MBC 638

LIVE SESSION WEEK 5

Agenda

Topic	Time	Wednesday Section
Introduction	5 min	9:00-9:05
Homework #2 Chapter 9 Quiz Quiz #2 Prep Review	35 min	9:05-9:40
Highlights from Week 5 Video	30 min	9:40-10:10
Breakout on Additional Example	10 min	10:10-10:20
Review of Upcoming Assignments and Open Question	10 min	10:20-10:30

Chapter 9 Online Practice Quiz

1. According to the U.S. Energy Information Administration, the average household expenditure for natural gas was $679 in the winter of 2013. Suppose that a random sample of 50 customers shows a mean expenditure of $712 and assume that the population standard deviation is $80. Test whether the population mean expenditure is greater than $679 using level of significance $\alpha = 0.05$. State the *p*-value and conclusion.

a. a) *p*-value = 0.0035; reject H_0: there is evidence at level of significance $\alpha = 0.05$ that the mean expenditure is greater than $679.

b. b) *p*-value = 0.9982; reject H_0: there is evidence at level of significance $\alpha = 0.05$ that the mean expenditure is greater than $679.

c. c) *p*-value = 0.0018; do not reject H_0: there is insufficient evidence at level of significance α = 0.05 that the mean expenditure is greater than $679.

d. d) *p*-value = 0.0018; mean expenditure is greater than $679.

and conclusion.

a. a) *p*-value = 0.0035; reject H_0: there is evidence at level of significance $\alpha = 0.05$ that the mean expenditure is greater than $679.

Incorrect. This *p*-value is the result of using a two-tailed test, whereas this situation calls for a right-tailed test with a *p*-value of 0.0018.
DATA SOURCE: Adapted from http://www.eia.gov/todayinenergy/detail.cfm?id=13311
Text Reference – pages 507-510
Section 9.3: *Z* Test for the Population Mean: *p*-value Method

b. b) *p*-value = 0.9982; reject H_0: there is evidence at level of significance $\alpha = 0.05$ that the mean expenditure is greater than $679.

Incorrect. This *p*-value is the result of using a left-tailed test, whereas this situation calls for a right-tailed test with a *p*-value of 0.0018.
DATA SOURCE: Adapted from http://www.eia.gov/todayinenergy/detail.cfm?id=13311
Text Reference – pages 507-510
Section 9.3: *Z* Test for the Population Mean: *p*-value Method

c. c) *p*-value = 0.0018; do not reject H_0: there is insufficient evidence at level of significance α = 0.05 that the mean expenditure is greater than $679.

Incorrect. Because the *p*-value is less than the significance level, we reject H_0 and conclude that there is evidence at level of significance $\alpha = 0.05$ that the mean expenditure is greater than $679.
DATA SOURCE: Adapted from http://www.eia.gov/todayinenergy/detail.cfm?id=13311
Text Reference – pages 507-510
Section 9.3: *Z* Test for the Population Mean: *p*-value Method

d. d) *p*-value = 0.0018; mean expenditure is greater than $679.

Correct. Because the *p*-value is less than the significance level, we reject H_0 and conclude that there is evidence at level of significance $\alpha = 0.05$ that the mean expenditure is greater than $679.
DATA SOURCE: Adapted from http://www.eia.gov/todayinenergy/detail.cfm?id=13311
Text Reference – pages 507-510
Section 9.3: *Z* Test for the Population Mean: *p*-value Method

#1

$\mu = 679$ 2013
$n = 50$
$\mu = 712$
$s = 80$

Upper/right-tail
$H_0: \mu \le \mu_0$
$H_a: \mu > \mu_0$

$H_0: \mu \le 679$
$H_a: \mu > 679$

Purple chart upper right tail

$n \geq 30$ so large

Large
$n \ge 30$
(or σ known)

$$Z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{712-679}{80/\sqrt{50}}$$

$$= \frac{712-679}{80/7.07} = \frac{712-679}{11.315}$$

$$= \frac{33}{11.315} = 2.916$$

pvalue of $Z = 2.916$ = .9982

But we want area to right of curve

Upper/right-tail
p = area right of Z or t

$1 - .9982 = .0018$
↳ NORM.S.DIST(2.916, TRUE)

P is lower than alpha = .05
so reject Ho
so yes Ha is true

D is the correct answer

Chapter 9 Online Practice Quiz

2. Which of the following alternative hypotheses best fits the following scenario?
The Bureau of Justice Statistics reports that in 2000, prison sentences for drug offenses had a mean length of 47 months. After an extended campaign to discourage such offenders by increasing the penalty for such crimes, a random sample is taken and finds a mean stay of 54 months. Has there been a significant increase in the length of stay?

a. a) H_a: $\mu < 47$

Incorrect. We are interested in finding an increase in the length of stay, so we consider $\mu > \mu_0$.
Text Reference: pages 488-491
Section 9.1: Introduction to Hypothesis Testing

b. b) H_a: $\mu \neq 47$

Incorrect. Consider the key phrase "significant increase."
Text Reference: pages 488-491
Section 9.1: Introduction to Hypothesis Testing

c. c) H_a: $\mu > 47$

Correct. This alternative hypothesis results in a right-tailed test.
Text Reference: pages 488-491
Section 9.1: Introduction to Hypothesis Testing

#2 $\mu = 47$ as a std
new $\mu = 54$

$H_a: \mu > 47$

ANSWER IS C

Chapter 9 Online Practice Quiz

3. In a recent hypothesis test, the null hypothesis was NOT rejected. Which of the following types of error could have been made, and what is the probability of such an error?

a. a) Type I error with probability of α

Incorrect. A Type I error is possible only if the null hypothesis is rejected.
Text Reference: pages 493-494
Section 9.1: Introduction to Hypothesis Testing

b. b) Type II error with probability of β

Correct.
Text Reference: pages 493-494
Section 9.1: Introduction to Hypothesis Testing

c. c) Type I error with probability of β

Incorrect. A Type I error is possible only if the null hypothesis is rejected. Also, the probability of a Type I error is constrained by the selection of α.
Text Reference: pages 493-494
Section 9.1: Introduction to Hypothesis Testing

d. d) Type II error with probability of α

Incorrect. The error type is correct, but the probability of a Type II error is β, not α.
Text Reference: pages 493-494
Section 9.1: Introduction to Hypothesis Testing

#3 Ho not rejected

	Ho true	Ho False
Fail to reject Ho	correct	Type II
reject Ho	Type I	correct

Type II is the possible error which is β

answer is B

Chapter 9 Online Practice Quiz

4. In 2006 a study found that 1% of American Internet users who were married or in a long-term relationship met on a blind date or through a dating service. A survey taken this year found 12 such couples in a random sample of 500 such Internet users. Has there been an increase in the proportion of " successful" matches via blind dates and dating services?

a. a) Yes, the sample gives $P(Z > 3.15) = 0.0008$, which provides extremely strong evidence of an increase in the proportion of " successful" such matches.

