$p < \alpha$	$p > \alpha$
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

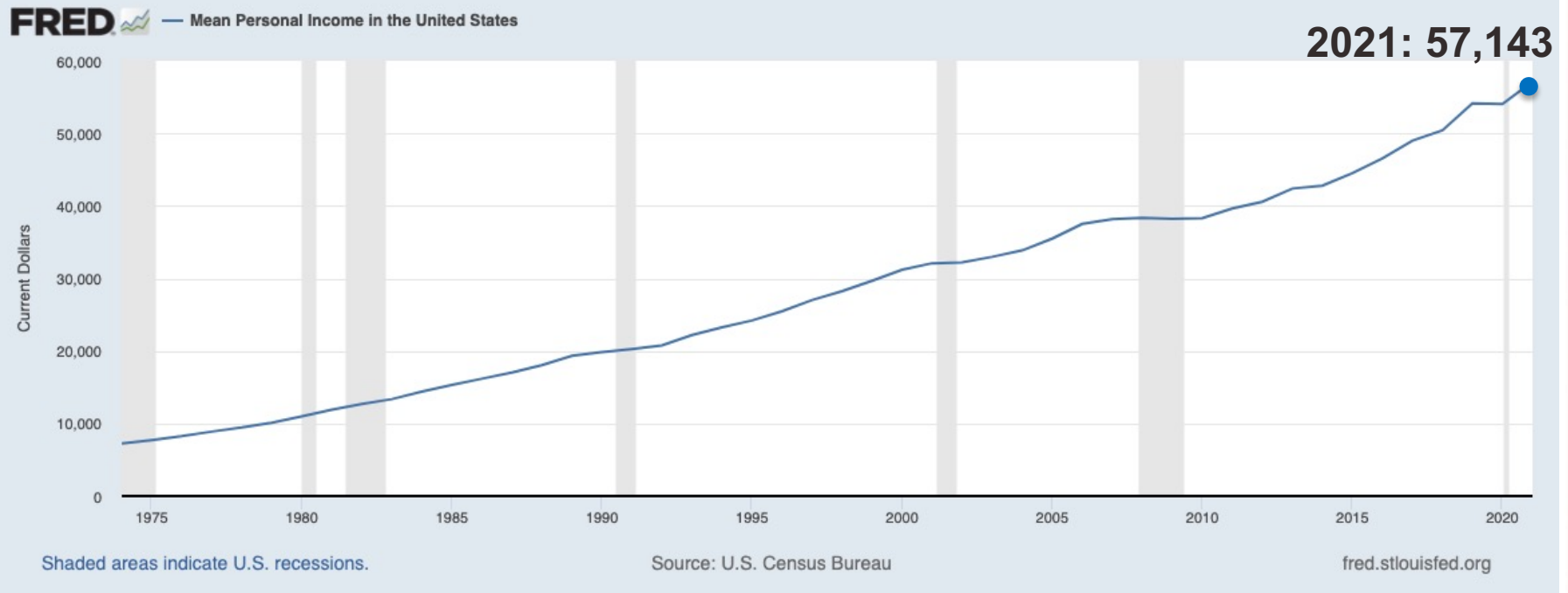
– If a statistical software gives only the two-tailed *p*-value, divide it by 2 to obtain the one-tailed *p*-value

