Two Sample Confidence Interval for a Difference in Population Means

[Navidi sections 5.4-5.7; Devore chapter 9]

From the **central limit theorem**, we know that, for sufficiently large sample sizes from two independent populations of means $\mu_1,\ \mu_2$ and variances $\sigma_1^{\,2},\ \sigma_2^{\,2}$, the sample means are distributed as

$$\bar{X}_1 \sim \mathrm{N}\left(\mu_1, \frac{\sigma_1^{\,2}}{n_1}\right), \qquad \bar{X}_2 \sim \mathrm{N}\left(\mu_2, \frac{\sigma_2^{\,2}}{n_2}\right), \text{ with } \quad \bar{X}_1 - \bar{X}_2 \sim$$

Also, for sufficiently large sample sizes, we may estimate unknown values of σ_1, σ_2 by s_1, s_2.

Example 12.01

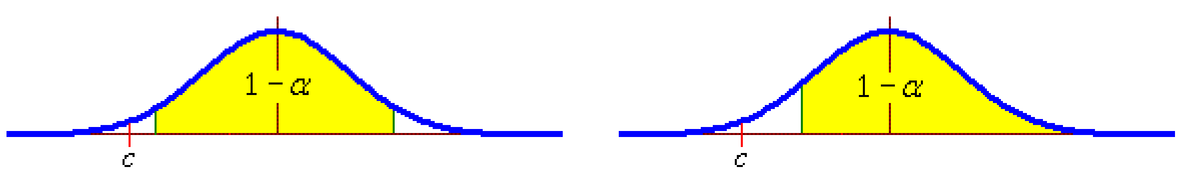
A large corporation wishes to determine the effectiveness of a new training technique. A random sample of 64 employees is tested after undergoing the new training technique and obtains a mean test score of 62.1 with a standard deviation of 5.12 . Another random sample of 100 employees, serving as a control group, is tested after undergoing the old training methods. The control group has a sample mean test score of 58.3 with a standard deviation of 6.30 .

(a) Use a two-sided confidence interval to determine whether the new training technique has led to a significant *change* in test scores.

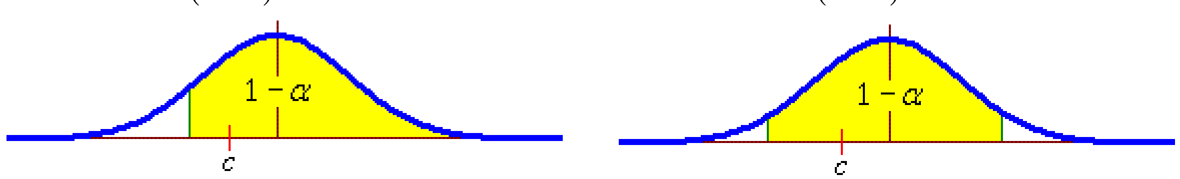
(b) Use a one-sided confidence interval to determine whether the new training technique has led to a significant *increase* in test scores.

Example 12.01 (continued)

If a two-sided $(1-\alpha)\times 100\%$ confidence interval does not include a value c, then the appropriate one-sided $(1-\alpha)\times 100\%$ confidence interval will not include c.



If a one-sided $(1-\alpha)\times 100\%$ CI includes c, then the two-sided $(1-\alpha)\times 100\%$ CI includes c.



Two Sample Confidence Interval for a Difference in Population Proportions
[for bonus questions only]

We know that when a random sample consists of n independent Bernoulli random quantities, with a probability p of success in each trial, the number X of successes in the random sample follows a binomial distribution with $\mu = np$ and $\sigma^2 = npq$.
Provided the expected numbers of successes (*np*) and failures (*nq*) are both more than 10, the sample proportion P of successes in the random sample follows a normal distribution to a good approximation:

$$P \sim \mathrm{N}\left(p, \frac{pq}{n}\right)$$

If we draw random samples of sizes n_X and n_Y (both large) from two independent populations whose true probabilities of success are p_X and p_Y, then the difference in sample proportions is

$$P_X - P_Y \sim \mathrm{N}\left(p_X - p_Y, \frac{p_X q_X}{n_X} + \frac{p_Y q_Y}{n_Y}\right)$$

This allows us to construct confidence intervals for the unknown $\left(p_X - p_Y\right)$. The traditional approach until the 1990's was to replace the unknown values of p_X and p_Y in the expression for the variance by the sample proportions, leading to the $(1-\alpha)\times 100\%$ confidence interval

$$\hat{p}_X - \hat{p}_Y \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_X \hat{q}_X}{n_X} + \frac{\hat{p}_Y \hat{q}_Y}{n_Y}}$$

However, simulations show that this confidence interval can be very inaccurate for sample sizes of about 30 or less.