Correct.
Text Reference: pages 546-547
Section 9.5: Z Test for the Population Proportion

b. b) No, the sample gives $P(Z > -3.15) = 0.9992$, which does not provide sufficient evidence to claim an increase in the proportion of " successful" such matches.

Incorrect. Be sure to use the sample value minus the null hypothesis, $(\hat{p} - p_0)$. This standardizes the value of $\hat{p}$ with respect to the distribution assumed by p_0.
Text Reference: pages 546-547
Section 9.5: Z Test for the Population Proportion

c. c) No, the sample gives $P(Z > 0.0063) = 0.4997$, which does not provide sufficient evidence to reject the null hypothesis.

Incorrect. Be sure to use the standard deviation of the sampling distribution, $\sqrt{p_0(1-p_0)/n}$, rather than the standard deviation, $\sqrt{np_0(1-p_0)}$.
Text Reference: pages 546-547
Section 9.5: Z Test for the Population Proportion

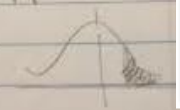
d. d) Yes, the sample gives $P(Z > 2.05) = 0.02$, which provides solid evidence of an increase in the proportion of " successful" such matches.

Incorrect. Be sure to use p_0 rather than $\hat{p}$ to calculate the standard deviation of the sampling distribution.
Text Reference: pages 546-547
Section 9.5: Z Test for the Population Proportion

#4 1% as a std

$H_0: p \le p_0$ orange chart

$H_a: p > p_0$ upper right



Upper/right-tail

$H_0: p \le p_0$

$H_a: p > p_0$

$p = \frac{12}{500} = .024$

$np \ge 5 \rightarrow 500\left(\frac{12}{500}\right) = 12$ yes ≥ 5 ✓

$n(1-p) \ge 5 = 500(1-.024)$
$= 488$ ✓

$n \ge 30$, yes

Must have	Where
$np \ge 5$	$p = \frac{X}{n}$
$n(1-p) \ge 5$	X = no. of items of interest in sample
$n \ge 30$	

$$Z = \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$Z = \frac{p-p_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

$$= \frac{.024 - .01}{\sqrt{\frac{.01(1-.01)}{500}}} = \frac{.014}{\sqrt{\frac{.01(.99)}{500}}}$$

$$Z = \frac{.014}{\sqrt{\frac{.0099}{500}}} = \frac{.014}{\sqrt{.0000198}} = \frac{.014}{.0044497} = 3.146$$

pvalue of Z = .9991724 and we want area to right
so 1 − .9991724 = .0008
so A is the answer, strong evidence to reject Ho

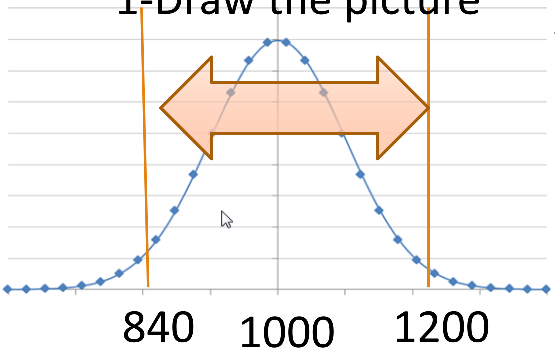
Upper/right-tail

p = area right of Z

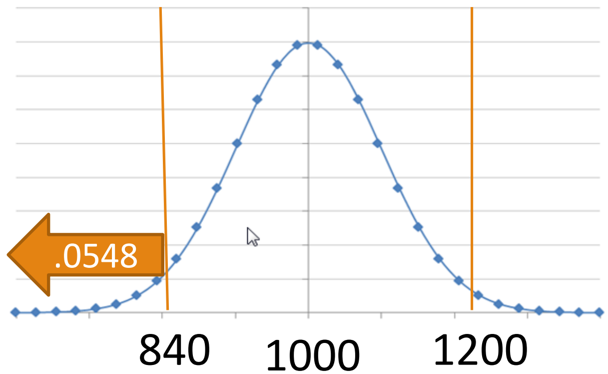
Quiz 2 Prep Question 1: Practice with Z calculations

The distribution of weekly incomes of supervisors at the ABC Company follows the normal distribution, with a mean of \$1000 and a standard deviation of \$100.
What percent of the supervisors have a weekly income between \$840 and \$1200?

1-Draw the picture



2-Think about what you are calculating related to the picture



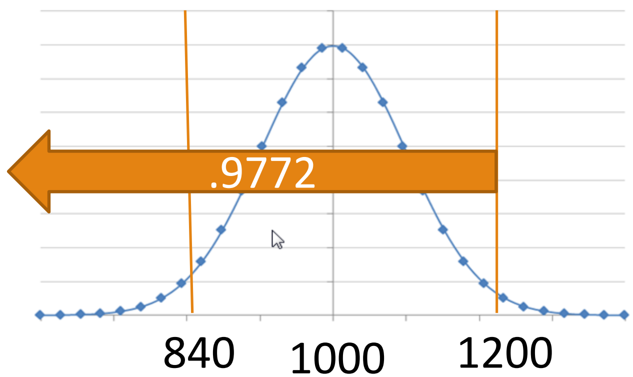
$$Z = \frac{X-\mu}{\sigma}$$

$$Z = \frac{840-1000}{100} = -1.6$$

Look up in tables, p = .0548

Or in Excel

=NORM.DIST(840,1000,100,TRUE)



$$Z = \frac{X-\mu}{\sigma}$$

$$Z = \frac{1200-1000}{100} = 2$$

Look up in tables, p = .9772

Or in Excel

=NORM.DIST(1200,1000,100,TRUE)

