

Confidence Intervals – 2 sample Means

1. Learning math. The Core Plus Mathematics Project (CPMP) is an innovative approach to teaching mathematics that engages students in group investigations and mathematical modeling. After field tests in 36 high schools over a three-year period, researchers compared the performances of CPMP students with those taught using a traditional curriculum. In one test, students had to solve applied algebra problems using calculators. Scores for 320 CPMP students were compared with those of a control group of 273 students in a traditional math program. Computer software was used to create a confidence interval for the difference in mean scores. Conf level: 95% Variable: $\mu(\text{CPMP}) - \mu(\text{Ctrl})$

Interval: (5.573, 11.427]

- What's the margin of error for this confidence interval?
- If we had created a 98% CI, would the margin of error be larger or smaller?
- Explain what the calculated interval means in this context.
- Does this result suggest that students who learn mathematics with CPMP will have significantly higher mean scores in algebra than those in traditional programs? Explain.

2. Stereograms. Stereograms appear to be composed entirely of random dots. However, they contain separate images that a viewer can "fuse" into a three-dimensional (3D) image by staring at the dots while defocusing the eyes. An experiment was performed to determine whether knowledge of the form of the embedded image affected the time required for subjects to fuse the images. One group of subjects (group NV) received no information or just verbal information about the shape of the embedded object. A second group (group W) received both verbal information and visual information (specifically, a drawing of the object). The experimenters measured how many seconds it took for the subject to report that he or she saw the 3D image.

2-Sample t-interval for $\mu_1 - \mu_2$ df = 70

Conf level = 90%

$\mu(\text{NV}) - \mu(\text{W})$ interval: (0.55, 5.47)

- Interpret your interval in context.
- Does it appear that viewing a picture of the image helps people "see" the 3D image in a stereogram?
- What's the margin of error for this interval?
- Explain carefully what the 90% confidence level means.
- Would you expect a 99% confidence level to be wider or narrower? Explain.
- Might that change your conclusion in part b? Explain.

3. Temperatures. The table below gives the average high temperatures in January and July for several European cities. Write a 90% confidence interval for the mean temperature difference between summer and winter in Europe. Be sure to check conditions for inference, and clearly explain what your interval means.

City	Jan.	July
Vienna	34	75
Copenhagen	36	72
Paris	42	76
Berlin	35	74
Athens	54	90
Rome	54	88
Amsterdam	40	69
Madrid	47	87
London	44	73
Edinburgh	43	65
Moscow	21	76
Belgrade	37	84

4. BST. Many dairy cows now receive injections of BST, a hormone intended to spur greater milk production. After the first injection, a test herd of 60 Ayrshire cows increased their mean daily production from 47 pounds to 61 pounds of milk. The standard deviation of the increases was 5.2 pounds. We want to estimate the mean increase a farmer could expect in his own cows.

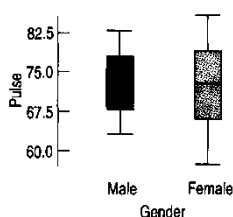
- Check the assumptions and conditions for inference.
- Write a 95% confidence interval.
- Explain what your interval means in this context.
- Given the cost of BST, a farmer believes he cannot afford to use it unless he is sure of attaining at least a 25% increase in milk production. Based on your confidence interval, what advice would you give him?

5. Commuting. A man who moves to a new city sees that there are two routes he could take to work. A neighbor who has lived there a long time tells him Route A will average 5 minutes faster than Route B. The man decides to experiment. Each day he flips a coin to determine which way to go, driving each route 20 days. He finds that Route A takes an average of 40 minutes with standard deviation 3 minutes, and Route B takes an average of 43 minutes with standard deviation 2 minutes. Histograms of travel times for the routes are roughly symmetric and show no outliers.

- Find a 95% confidence interval for the difference in average commuting time for the two routes.
- Should the man believe the old-timer's claim that he can save an average of 5 minutes a day by always driving Route A? Explain.

6. Pulse rates. A researcher wanted to see whether there is a significant difference in resting pulse rates for men and women. The data she collected are displayed in the box-plots and summarized below.

Gender	Male	Female
Count	28	24
Mean	72.75	72.625
Median	73	73
StdDev	5.37225	7.69987
Range	20	29
IQR	9	12.5



- What do the boxplots suggest about any gender differences in pulse rates?
- Is it appropriate to analyze these data using the methods of inference discussed in this chapter? Explain.
- Create a 90% confidence interval for the difference in mean pulse rates.
- Does the confidence interval confirm your answer to part a? Explain.