A more accurate interval (the Agresti-Caffo interval) can be achieved by adding one to the observed numbers of successes, $x^* = x+1$ and $y^* = y+1$ and adding two to the sample sizes, $n_X^* = n_X + 2$ and $n_Y^* = n_Y + 2$.

It then follows that $p_X^* = \dfrac{x^*}{n_X^*}$, $q_X^* = 1 - p_X^*$, $p_Y^* = \dfrac{y^*}{n_Y^*}$, $q_Y^* = 1 - p_Y^*$

The two-sided $(1-\alpha)\times 100\%$ confidence interval for $\left(p_X - p_Y\right)$ is therefore

$$\boxed{p_X^* - p_Y^* \pm z_{\alpha/2}\sqrt{\frac{p_X^* q_X^*}{n_X^*} + \frac{p_Y^* q_Y^*}{n_Y^*}}}$$

If the lower bound is below –1, then replace it by –1.
If the upper bound is above +1, then replace it by +1.
This interval works well even for small sample sizes.

Example 12.02 [not examinable except for bonus]

A random sample of 25 components (produced by one machine) yields 15 components that are longer than 10.0 cm. Another random sample of 30 components (produced by another machine) yields 12 components that are longer than 10.0 cm. Construct a 95% confidence interval for the difference in population proportions of components that are longer than 10.0 cm. Can one conclude that the two population proportions are different?

Using the Agresti-Caffo 95% confidence interval for $p_1 - p_2$,

$$x_1 = 15, \quad n_1 = 25 \quad \Rightarrow \quad p_1{*} = \frac{x_1 + 1}{n_1 + 2} = \frac{16}{27} \quad \Rightarrow \quad q_1{*} = \frac{11}{27}$$

$$x_2 = 12, \quad n_2 = 30 \quad \Rightarrow \quad p_2{*} = \frac{13}{32} \quad \Rightarrow \quad q_2{*} = \frac{19}{32}$$

$$\Rightarrow \quad (p_1{*} - p_2{*}) = \frac{16}{27} - \frac{13}{32} = \frac{161}{864} = .186342\ldots$$

leading to the standard error

$$s{*} \;=\; \sqrt{\frac{p_1{*}\,q_1{*}}{n_1{*}} + \frac{p_2{*}\,q_2{*}}{n_2{*}}} \;=\; \sqrt{\frac{16\times 11}{27^3} + \frac{13\times 19}{32^3}} \;=\; .128372\ldots$$

$$z_{.025} = 1.959\ldots$$

Therefore the 95% CI is

$$(p_1{*} - p_2{*}) \;\pm\; z_{.025}\times s \;=\; .186\ldots \;\pm\; 1.959\ldots\times .128\ldots \;=\; .186\ldots \;\pm\; .251\ldots$$

Correct to 3 s.f., the 95% CI is

$$-.065 \le p_1 - p_2 < +.438$$

This confidence interval includes zero. Therefore $p_1 = p_2$ is plausible from these data.

<u>**NO**</u>, we cannot conclude that the two population proportions are different,
(even though the sample proportions differ by 20%!)
We need much larger sample sizes in order for a 20% difference in sample proportions to be significant.
[Note: the traditional 95% CI is .200 ± .260 = [–.060, +.460]]

Example 12.03 [not examinable except for bonus]

Suppose that random samples of equal size n are drawn from two independent populations. Find the smallest value of n for which an observed difference in sample proportions of 10% would allow us to conclude, at a level of confidence of 95%, that the population proportions are different.

The expression for the standard error in $\hat{P}_1 - \hat{P}_2$ is $\sigma = \sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}} = \dfrac{\sqrt{p_1 q_1 + p_2 q_2}}{\sqrt{n}}$

where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$. The maximum possible value of σ occurs when both $p_1(1-p_1)$ and $p_2(1-p_2)$ are as large as they can be.