.9772-.0548 = .9224, so 92.24% have a weekly income between \$840 and \$1200

Table C Standard normal distribution (cont

Z	0.00	0.01	0.02
0.0	0.5000	0.5040	0.5080
0.1	0.5398	0.5438	0.5478
0.2	0.5793	0.5832	0.5871
0.3	0.6179	0.6217	0.6255
0.4	0.6554	0.6591	0.6628
0.5	0.6915	0.6950	0.6985
0.6	0.7257	0.7291	0.7324
0.7	0.7580	0.7611	0.7642
0.8	0.7881	0.7910	0.7939
0.9	0.8159	0.8186	0.8212
1.0	0.8413	0.8438	0.8461
1.1	0.8643	0.8665	0.8686
1.2	0.8849	0.8869	0.8888
1.3	0.9032	0.9049	0.9066
1.4	0.9192	0.9207	0.9222
1.5	0.9332	0.9345	0.9357
1.6	0.9452	0.9463	0.9474
1.7	0.9554	0.9564	0.9573
1.8	0.9641	0.9649	0.9656
1.9	0.9713	0.9719	0.9726
2.0	0.9772	0.9778	0.9783

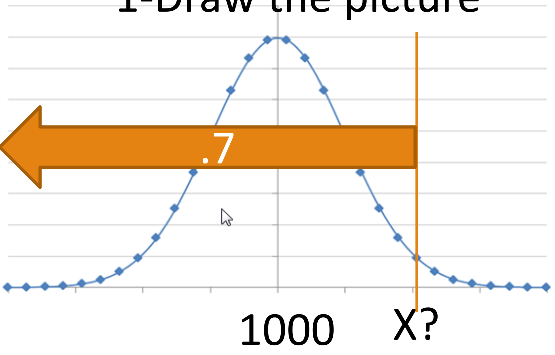
Table C Standard normal distribution

Z	0.00	0.01	0.02	0.03
–3.4	0.0003	0.0003	0.0003	0.0003
–3.3	0.0005	0.0005	0.0005	0.0004
–3.2	0.0007	0.0007	0.0006	0.0006
–3.1	0.0010	0.0009	0.0009	0.0009
–3.0	0.0013	0.0013	0.0013	0.0012
–2.9	0.0019	0.0018	0.0018	0.0017
–2.8	0.0026	0.0025	0.0024	0.0023
–2.7	0.0035	0.0034	0.0033	0.0032
–2.6	0.0047	0.0045	0.0044	0.0043
–2.5	0.0062	0.0060	0.0059	0.0057
–2.4	0.0082	0.0080	0.0078	0.0075
–2.3	0.0107	0.0104	0.0102	0.0099
–2.2	0.0139	0.0136	0.0132	0.0129
–2.1	0.0179	0.0174	0.0170	0.0166
–2.0	0.0228	0.0222	0.0217	0.0212
–1.9	0.0287	0.0281	0.0274	0.0268
–1.8	0.0359	0.0351	0.0344	0.0336
–1.7	0.0446	0.0436	0.0427	0.0418
–1.6	0.0548	0.0537	0.0526	0.0516

Quiz 2 Prep Question 1 Extended: Practice with Z calculations

The distribution of weekly incomes of supervisors at the ABC Company follows the normal distribution, with a mean of \$1000 and a standard deviation of \$100.
Management wants to give bonuses to those supervisors within the top 30% of weekly incomes. What is the weekly income cut off, the lowest weekly income a supervisor can have and still receive the bonus?

1-Draw the picture



2-Think about what you are calculating related to the picture

$$Z = \frac{X - \mu}{\sigma} \qquad Z = \frac{x - 1000}{100} = ?$$

We know p = .7, what Z goes with that p value…look it up from the table = .52 or in Excel=NORM.S.INV(0.7)=.52

$$.52 = \frac{x - 1000}{100}$$

$$.52 * 100 = X - 1000$$

52 = X – 1000

52 + 1000 = X

1052 = X

\$1052 is the lowest weekly income a supervisor can have and still receive a bonus

Quiz 2 Review

Highlights: Video Segment 3.5: Binomial(Discrete Data)

No set shape
Mean: μ=np, n = number of trials, p=probability of success
Variance: σ^2= n x p(1-p)

<u>A binomial experiment is an experiment which satisfies these four conditions:</u>
A fixed number of trials
Each trial is independent of the others
There are only two outcomes
The probability of each outcome remains constant from trial to trial.

Quiz 2 Prep

You're taking a quiz with five true/false questions. You didn't study and plan to guess. What's the probability you get three questions correct?

Find P(X = 3), the probability that the number of successes is equal to three.

- $n = 5$
- $p = 0.5$

Example: Binomial Table

		p (probability of a success)						
n	*X*	**0.10**	**0.15**	**0.20**	...	**0.40**	**0.45**	**0.50**
⋮	⋮				...			
4	0	0.6561	0.5220	0.4096		0.1296	0.0915	0.0625
	1	0.2916	0.3685	0.4096		0.3456	0.2995	0.2500
	2	0.0486	0.0975	0.1536		0.3456	0.3675	0.3750
	3	0.0036	0.0115	0.0256		0.1536	0.2005	0.2500
	4	0.0001	0.0005	0.0016		0.0256	0.0410	0.0625
5	0	0.5905	0.4437	0.3277	...	0.0778	0.0503	0.0312
	1	0.3280	0.3915	0.4096		0.2592	0.2059	0.1562
	2	0.0729	0.1382	0.2048		0.3456	0.3369	0.3125
	3	0.0081	0.0244	0.0512		0.2304	0.2757	0.3125

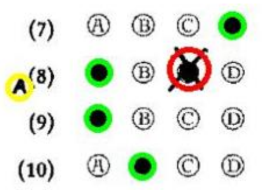
Use the tables in the back of the book or Excel to calculate probabilities for a binomial distribution for discrete data or use Excel: =binom.dist(3,5,.5,False) = .3125

Means give me the probability that I get 3 successes, out of 5 trials, when each has a probability of success of .5, and False means exactly 3 successes(True would mean all those probabilities up to and including 3 successes:1,2,3 successes all added together)

Quiz 2 Review

Highlights: Video Segment 3.5: Binomial(Discrete Data)

Another Example:



For a multiple choice test that you are guessing on, you want to know the probability you get at least 3 correct, on a test that has 5 multiple choice questions, and each has 4 choices.

n=5 test questions, p=.25(chance of answering correctly on each problem(1/4)), x>=3

=BINOM.DIST(3,5,0.25,FALSE)

This formula means, the probability that I get 3 questions correct, with 5 questions on the test, and 4 answers for each question so a ¼=.25 chance of getting each correct, and false means I don't want the cumulative percent because I won't pass the test if I get 1 or 2 correct. This will give me probability of getting exactly 3 correct, then I would do the same with the probability at 4 and 5 correct and add the three probabilities together – because I want to know the probably of getting at least 3 correct, which means 3 or more correct.