7. Cereal. The data below show the sugar content (as a percentage of weight) of several national brands of children's and adults' cereals. Create and interpret a 95% confidence interval for the difference in mean sugar content. Be sure to check the necessary assumptions and conditions.

Children's cereals: 40.3, 55, 45.7, 43.3, 50.3, 45.9, 53.5, 43, 44.2, 44, 47.4, 44, 33.6, 55.1, 48.8, 50.4, 37.8, 60.3, 46.6

Adults' cereals: 20, 30.2, 2.2, 7.5, 4.4, 22.2, 16.6, 14.5, 21.4, 3.3, 6.6, 7.8, 10.6, 16.2, 14.5, 4.1, 15.8, 4.1, 2.4, 3.5, 8.5, 10, 1, 4.4, 1.3, 8.1, 4.7, 18.4

8. Egyptians. Some archaeologists theorize that ancient Egyptians interbred with several different immigrant populations over thousands of years. To see if there is any indication of changes in body structure that might have resulted, they measured 30 skulls of male Egyptians dated from 4000 B.C.E. and 30 others dated from 200 B.C.E.

- Are these data appropriate for inference? Explain.
- Create a 95% confidence interval for the difference in mean skull breadth between these two eras.
- Do these data provide evidence that the mean breadth of males' skulls changed over this time period? Explain.

Maximum Skull Breadth			
4000 B.C.E	4000 B.C.E.	200 B.C.E.	200 B.C.E.
131	131	141	131
125	135	141	129
131	132	135	136
119	139	133	131
136	132	131	139
138	126	140	144
139	135	139	141
125	134	140	130
131	128	138	133

134	130	132	138
129	138	134	131
134	128	135	136
126	127	133	132
132	131	136	135
141	124	134	141

9. Reading. An educator believes that new reading activities for elementary school children will improve reading comprehension scores. She randomly assigns third graders to an eight-week program in which some will use these activities and others will experience traditional teaching methods. At the end of the experiment, both groups take a reading comprehension exam. Their scores are shown in the back-to-back stem-and-leaf display. Do these results suggest that the new activities are better? Find an appropriate confidence interval and state your conclusion.

New Activities	Control
	1 07
4	2 068
3	3 377
96333	4 122238
9876432	5 355
721	6 02
1	7
	8 5

10. Marathons. Shown are the winning times (in minutes) for men and women in the New York City Marathon between 1978 and 1998. Assuming that performances in the Big Apple resemble performances elsewhere, we can think of these data as a sample of performance in marathon competitions. Create a 90% confidence interval for the mean difference in winning times for male and female marathon competitors. (Chance, 12, no. 4, [1999])

Year	Men	Women
1978	132.2	152.5
1979	131.7	147.6
1980	129.7	145.7
1981	128.2	145.5
1982	129.5	147.2
1983	129.0	147.0
1984	134.9	149.5
1985	131.6	148.6
1986	131.1	148.1
1987	131.0	150.3
1988	128.3	148.1
1989	128.0	145.5
1990	132.7	150.8
1991	129.5	147.5
1992	129.5	144.7
1993	130.1	146.4
1994	131.4	147.6
1995	131.0	148.1
1996	129.9	148.3
1997	128.2	148.7
1998	128.8	145.3

- 11. Push-ups.** Every year the students at Gossett High School take a physical fitness test during their gym classes. One component of the test asks them to do as many push-ups as they can. Results for one class are shown below, according to gender. Assuming that students at Gossett are assigned to gym classes at random, create a 90% confidence interval for how many more pushups boys can do than girls, on average, at that high school.

Boys	17	27	31	17	25	32	28	23	25	16	11	34
Girls	24	7	14	16	2	15	19	25	10	27	31	8

- 12. Exercise.** An August 2001 article in the journal Medicine and Science in Sports and Exercise compared how long it would take men and women to burn 200 calories during light or heavy workouts on various kinds of exercise equipment. The results summarized in the table are the average times for a group of physically active young men and women whose performances were measured on each type of equipment.

- On average, how many minutes longer than a man must a woman exercise at a light exertion rate in order to burn 200 calories? Give a 95% confidence interval.
- Estimate the average number of minutes longer a woman must work out at a light exertion than at heavy exertion to get the same benefit. Give a 95% confidence interval.
- These data are actually averages rather than individual times. How might this affect the margins of error in these confidence intervals?