But $y = x(1-x) = x - x^2$ is a quadratic function whose unique maximum occurs when $\dfrac{dy}{dx} = 1 - 2x = 0$, that is, at $x = \dfrac{1}{2}$.

$\Rightarrow$ the maximum value of σ occurs when $p_1 = q_1 = p_2 = q_2 = \dfrac{1}{2}$:

$$\Rightarrow \quad \sigma \le \frac{1}{\sqrt{n}} \sqrt{\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2}} = \frac{1}{\sqrt{2n}}$$

With $\hat{p}_1 - \hat{p}_2 = .10$, the lower bound of the traditional 95% confidence interval for $p_1 - p_2$ is

$$.10 - 1.95\ldots\sigma \ge .10 - 1.95\ldots\frac{1}{\sqrt{2n}}$$

For the CI to be guaranteed to exclude zero,

$$.10 - 1.95\ldots\frac{1}{\sqrt{2n}} > 0 \quad \Rightarrow \quad .10 > 1.95\ldots\frac{1}{\sqrt{2n}} \quad \Rightarrow \quad \sqrt{2n} > \frac{1.95\ldots}{.10} = 19.5\ldots$$

$$\Rightarrow \quad n > \frac{(19.5\ldots)^2}{2} = 192.07\ldots$$

Therefore a minimum sample size of **<u>193</u>** items from each population will guarantee that an observed difference of sample proportions of 10% will generate a 95% confidence interval for $p_1 - p_2$ that excludes zero.

With such large sample sizes, the Agresti-Caffo CI will be very close to the traditional CI.

Small Sample Confidence Interval for the Difference between Two Means

Suppose $\left\{X_1, X_2, \ldots, X_{n_X}\right\}$ is a random sample of size n_X
drawn from a population $X \sim \mathrm{N}\left(\mu_X, \sigma_X^2\right)$
and that $\left\{Y_1, Y_2, \ldots, Y_{n_Y}\right\}$ is a random sample of size n_Y
drawn from a population $Y \sim \mathrm{N}\left(\mu_Y, \sigma_Y^2\right)$, with Y independent of X, then

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_Y\right)}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}} \sim \mathrm{N}(0,1)$$

If at least one of the sample sizes is small (less than 30 or so), then the central limit theorem cannot be invoked. Z will be normal (exactly or to an acceptable approximation) only if both populations are exactly or nearly normal.

If, in addition, both population variances are known, then Z follows the standard normal distribution (exactly or approximately) no matter how small the sample sizes may be.

However, if the population variances are not known, then they must be replaced by their point estimators from the data (the sample variances):

$$\boxed{T = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_Y\right)}{\sqrt{\dfrac{S_X^2}{n_X} + \dfrac{S_Y^2}{n_Y}}}}$$

The additional uncertainty, (introduced by the fact that S_X^2 and S_Y^2 are themselves random quantities), results in T following Student's t-distribution instead of the standard normal distribution. For any particular pair of random samples, the number of degrees of freedom is

$$\boxed{\nu = \mathrm{INT}\left(\frac{\left(\dfrac{s_X^2}{n_X} + \dfrac{s_Y^2}{n_Y}\right)^2}{\dfrac{1}{n_X - 1}\left(\dfrac{s_X^2}{n_X}\right)^2 + \dfrac{1}{n_Y - 1}\left(\dfrac{s_Y^2}{n_Y}\right)^2}\right)}$$

Example 12.04

An investigator wants to know which of two electric toasters has the greater ability to resist the abnormally high electrical currents that occur during an unprotected power surge. Random samples of six toasters from factory A and five toasters from factory B were subjected to a destructive test, in which each toaster was subjected to increasing currents until it failed. The distribution of currents at failure (measured in amperes) is known to be approximately normal for both products. The results are as follows:

Factory A:	20	28	24	26	23	26
Factory B:	21	18	19	17	22	

(a) State the assumptions that you are making.
(b) Construct a 95% confidence interval for the mean difference in failure currents.
(c) Is there any significant difference between the failure currents of the two types of toaster?