=BINOM.DIST(3,5,0.25,FALSE)= .088
=BINOM.DIST(4,5,0.25,FALSE)= .015
=BINOM.DIST(5,5,0.25,FALSE)= <u>.0010</u>
.104 or 10.4% probability that you get at least 3 correct

Or =BINOM.DIST(2,5,0.25,TRUE) means probability I get 1 or 2 correct, then subtract from 1 to get probability of 3,4,5 correct

=.896, so 1-.896=.1035 or .104, so 10.4% probability that you get at least 3 correct

Quiz 2 Prep Question 2:

Twenty percent of the employees of ABC Company use direct deposit and have their wages sent directly to the bank. Assume we random sample five employees. What is the probability that all five employees use direct deposit?

Solution:

n=5, trials is 5
p=.2, 20% chance they use direct deposit(yes or no)
X=5, we want to know the probability 5 employees used direct deposit
=Table B in your book, n=5,x=5,.2 column **= .0003 is the probability all 5 employees use direct deposit**

Or in Excel =BINOM.DIST(5,5,0.2,FALSE)
BINOM.DIST(successes, trials, probability, cumulative), false because we want exactly 5 successes not cumulative)

Quiz 2 Review

Table B Binomial distribution

		p								
n	***X***	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
2	0	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
	2	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
	2	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
	3	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
	2	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
	3	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
	4	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0312
	1	0.3280	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1562
	2	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
	3	0.0081	0.0244	0.0512	0.0879	0.1323	0.1811	0.2304	0.2757	0.3125
	4	0.0004	0.0022	0.0064	0.0146	0.0284	0.0488	0.0768	0.1128	0.1562
	5		0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0312
6	0	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0938
	2	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344
	3	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	[illegible]	0.3125
	4	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344
	5	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938
	6			0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156
7	0	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.3720	0.3960	0.3670	0.3115	0.2471	0.1848	0.1306	0.0872	0.0547
	2	0.1240	0.2097	0.2753	0.3115	0.3177	0.2985	0.2613	0.2140	0.1641
	3	0.0230	0.0617	0.1147	0.1730	0.2269	0.2679	0.2903	0.2918	0.2734
	4	0.0026	0.0109	0.0287	0.0577	0.0972	0.1442	0.1935	0.2388	0.2734
	5	0.0002	0.0012	0.0043	0.0115	0.0250	0.0466	0.0774	0.1172	0.1641
	6		0.0001	0.0004	0.0013	0.0036	0.0084	0.0172	0.0320	0.0547
	7				0.0001	0.0002	0.0006	0.0016	0.0037	0.0078
8	0	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.3826	0.3847	0.3355	0.2670	0.1977	0.1373	0.0896	0.0548	0.0312
	2	0.1488	0.2376	0.2936	0.3115	0.2965	0.2587	0.2090	0.1569	0.1094
	3	0.0331	0.0839	0.1468	0.2076	0.2541	0.2786	0.2787	0.2568	0.2188
	4	0.0046	0.0185	0.0459	0.0865	0.1361	0.1875	0.2322	0.2627	0.2734
	5	0.0004	0.0026	0.0092	0.0231	0.0467	0.0808	0.1239	0.1719	0.2188
	6		0.0002	0.0011	0.0038	0.0100	0.0217	0.0413	0.0703	0.1094
	7			0.0001	0.0004	0.0012	0.0033	0.0079	0.0164	0.0313
	8					0.0001	0.0002	0.0007	0.0017	0.0039

Note: Blank entries indicate a binomial probability of less than 0.00005.

Agenda

Topic	Time	Wednesday Section
Introduction	5 min	9:00-9:05
Homework #2 Chapter 9 Quiz Quiz #2 Prep Review	35 min	9:05-9:40
Highlights from Week 5 Video	30 min	9:40-10:10
Breakout on Additional Example	10 min	10:10-10:20
Review of Upcoming Assignments and Open Question	10 min	10:20-10:30

Highlights: Video Segment 5.3:Confidence Intervals for Continuous Data

Another Statistical Inference Method

1

Drawing a conclusion about a population

Takes into account the *natural variability* in the data

...based on a sample

Reasons to sample:

Too time consuming

Too expensive

May require destruction

A **confidence interval** is an estimate of a parameter consisting of an interval of numbers based on a point estimate, together with a **confidence level** specifying the probability that the interval contains the parameter.

2

How Is a Confidence Interval Useful?

3

- Estimates an unknown population parameter (e.g., mean, standard deviation, variance, proportion)
- Gives an indication of how accurate that estimate is
- Also indicates how confident we are that the results are correct

What Is a Confidence Interval?

4

Confidence interval: a range of values (from sample data) in which we expect the population parameter to occur

Parameter estimate ± margin of error (E)

↓ $\bar{x}, s, p$ (estimates of population parameters)

↓ margin of error (E)

Highlights: Video Segment 5.3:Confidence Intervals for Continuous Data

Confidence Interval Formulas for Mean

Boundaries of interval (two sided)	When population standard deviation is *known* (not often)	When population standard deviation is *unknown* and sample size n is *large* (≥ 30)	When population standard deviation is *unknown* and sample size n is *small* (< 30)
Upper confidence limits for μ	$U = \bar{x} + z^* \frac{\sigma}{\sqrt{n}}$	$U = \bar{x} + z^* \frac{s}{\sqrt{n}}$	$U = \bar{x} + t \frac{s}{\sqrt{n}}$
Lower confidence limits for μ	$L = \bar{x} - z^* \frac{\sigma}{\sqrt{n}}$	$L = \bar{x} - z^* \frac{s}{\sqrt{n}}$	$L = \bar{x} - t \frac{s}{\sqrt{n}}$
			$df = n - 1$

- In all cases, assume a normal distribution.
- Find z^* and t-values in Table D, p. T-11.

Note: The Z value is for $\frac{\alpha}{2}$

Margin of Error

Confidence intervals for the population mean μ take the form:

point estimate $\pm$ margin of error E

The **margin of error *E*** is a measure of the precision of the confidence interval estimate. For the Z interval, the margin of error takes the form $E = Z_{\alpha/2}(\sigma/\sqrt{n})$.