Average Minutes to Burn 200 Calories

	Hard Exertion		Light Exertion	
	Men	Women	Men	Women
Treadmill	12	17	14	22
X-C skier	12	16	16	23
Stair climber	13	18	20	37
Rowing machine	14	16	21	25
Exercise rider	22	24	27	36
Exercise bike	16	20	29	44

- 13. Baseball.** American League baseball teams play their games with the designated hitter rule, meaning that pitchers do not bat. The league believes that replacing the pitcher, traditionally a weak hitter, with another player in the batting order produces more runs and generates more interest among fans. Below are the average numbers of runs scored in American League and National League stadiums for the first half of the 2001 season.

American			National		
11.1	10.8	10.8	14.0	11.6	10.4
10.3	10.3	10.1	10.3	10.2	9.5
10.0	9.5	9.4	9.5	9.5	9.5
9.3	9.2	9.2	9.1	8.8	8.4
9.0	8.3		8.3	8.2	8.1
					7.9

- Create an appropriate display of these data. What do
- With a 95% confidence interval, estimate the mean number of runs scored in American League games.
- What concerns do you have about making a similar confidence interval for National League games? What could you do?
- Coors Field, in Denver, stands a mile above sea level, an altitude far greater than that of any other major league ball park. Some believe that the thinner air makes it harder for pitchers to throw curve balls and easier for batters to hit the ball a long way. Do you think the 14 runs scored per game at Coors are unusual? Explain.
- Explain why you should not use two separate confidence intervals to decide whether the two leagues differ in average number of runs scored.

- 14. Handy.** A factory hiring people to work on an assembly line gives job applicants a test of manual agility. This test counts how many strangely shaped pegs the applicant can fit into matching holes in a one-minute period. The table below summarizes the data by gender of the job applicant. Assume that all conditions necessary for inference are met.

	Male	Female
Number of subjects	50	50
Pegs placed:		
Mean	19.39	17.91
Std Dev	2.52	3.39

- Find 95% confidence intervals for the average number of pegs that males and females can each place.
- Those intervals overlap. What does this suggest about any gender-based difference in manual agility?
- Find a 95% confidence interval for the difference in the mean number of pegs that could be placed by
- What does this interval suggest about any gender-based difference in manual agility?
- The two results seem contradictory. Which method is correct: doing two-sample inference, or doing one-sample inference twice?
- Why don't the results agree?

- 15. Double header.** Do the data in Exercise 13 suggest that the American's designated hitter rule lead to more runs?

- Using a 95% confidence interval, estimate the difference between the mean number of runs scored in American and National League games.
- Interpret your interval.
- Does that interval suggest that the two leagues may differ in average number of runs scored per game?
- Does omitting the 14 runs scored per game at Coors Field affect your decision?

- 16. Braking.** In a test of braking performance, a tire manufacturer measured the stopping distance for one of its tire models. On a test track, a car made repeated stops from 60 miles per hour. The test was run on both dry and wet pavement, with results as shown in the table. (Note that actual braking distance, which takes into account the driver's reaction time, is much longer, typically nearly 300 feet at 60 mph!)

Write a 95% confidence interval for the mean increase in stopping distance on wet pavement from dry pavement. Be sure to check the appropriate assumptions and conditions, and explain what your interval means.

Stopping Distance (ft.)											
Dry Pavement	145	152	141	143	131	148	126	140	135	133	
Wet Pavement	211	191	220	207	198	208	206	177	183	223	

- 17. Braking, test 2.** For another test of the tires in Exercise 16, the company tried them on 10 different cars, recording the stopping distance for each car on both wet and dry pavement. Results are shown in the table.

STOPPING DISTANCE (FT)		
Car #	Dry pavement	Wet pavement
1	150	201
2	147	220
3	136	192
4	134	146
5	130	182
6	134	173
7	134	202
8	128	180
9	136	192
10	158	206

Write a 95% confidence interval for the mean increase in stopping distance on wet pavement from dry pavement. Be sure to check the appropriate assumptions and conditions, and explain what your interval means.

- 18. Tuition.** How much more do public colleges and universities charge out-of-state students for tuition per semester? A random sample of 19 public colleges and universities listed in the Information Please Almanac found the data shown below. Tuition figures per semester are rounded to the nearest hundred dollars.