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
Pr(Z < z) = 0.0000

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

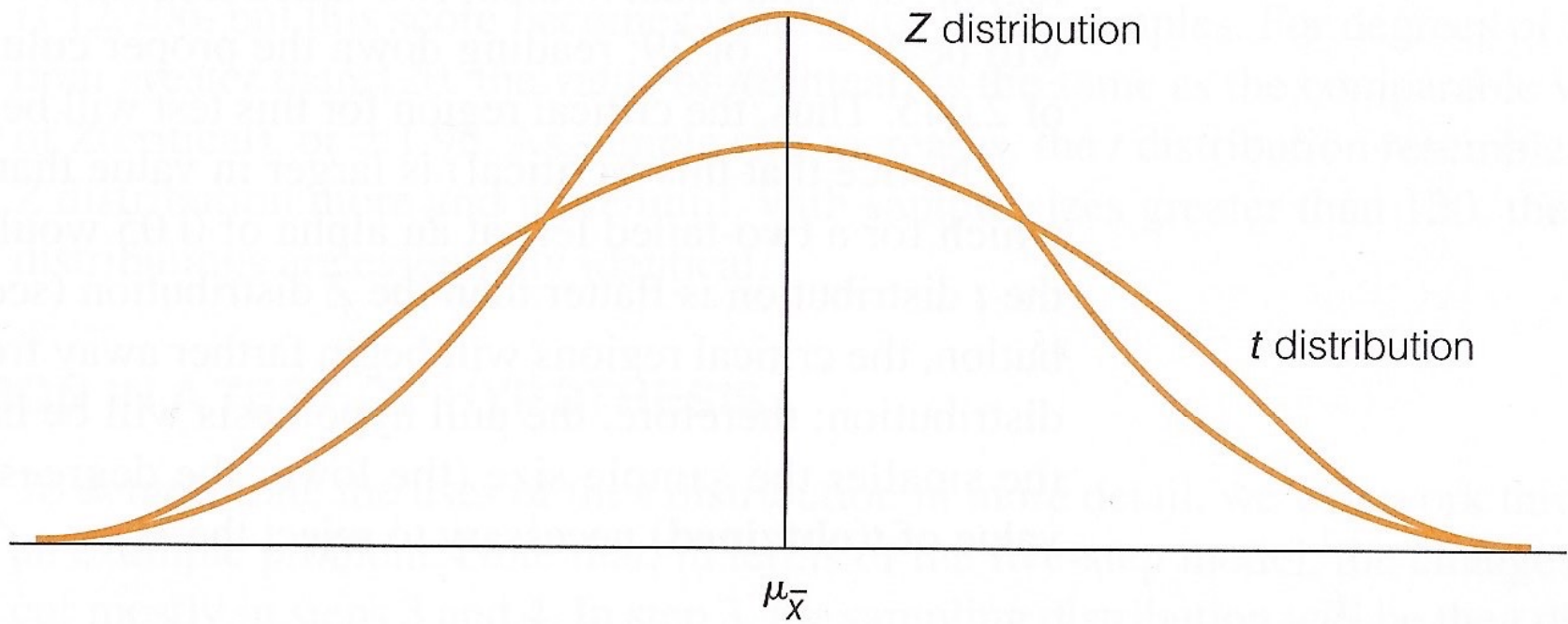
Ha: mean > 57143
Pr(Z > z) = 1.0000

The Student's t distribution

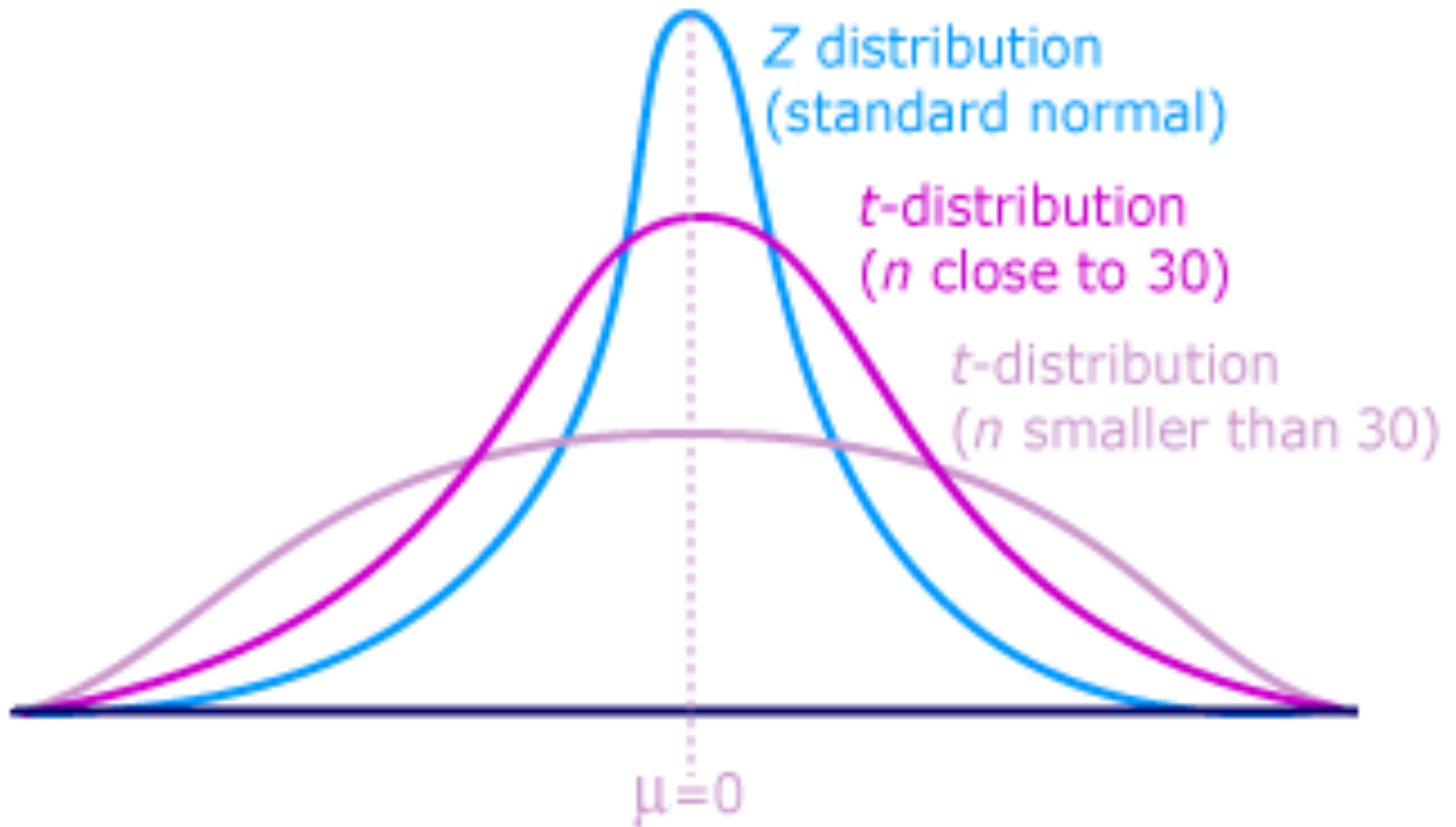
- How can we test a hypothesis when the population standard deviation (σ) is unknown, as is usually the case?
- For large samples ($n \geq 100$), we can use the sample standard deviation (s) as an estimator of the population standard deviation (σ)
 - Use standard normal distribution (Z)
- For small samples,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



t and Z distributions



t and Z distributions



Source: <https://joejeong33.wordpress.com/2013/06/03/t-distribution-in-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



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
Pr(T < t) = 0.0052
```

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

```
Ha: mean > 57143
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 ($\alpha = 0.10$)



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 $Z(\text{critical}) = \pm 1.65$
 - 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

- $Z(\text{obtained}) = 0.91$
 - $Z(\text{obtained})$ did not fall in the critical region delimited by $Z(\text{critical}) = \pm 1.65$, 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  
Pr(Z < z) = 1.0000
```

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

```
Ha: p > 0.505  
Pr(Z > z) = 0.0000
```



TEXAS A&M
UNIVERSITY.