In our example, $E = Z_{\alpha/2}(\sigma/\sqrt{n}) = 32.9$. Therefore, the confidence interval has the form:

$$510 \pm 32.9$$

We would like our confidence interval estimates to be as precise as possible. Therefore, we would like the margin of error to be as small as possible. There are two strategies to decrease *E:*

- ***Decrease the confidence level***
- ***Increase the sample size***

Highlights: Video Segment 5.4 Confidence Interval Example

Example: Hank's Process

- Find 95% confidence interval about the mean
  - Need lower and upper limits
- Use data given by example:
  - $n = 30$
  - $\bar{x} = 17.23$
  - $s = 4.52$
- Assume $\alpha = 0.05$

When population standard deviation is *unknown* and sample size n is *large* (≥ 30)

$$U = \bar{x} + z^* \frac{s}{\sqrt{n}}$$

$$L = \bar{x} - z^* \frac{s}{\sqrt{n}}$$

$$U \,\&\, L = \bar{x} \pm z^* \frac{s}{\sqrt{n}} = 17.23 \pm (1.96) \frac{4.52}{\sqrt{30}}$$

Highlights: Video Segment 5.4 Confidence Interval Example

Confidence Interval for Hank's Process

$$U \& L = \bar{x} \pm z^* \frac{s}{\sqrt{n}} = 17.23 \pm (1.96)\frac{4.52}{\sqrt{30}}$$

- Use data given by example: $n = 30$; $\bar{x} = 17.23$; $s = 4.52$.
- Plus/minus gives margin of error above and below the parameter estimate.
- In Table D, find z^* at 95% confidence ($p = 0.95$).

OR

=CONFIDENCE.NORM(alpha, std dev, sample size)

=CONFIDENCE.NORM(0.05,4.52,30)
=1.617432 (the margin or error)

=17.23(+ or –) 1.62

= 15.61 to 18.85

How to get the 1.96:
1.96 from last row of table D under 95% confidence or
enter in Excel=NORM.S.INV(0.975) or
lookup .975 inside the z table
it is .975 because you want + or – so you need to split the .05 in half which then equals 1-(.05/2) = .975

- We are 95% confident that the mean of the population (μ_{pop}) is between 15.61 and 18.85 days.

Highlights: Video Segment 5.4 Confidence Interval Example

Table D *t*-Distribution

		Confidence level				
		80%	90%	95%	98%	99%
		Area in one tail				
		0.10	0.05	0.025	0.01	0.005
		Area in two tails				
		0.20	0.10	0.05	0.02	0.01
df	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750
	31	1.309	1.696	2.040	2.453	2.744
	32	1.309	1.694	2.037	2.449	2.738
	33	1.308	1.692	2.035	2.445	2.733
	34	1.307	1.691	2.032	2.441	2.728
	35	1.306	1.690	2.030	2.438	2.724
	36	1.306	1.688	2.028	2.435	2.719
	37	1.305	1.687	2.026	2.431	2.715
	38	1.304	1.686	2.024	2.429	2.712
	39	1.304	1.685	2.023	2.426	2.708
	40	1.303	1.684	2.021	2.423	2.704
	50	1.299	1.676	2.009	2.403	2.678
	60	1.296	1.671	2.000	2.390	2.660
	70	1.294	1.667	1.994	2.381	2.648
	80	1.292	1.664	1.990	2.374	2.639
	90	1.291	1.662	1.987	2.368	2.632
	100	1.290	1.660	1.984	2.364	2.626
	1000	1.282	1.646	1.962	2.330	2.581
	z	1.282	1.645	1.960	2.326	2.576

Z Star

i.e. 95% confident = 1.96

Highlights: Video Segment 5.4 Confidence Interval Example

Regular Z table

Must use 1 - α/2

i.e. 95% confident,

α is .05, .05/2 = .025

1-.025 = .975

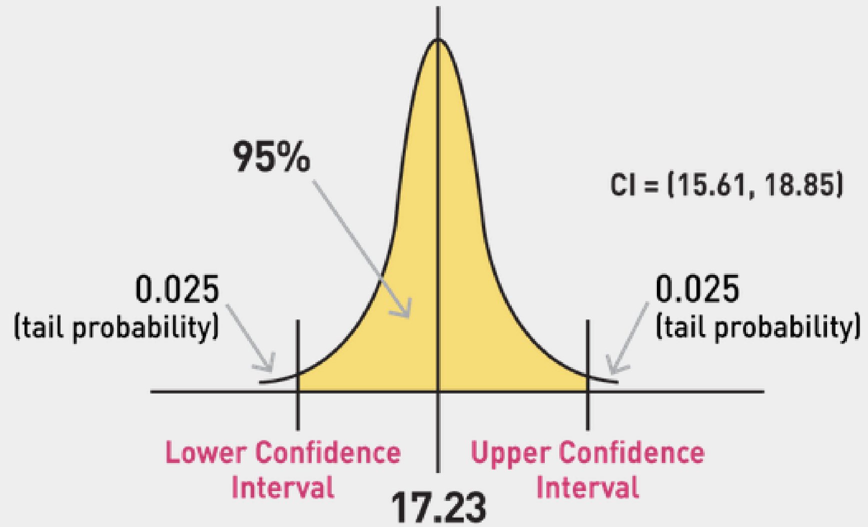
Z value = 1.96

Table C Standard normal distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Highlights: Video Segment 5.4
Confidence Interval Example

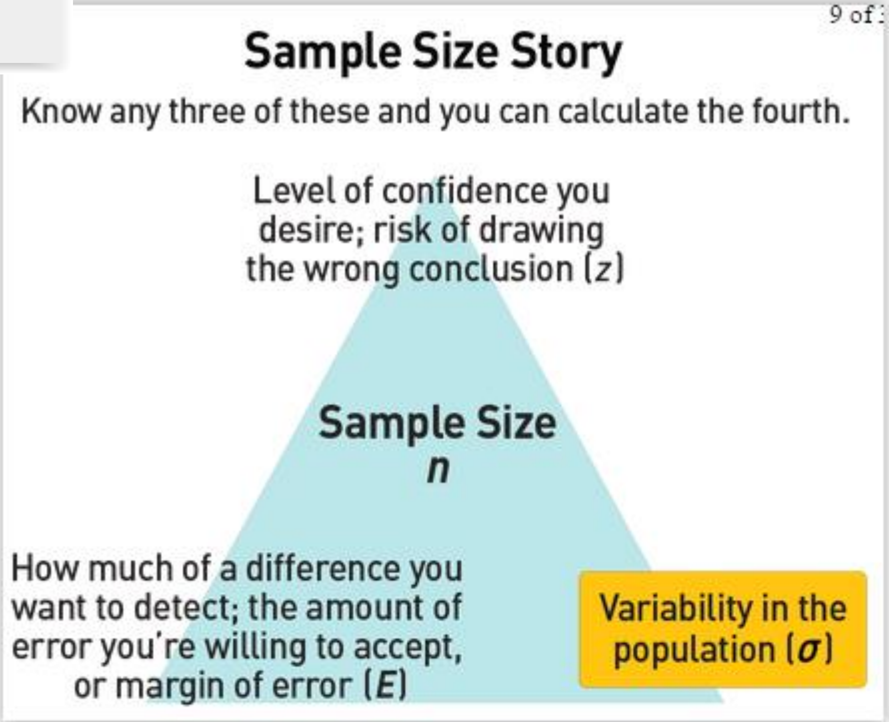
Other Ways to Interpret This CI



- The above distribution is a distribution of sample means, so the probability of the true mean being within the confidence interval is 95%.
- 95% of all $\bar{x}$s (calculated from samples) will fall into the shaded region.