Institution	Resident	Nonresident
UAkron(OH)	3200	7900
Athens State (GA)	3800	7500
Ball State (IN)	3000	7800
Bloomsburg U (PA)	3700	8500
UC Irvine (CA)	4300	12,000
Central State (OH)	2900	6400
Clarion U (PA)	3900	8600
Dakota State	2500	4600
Fairmont State (WV)	1800	4500

Johnson State (VT)	4000	8600
Lock Haven U (PA)	7400	12,100
New College of S.Fla.	2000	7900
Oakland U (MI)	3100	8700
U Pittsburgh	5400	11,200
Savannah State (GA)	2000	5200
SW Louisiana	1900	4900
W Liberty State (WV)	1900	4500
W Texas State	1600	6000
Worcester State (MA)	2700	6200

- a) Create a 90% confidence interval for the mean difference in cost. Be sure to justify the procedure you use.
- b) Interpret your interval in context.
- c) A national magazine claims that public institutions charge state residents an average of \$4000 less for tuition each semester. What does your confidence interval indicate about the validity of this assertion?

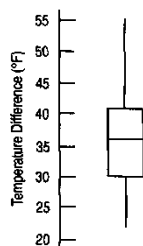
- 21. Sex and violence.** In June 2002, the Journal of Applied Psychology reported on a study that examined whether the content of TV shows influenced the ability of viewers to recall brand names of items featured in the commercials. The researchers randomly assigned volunteers to nine commercials. One of the programs had violent content, another sexual content, and the third neutral content. After the shows ended the subjects were asked to recall the brands of products that were advertised. Results are summarized below.

	Program Type		
	Violent	Sexual	Neutral
No. of subjects	101	106	103
Brands recalled			
Mean	3.02	2.72	4.65
St. Dev	1.61	1.85	1.62

Find a 95% confidence interval for the difference in mean number of brand names remembered between the groups watching shows with sexual content and those watching neutral shows. Interpret your interval in this context.

Answers:

1. a) 2.927
b) Larger
c) Based on this sample, we are 95% confident that students who learn math using the CPMP method will score on average between 5.57 and 11.43 points better on a test solving applied algebra problems with a calculator than students who learn by traditional methods.
d) Yes; 0 is not in the interval.
2. a) Based on this sample, we are 90% confident that people who receive either no or only verbal information about the image in a stereogram will take between 0.55 and 5.47 seconds longer, on average, to report they saw the image than people who receive both verbal and visual information.
b) Yes, since 0 is not in the interval.
c) 2.46
d) 90% of all random samples will produce intervals that contain the true value of the mean difference between the times for these two groups.
e) Wider. More confidence means less precision.
f) Possibly. The wider interval may contain 0.

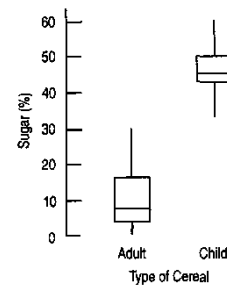


Data are paired for each city; cities are independent of each other; less than 10% of all European cities; boxplot shows the temperature differences are symmetric with no outliers. This is probably not a random sample, so we might be wary of inferring that this difference applies to all European cities. Based on these data, we are 90% confident that the average temperature in European cities in July is between 32.3° and 41.3°F higher than in January.

4. a) Same cows before and after injection; the cows should be representative of others of their breed; cows are independent of each other; less than 10% of all cows; don't know about Nearly Normal differences.
b) (12.66, 15.34)
c) Based on this sample, with 95% confidence, the average increase in milk production for Ayrshire cows given BST is between 12.66 and 15.34 pounds per day.
d) $0.25(47) = 11.75$. The average increase is much more than this, so we would recommend he go to the extra expense.
5. a) (1.313, 4.687)
b) No; 5 minutes is beyond the high end of the interval.
6. a) The mean rates are roughly equal, but females are more variable.

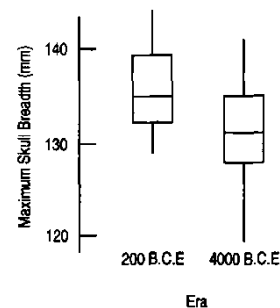
- b) Yes; boxplots look symmetric.
- c) (-3.025, 3.275)
- d) Yes; 0 is in the interval.

7.