(a) Given in the question:

Assumption:

(b) The **summary statistics** are

$$n_A = 6 \qquad \bar{x}_A = 24.5 \qquad s_A^2 = 7.9 \qquad \Rightarrow \quad \frac{s_A^2}{n_A} = 1.31\dot{6}$$

$$n_B = 5 \qquad \bar{x}_B = 19.4 \qquad s_B^2 = 4.3 \qquad \Rightarrow \quad \frac{s_B^2}{n_B} = 0.86$$

$$\Rightarrow \quad (s.e.)^2 = \frac{s_A^2}{n_A} + \frac{s_B^2}{n_B} = 2.17\dot{6} \qquad \Rightarrow \quad s.e. = 1.475353\ldots$$

The most tedious part of this set of calculations is for the number of degrees of freedom:

Example 12.04 (continued)

$$\nu = \text{INT}\left(\frac{\left(\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B} \right)^2}{\dfrac{1}{n_A - 1}\left(\dfrac{s_A^2}{n_A} \right)^2 + \dfrac{1}{n_B - 1}\left(\dfrac{s_B^2}{n_B} \right)^2} \right)$$

$$= \text{INT}\left(\frac{\left(2.17\dot{6}\right)^2}{\dfrac{\left(1.31\dot{6}\right)^2}{5} + \dfrac{\left(0.86\right)^2}{4}} \right) = \text{INT}\left(8.912\ldots\right) = 8$$

For a 95% two-sided confidence interval with 8 degrees of freedom, we need

$$t_{.025,\,8} = 2.306\,00$$

$$\left(\bar{x}_A - \bar{x}_B\right) \pm t_{.025,\,8} \times \left(s.e.\right) = \left(24.5 - 19.4\right) \pm 2.306\,00 \times 1.475\,353\ldots$$
$$= 5.1 \pm 3.402\ldots$$

Therefore the 95% CI for $\left(\mu_A - \mu_B\right)$ is, correct to 1 d.p.,

$$\left[\, 1.7, 8.5 \,\right] \quad \text{A}$$

An Excel spreadsheet for this example is available at
" `www.engr.mun.ca/~ggeorge/4421/demos/CI2small_n.xls` "

The textbooks (Navidi section 5.6; Devore section 9.2) provide an alternative more precise confidence interval that may be used only when the population variances can safely be assumed to be equal (“pooled *t* procedures”). We shall not employ that interval in this course.

Confidence Intervals for Paired Data

Example 12.05

Nine volunteers are tested before and after a training programme. Based on the data below, can you conclude that the programme has improved test scores?

Volunteer:	1	2	3	4	5	6	7	8	9
After training:	75	66	69	45	54	85	58	91	62
Before training:	72	65	64	39	51	85	52	92	58

Let X_A = score **a**fter training and X_B = score **b**efore training.

INCORRECT METHOD:

The summary statistics are

$n_A = n_B = 9$,

$$\bar{x}_A = 67.\dot{2}, \qquad s_A^2 = \frac{1943.5}{9} = 215.9\dot{4} \qquad \Rightarrow \quad \frac{s_A^2}{n_A} = \frac{1943.5}{81} = 23.993\ldots$$

$$\bar{x}_B = 64.\dot{2}, \qquad s_B^2 = \frac{2546.5}{9} = 282.9\dot{4} \qquad \Rightarrow \quad \frac{s_B^2}{n_B} = \frac{2546.5}{81} = 31.438\ldots$$

$$\Rightarrow \quad (s.e.)^2 = \frac{s_A^2}{n_A} + \frac{s_B^2}{n_B} = \frac{4490}{81} = 55.432\ldots \qquad \Rightarrow \quad s.e. = 7.445273\ldots$$

$$\nu = \text{INT}\left(\frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B} \right)^2}{\frac{1}{n_A - 1}\left(\frac{s_A^2}{n_A} \right)^2 + \frac{1}{n_B - 1}\left(\frac{s_B^2}{n_B} \right)^2} \right) = \text{INT}\left(\frac{(55.432\ldots)^2}{\frac{1}{8}(23.993\ldots)^2 + \frac{1}{8}(31.438\ldots)^2} \right)$$

$$= \text{INT}(15.7\ldots) = 15$$

$t_{.050,15} = 1.75305$

We need a one-sided CI because we are looking for evidence of an increase, not a change.