Highlights: Video Segment 5.5 Sample Size for Continuous

The only way to have both high confidence and a tight interval is to increase sample size.



Note: Sample Size Calculation for projects

Highlights: Video Segment 5.5 Sample Size for Continuous

Sample Size Formula for Continuous Data

$$n = \left(\frac{z^* \hat{\sigma}}{E}\right)^2$$

Example: Time to Complete Job

- Suppose you have collected a simple random sample of data and found the standard deviation to be three minutes.
- How many samples are needed to detect a change in job completion time after a process improvement project is implemented?
  - You are okay with a margin of error of two minutes.
  - Assume you want 95% confidence.

Example: Time to Complete Job (cont.)

$$n = \left(\frac{1.96(3)}{2}\right)^2$$

$$= 8.6 \cong 9$$

- z^* at 95% confidence = 1.96
- $\hat{\sigma} = 3$
  - Estimated population standard deviation
  - Equivalent to sample standard deviation, s
- $E = 2$

Video Sl

Highlights: Video Segment 5.6 Sample Size for Discrete Data

Sample Size for Discrete Data

$$n = \frac{(z^*)^2 \, p(1-p)}{E^2}$$

Highlights: Video Segment 5.6 Confidence Interval and Sample Size for Discrete Data

CI Example: Candidate A Voters

- Of a sample of 300 voters, 164 want to vote for Candidate A.
- Find the 99% confidence interval for the proportion of voters planning to vote for Candidate A.

Confidence Interval for Discrete Data

For a population proportion (e.g., cosmetic defect):

$$U = p + z^* \sqrt{\frac{p(1-p)}{n}}$$

$$L = p - z^* \sqrt{\frac{p(1-p)}{n}}$$

- Upper and lower confidence limits for p
- Where:
  - p = sample proportion (e.g., percent defective)
  - n = sample size

CI Example: Candidate A Voters (cont.)

$$\begin{aligned} U \,\&\, L &= p \pm z^* \sqrt{\frac{p(1-p)}{n}} \\ &= 0.547 \pm 2.576\sqrt{\frac{0.547(1-0.547)}{300}} \\ &= 0.547 \pm 0.077 \end{aligned}$$

- $p = 164/300 = 54.7\%$
- z^* at 99% = 2.576 (at the bottom of p. T-11)
- $n = 300$

We are 99% confident that the true percentage of voters for Candidate A is between 47.0% and 62.4% (i.e., $0.47 \le \pi \le 0.624$).

Note: ALWAYS round up

Highlights: Video Segment 5.6 Exercises

You would like to start a new business providing Internet service and need to estimate the average Internet usage of households during one week for your business plan. How many households must you select to be 95 percent sure that the sample mean is within one minute (*E* = 1) of the population mean? Assume a previous survey of household usage has shown that sigma = 6.95 minutes.

Sample Size Formula for Continuous Data

$$n = \left(\frac{z^* \hat{\sigma}}{E}\right)^2$$

n= $(\frac{1.96\ X\ 6.95}{1})^2$

n = 185.55, round up to a whole number for a sample

n = 186

Note: T versus Z based on sample size

Highlights: Video Segment 5.6 Exercises

You are interested in how long it takes to get your food at a takeout restaurant (the time it takes from placing your order to when the food arrives). Over the next month, you decide to sample 20 times to determine the actual average. You find that xbar = 15.8 minutes and *s* = 2.5 minutes. Find the 95 percent confidence interval for the true mean.

When population standard deviation is *unknown* and sample size *n* is *small* (< 30)

$$U = \bar{x} + t\frac{s}{\sqrt{n}}$$

$$L = \bar{x} - t\frac{s}{\sqrt{n}}$$

$df = n - 1$

Df = n-1 = 20-1 = 19

Find t value look at df = 19 and under the 95% confidence level column = 2.093

$$15.8 \pm 2.093\left(\frac{2.5}{\sqrt{20}}\right)$$

$$15.8 \pm 2.093\left(\frac{2.5}{4.472}\right)$$

$$15.8 \pm 2.093(.559)$$

$$15.8 \pm 1.17 = 14.63 \text{ to } 16.97$$

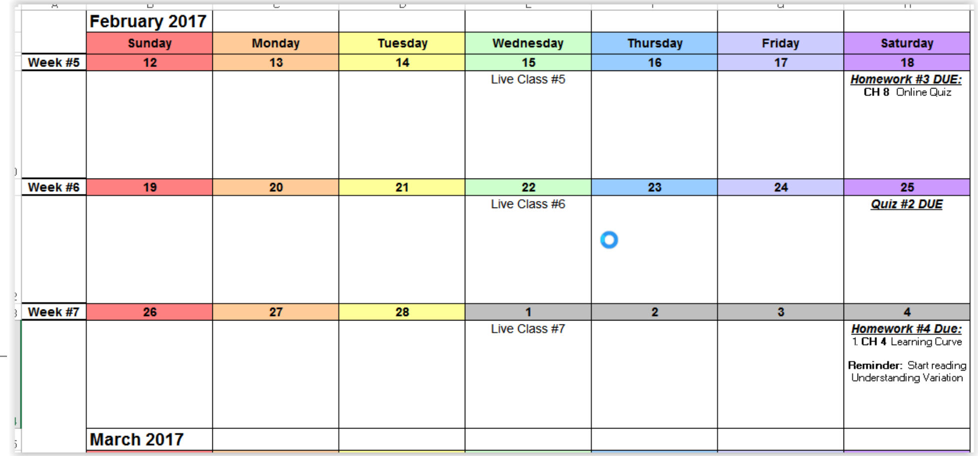
Or in Excel use =CONFIDENCE.T(0.05,2.5,20)=1.17 then +- to 15.8 to get interval