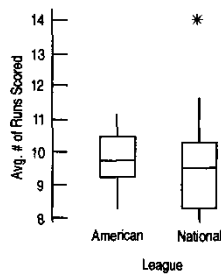
Random sample—questionable, but probably representative, independent samples, less than 10% of all cereals; boxplot shows *no* outliers—not exactly symmetric, but these are reasonable sample sizes. Based on these samples, with 95% confidence, children's cereals average between 32.49% and 40.80% more sugar content than adult's cereals.

8. a) Random sample (we assume), less than 10% of all Egyptians at those times, independent samples, boxplots look symmetric.



- b) (1.88, 6.66) mm
- c) These data provide evidence that mean maximum skull breadth in Egyptians in 200 B.C.E. was between 1.88 and 6.66 mm larger than that in 4000 B.C.E.
9. For a 95% C.I. (.2042, 19.64). These data do suggest that new activities are better since 0 is not in the interval. The mean reading comprehension score for the group with new activities is significantly higher than the mean score for the control group.
10. We are 90% confident that the average women's winning marathon time is between 16.86 and 18.06 minutes higher than the men's based on this sample.
11. Based on these data we are 90% confident that boys, on average, can do between 1.6 and 13.0 more push-ups than girls, (independent samples—not paired).
12. a) (4.8, 15.2) minutes
b) (4.9, 20.4) minutes
c) Margin of error for raw data will be higher.

13. a.



Both are reasonably symmetric, except for the outlier in the National League.

b) Based on these data, the average number of runs in an American League stadium is between 9.34 and 10.27, with 95% confidence.

c) There is an outlier. It could be deleted, and mentioned separately.

d) Yes. The boxplot indicates it is an outlier.

e) We want to work directly with the average difference. The two separate confidence intervals do not answer questions about the difference. The difference has a different standard deviation.

14. a) Males: (18.67,20.11) pgs; females: (16.95,18.87) pgs.

b) It suggests that there is no evidence of a difference, but the method is incorrect.

c) (0.29,2.67) pgs

d) There is a difference—males do better on average by between 0.29 and 2.67 pgs.

e) Two-sample inference

f) If we examine the gap between the means using two separate confidence intervals, we are essentially adding the margins of error, which are based on standard deviations. To be correct we must instead add variances. That's exactly what happens when we create the confidence interval for the difference in means—the proper approach.

15. a) (-0.69,1.14)

b) Based on these data, with 95% confidence, American League stadiums average between -0.69 and 1.14 more runs than National League stadiums.

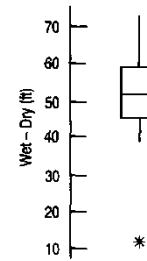
c) No; 0 is in the interval.

d) No. The interval is now (-0.19,1.23). It still contains 0.

16. Not a simple random sample, but most likely representative; stops most likely independent of each other; less than 10% of all possible wet stops; a normal probability plot is relatively straight.

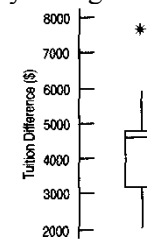
Based on these data, with 95% confidence, the average increase in distance for these tires on wet pavement is between 50.76 and 75.24 feet.

17. Data are paired by car; cars were probably not randomly chosen, but representative; less than 10% of all cars who might use these tires; boxplot shows an outlier (car 4) with a difference of 12. With deletion of that car, a Normal probability plot of the differences is relatively straight.



Having deleted an outlier difference of 12 feet, we estimate with 95% confidence that the average braking distance is between 38.784 and 62.616 feet more on wet pavement than on dry, based on this sample. (With the outlier, the confidence interval is 38.8 to 62.6 feet.)

18 a) Data are paired by college.



UC Irvine is an outlier. Without UC Irvine the 90% confidence interval for the mean difference in nonresident and resident tuition is (\$3662.65 to \$4592.91). With UC Irvine included, it is \$3769.30 to \$4862.30.

b) Having deleted the observation for UC Irvine, whose difference of \$7700 was an outlier, we are 90% confident based on the remaining data, that nonresidents pay, on average, between \$3662.60 and \$4592.90 more than residents. With UC Irvine, the interval is \$3769.30 to \$4862.30.

c) Assertion is reasonable; \$4000 is in the confidence interval.

21. With 95% confidence, the average number of brand names remembered 24 hours later is between 1.46 and 2.40 higher for viewers of neutral content shows than for viewers of sexual content shows, based on these data.