Example 12.05 (continued)

The lower boundary of the 95% one-sided CI for $(\mu_A - \mu_B)$ is

$(\bar{x}_A - \bar{x}_B) - t_{.050,15} \times (s.e.) = (67.\dot{2} - 64.\dot{2}) - 1.75305 \times 7.445273\ldots = 3.0 - 13.051\ldots$

We are "95% sure that $\mu_A - \mu_B \geq -10.1$".

Clearly zero is a very plausible value.

We would conclude that the data are consistent with $\mu_A = \mu_B$ and therefore there is nowhere near enough evidence to conclude that the training has improved test scores.

But look again at the test scores:

Volunteer:	1	2	3	4	5	6	7	8	9
After training:	75	66	69	45	54	85	58	91	62
Before training:	72	65	64	39	51	85	52	92	58

There is a clear pattern of increases from the before score to the after score for most individuals. There is considerable variation between individuals, which is swamping the individual increases.

The error in the analysis above is that

CORRECT METHOD:

The remedy is to explore only the differences, not the two sets of scores.

Volunteer:	1	2	3	4	5	6	7	8	9
Difference d :									

Now we have a single sample, from which to construct a one-sided confidence interval for the true mean difference μ_D.

$$n = 9, \qquad \bar{d} = 3.0, \qquad s_D^2 = \frac{9 \times 133 - 27^2}{9 \times 8} = \frac{13}{2} = 6.5 \qquad \Rightarrow \qquad s_D = 2.549\ldots$$

$$s.e. = \frac{s_D}{\sqrt{n}} = 0.849836\ldots \qquad\qquad t_{.050,8} = 1.85955$$

$$\mu_D > \bar{d} - t_{.050,8}(s.e.) = 3.0 - 1.85\ldots \times 0.84\ldots = 3.0 - 1.58\ldots \approx 1.42$$

Paired vs. Unpaired Confidence Intervals

When the sample sizes are equal, the question arises as to which type of confidence interval to use: two-sample (unpaired) or the interval based on differences (paired).

If the samples are pairs of observations of two different effects on the ***same set of individuals***, then independence between the populations is unlikely and one should use the **paired** confidence interval. Otherwise use the unpaired confidence interval.

The paired confidence interval is valid even if the two populations are strongly correlated, whereas the unpaired confidence interval is based on the assumption that the two populations are independent (or at least uncorrelated).

<u>Example 12.06</u> (Navidi textbook, exercises 5.7, page 373, question 6)

A sample of 10 diesel trucks were run both hot and cold to estimate the difference in fuel economy. The results, in miles per gallon, are presented in the following table. (from "In-use Emissions from Heavy-Duty Diesel Vehicles", J. Yanowitz, Ph.D. thesis, Colorado School of Mines, 2001.)

Truck	**Hot**	**Cold**
1	4.56	4.26
2	4.46	4.08
3	6.49	5.83
4	5.37	4.96
5	6.25	5.87
6	5.90	5.32
7	4.12	3.92
8	3.85	3.69
9	4.15	3.74
10	4.69	4.19

Find a 98% confidence interval for the difference in mean fuel mileage between hot and cold engines.

Choice of type of confidence interval:

Example 12.06 (continued)

Truck	Hot	Cold	Difference
1	4.56	4.26	0.30
2	4.46	4.08	0.38
3	6.49	5.83	0.66
4	5.37	4.96	0.41
5	6.25	5.87	0.38
6	5.90	5.32	0.58
7	4.12	3.92	0.20
8	3.85	3.69	0.16
9	4.15	3.74	0.41
10	4.69	4.19	0.50

Summary statistics:

$n = 10, \quad \sum d = 3.98, \quad \sum d^2 = 1.8026 \qquad \Rightarrow \quad \bar{d} = \frac{\sum d}{n} = \frac{3.98}{10} = 0.398,$

$$s_D^2 = \frac{n\sum d^2 - \left(\sum d\right)^2}{n(n-1)} = \frac{10 \times 1.8026 - (3.98)^2}{10 \times 9} = \frac{2.1856}{90} = 0.02428\dot{4}$$

$$\Rightarrow \quad s_D = 0.155834\ldots \qquad \Rightarrow \quad s.e. = \frac{s_D}{\sqrt{n}} = \frac{0.15\ldots}{\sqrt{10}} = 0.049279\ldots$$

[End of Chapter 12]