Confidence.t(alpha, std deviation, sample)

Agenda

Topic	Time	Wednesday Section
Introduction	5 min	9:00-9:05
Homework #2 Chapter 9 Quiz Quiz #2 Prep Review	35 min	9:05-9:40
Highlights from Week 5 Video	30 min	9:40-10:10
Breakout on Additional Example	10 min	10:10-10:20
Review of Upcoming Assignments and Open Question	10 min	10:20-10:30

Review of Upcoming Assignments: Wednesday

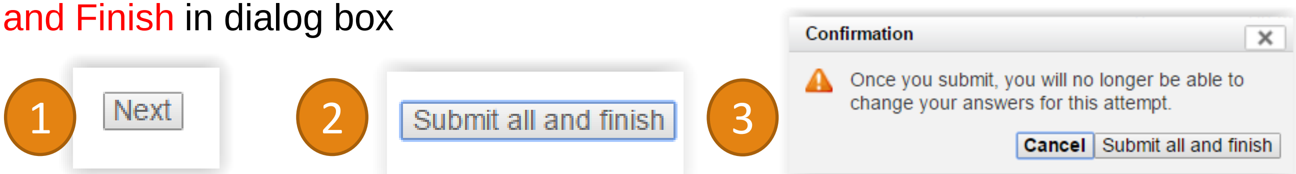


	February 2017						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Week #5	12	13	14	15	16	17	18
				Live Class #5			Homework #3 DUE: CH 8 Online Quiz
Week #6	19	20	21	22	23	24	25
				Live Class #6			Quiz #2 DUE
Week #7	26	27	28	1	2	3	4
				Live Class #7			Homework #4 Due: 1. CH 4 Learning Curve Reminder: Start reading Understanding Variation
	March 2017						

1. Homework #3 due ,2/18 midnight EST in LaunchPad, Ch 8 Online Practice Quiz

2. Quiz #2 Prep file and answers are uploaded– optional

3. Quiz #2 is 5 calculation questions and 1, 10 part definitional question, due , 2/25 midnight EST, in the learning management system
   - DO NOT LEAVE ANY BLANK
   - You can not start and stop the Quiz.
   - There is a 2 hr. time limit. There is no timer, you must keep track of your own time.
   - Password for the Quiz is: TestTime101
   - At the end must click, NEXT at the bottom of the questions, then click Submit all and Finish, then click AGAIN Submit all and Finish in dialog box



4. Projects
   - Measure should be wrapping up/wrapped up
   - Analyze should be started
   - Sample size calculation would be a good add to your measure/analyze phase to make sure you have enough data or to let you know your risk

Reference

Z Interval

When a random sample of size *n* is taken from a population,
a (100 – *α*)% confidence interval is given by:

$$\text{lower bound} = \bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n})$$

$$\text{upper bound} = \bar{x} + Z_{\alpha/2}(\sigma/\sqrt{n})$$

The *Z* interval can also be written as $\bar{x} \pm Z_{\alpha/2}(\sigma/\sqrt{n})$

α	Confidence Level	$\alpha/2$	$Z_{\alpha/2}$
0.10	90%	0.05	1.645
0.05	95%	0.025	1.96
0.01	99%	0.005	2.576

t Interval

When a random sample of size n is taken from a population, a (100 – α)% confidence interval is given by:

$$\text{lower bound} = \bar{x} - t_{\alpha/2}(s/\sqrt{n})$$

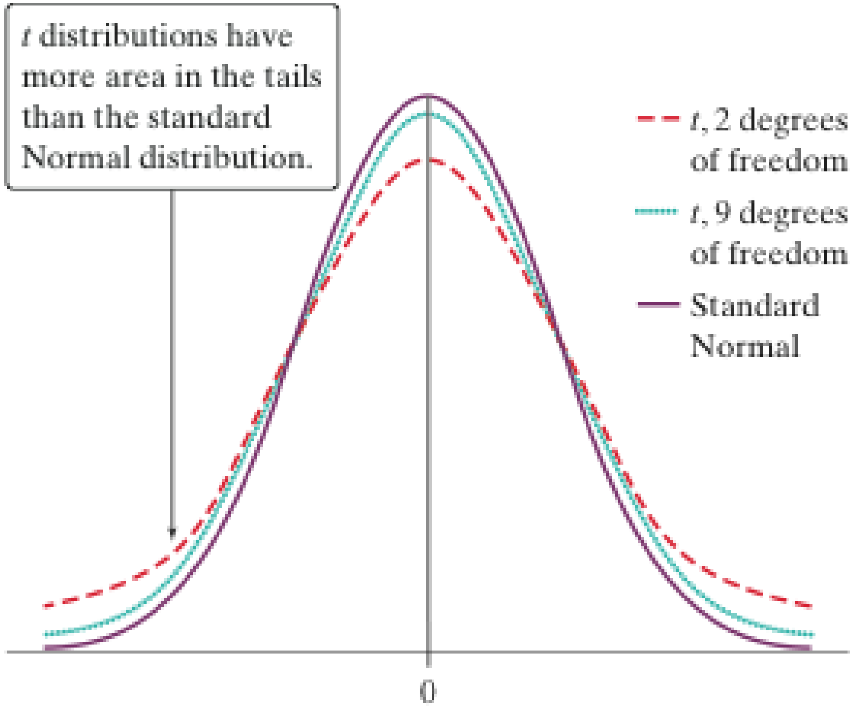
$$\text{upper bound} = \bar{x} + t_{\alpha/2}(s/\sqrt{n})$$

The t interval can also be written as: $\bar{x} \pm t_{\alpha/2}(s/\sqrt{n})$

For the t interval, the margin of error takes the form $E = t_{\alpha/2}(s/\sqrt{n})$. The interval can be interpreted as "We can estimate μ to within E units with (1 – α)% confidence."

Characteristics of the *t* Distribution

- Centered at zero. The mean of *t* is 0, just as with *Z*.
- Symmetric about its mean 0, just as with *Z*.
- As df decreases, the *t* curve gets flatter, and the area under the *t* curve decreases in the center and increases in the tails.
  - That is, the *t* curve has heavier tails than the *Z* curve.
- As df increases toward infinity, the *t* curve approaches the *Z* curve.

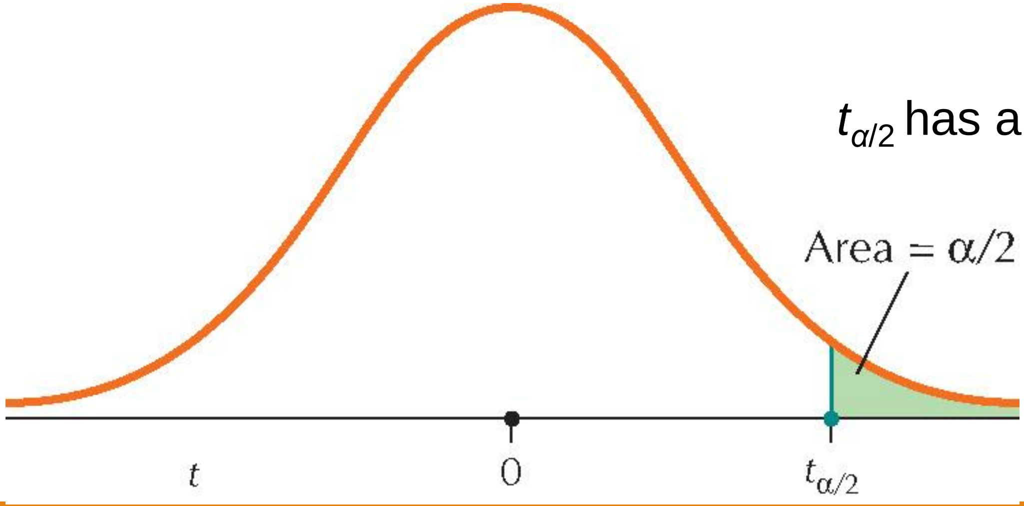


Finding $t_{\alpha/2}$

Step 1: Go across the row marked "Confidence level" in the *t* table until you find the column with the desired confidence level at the top. The $t_{\alpha/2}$ value is in this column somewhere.

Step 2: Go down the column until you see the correct number of degrees of freedom on the left. The number in that row and column is the desired value of $t_{\alpha/2}$.

$t_{\alpha/2}$ has area $\alpha/2$ to the right of it.



Example

Find the value of $t_{\alpha/2}$ that will produce a 95% confidence interval for μ if the sample size is $n = 20$.

Step 1
Go across the row labeled "Confidence level" in the *t* table until we see the 95% confidence level.

Step 2
df = $n - 1 = 20 - 1 = 19$

Go down the column until you see 19 on the left.

The number in that row is $t_{\alpha/2} = 2.093$.

t-Distribution

	Confidence level				
	80%	**90%**	**95%**	**98%**	**99%**
	Area in one tail				
	0.10	**0.05**	**0.025**	**0.01**	**0.005**
	Area in two tails				
	0.20	**0.10**	**0.05**	**0.02**	**0.01**
df 1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831

Calculate a Point Estimate

Recall that characteristics of the sample are called **statistics,** while characteristics of the population are called **parameters.** We have dealt with interval estimates of *μ,* but we may also be interested in the interval estimate for the population proportion of successes, *p*.

$$\hat{p} = \frac{x}{n} = \frac{\text{number of successes}}{\text{sample size}}$$

is a point estimate of the population proportion *p*.

Z Interval for the Population Proportion *p*

Recall the Central Limit Theorem for Proportions

Central Limit Theorem for Proportions
The sampling distribution of the sample proportion follows an approximately normal distribution with mean *p* and standard deviation

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

when both $np \geq 5$ and $n(1-p) \geq 5$.

We can use the CLT for Proportions to construct confidence intervals for the population proportion *p*.

Z Interval for *p*

Z Interval for the Population Proportion p

The Z interval for p may be constructed only when both of the following two conditions are met: $n(p\text{-}hat) \geq 5$ and $n(1 - p\text{-}hat) \geq 5$.

When a random sample of size n is taken from a population, a $(100 - \alpha)\%$ confidence interval is given by:

$$\text{lower bound} = \hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\text{upper bound} = \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The Z interval can also be written as: $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

α	Confidence Level	$\alpha/2$	$Z_{\alpha/2}$
0.10	90%	0.05	1.645
0.05	95%	0.025	1.96
0.01	99%	0.005	2.576

Margin of Error

Confidence intervals for the population proportion *p* take the form:

point estimate ± margin of error *E*

The **margin of error *E*** is a measure of the precision of the confidence interval estimate. For the *Z* interval, the margin of error takes the form:

$$E = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$$

The margin of error *E* for a (1– α)% Z interval for *p* can be interpreted as follows:

- "We can estimate *p* to within *E* with (1 – α)100% confidence."

Example

There is hardly a day that goes by without some new poll coming out. Especially during election campaigns, polls influence the choice of candidates and the direction of their policies. In October 2004, the Gallup organization polled 1012 American adults, asking them, "Do you think there should or should not be a law that would ban the possession of handguns, except by the police and other authorized persons?" Of the 1012 randomly chosen respondents, 638 said that there should NOT be such a law.

a. Check that the conditions for the Z interval for p have been met.

b. Find and interpret the margin of error E.

c. Construct and interpret a 95% confidence interval for the population proportion of all American adults who think there should not be such a law.

Solution

- Sample size is n = 1012
- Observed proportion is: $$\hat{p} = \frac{638}{1012} \approx 0.63$$

$$n\hat{p} = (1012)\cdot(0.63) = 637.56 \geq 5$$

$$n(1-\hat{p}) = (1012)\cdot(0.37) = 374.44 \geq 5$$

- The confidence level of 95% implies that our $Z_{\alpha/2}$ equals 1.96.

$$E = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \cdot \sqrt{\frac{0.63(0.37)}{1012}} \approx 0.03$$

Solution

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$$

$$= \hat{p} \pm E$$

$$= 0.63 \pm 0.03$$

$$= (\text{lower bound } \ 0.60, \text{ upper bound } \ 0.66)$$

Thus, we are 95% confident that the population proportion of all American adults who think that there should not be such a law lies between 60% and 66%.

Sample Size for Estimating *p*

A natural question when constructing a confidence interval is "*How large a sample size do I need* to get a tight confidence interval with a high confidence level?"

Sample Size for Estimating the Population Proportion

The sample size for a Z interval that estimates p to within a margin of error E with confidence $100(1 - \alpha)\%$ is given by:

$$n = \hat{p}(1 - \hat{p})\left(\frac{Z_{\alpha/2}}{E}\right)^2$$

Whenever this formula yields a sample size with a decimal, *always round up to the next whole number*.

When *p-hat* is unknown, use $$n = \left(\frac{0.5 \cdot Z_{\alpha/2}}{E}\right)^2$$