You might find this additional information useful...

Supplemental material for this article can be found at:
http://ajpadvan.physiology.org/cgi/content/full/33/2/87/DC1

This article cites 21 articles, 10 of which you can access free at:
http://ajpadvan.physiology.org/cgi/content/full/33/2/87#BIBL

Updated information and services including high-resolution figures, can be found at:
http://ajpadvan.physiology.org/cgi/content/full/33/2/87

Additional material and information about *Advances in Physiology Education* can be found at:
http://www.the-aps.org/publications/advan

This information is current as of August 27, 2009 .

Explorations in statistics: confidence intervals

Douglas Curran-Everett

*Division of Biostatistics and Bioinformatics, National Jewish Health, and Department of Biostatistics and Informatics
and Department of Physiology and Biophysics, School of Medicine, University of Colorado Denver, Denver, Colorado*

Curran-Everett D. Explorations in statistics: confidence intervals.
Adv Physiol Educ 33: 87–90, 2009; doi:10.1152/advan.00006.2009.—
Learning about statistics is a lot like learning about science: the learning
is more meaningful if you can actively explore. This third installment of
Explorations in Statistics investigates confidence intervals. A confidence
interval is a range that we expect, with some level of confidence, to
include the true value of a population parameter such as the mean. A
confidence interval provides the same statistical information as the *P*
value from a hypothesis test, but it circumvents the drawbacks of that
hypothesis test. Even more important, a confidence interval focuses our
attention on the scientific importance of some experimental result.

estimation; R; software

THIS THIRD ARTICLE in *Explorations in Statistics* (see Refs. 11
and 12) provides an opportunity to explore confidence inter-
vals. A confidence interval estimates our uncertainty about the
true value of some population parameter.[1] For example, when
we construct a confidence interval for the mean of some
population, we expect, with some level of confidence, that the
true value of the population mean will fall within that interval.
A confidence interval provides the same statistical information
as the *P* value from a hypothesis test, it circumvents the
drawbacks inherent to that hypothesis test, and it provides
information about scientific importance. The routine reporting
of confidence intervals is recommended (1–3, 9, 13, 15, 18),
but the meaning of a confidence interval is difficult to under-
stand (7, 16). To be blunt, the meaning of a confidence interval
is nearly impossible to understand unless you observe the
development of its underlying concept. In this exploration, we
will.

A Brief History of Confidence Intervals

Unlike hypothesis tests whose origins can be traced to 1279
(25), confidence intervals are a recent development: Jerzy
Neyman derived them in the 1930s (20–22). There would be a
50-year lag before medical journals advocated the use of
confidence intervals (4, 5, 17, 18). It would be just 10 years
before George Snedecor added confidence intervals to his
historic *Statistical Methods* (24).[2] In 1913, 6 years before
Fisher went to Rothamsted Station (11, 26), Snedecor arrived
at Iowa State College as an assistant professor of mathematics
(6, 8, 10, 19). In his courses, Snedecor derived examples based
on agricultural and biological data from researchers at Iowa
State. These collaborations led Snedecor to create the Mathe-

matics Statistical Service (1927) and then the Statistical Lab-
oratory (1933) at what is now Iowa State University. The
hallmark of *Statistical Methods* is its focus on the application
of statistical methods to actual scientific problems and data.

R: Basic Operations

In the inaugural article (12) of this series, I summarized the
freeware package R (23) and outlined its installation. For this
exploration, there is just one additional step: download the
script Advances_Statistics_Code_CI.R[3] to your Advances
folder (see Ref. 12).

If you use a Mac, highlight the commands in Advances_Sta-
tistics_Code_CI.R you want to submit and then press
⌘↵ (command key+enter). If you use a PC, highlight the
commands you want to submit, right-click, and then click Run
line or selection. Or, highlight the commands you want to
submit and then press Ctrl+R.

The Simulation: Observations and Sample Statistics

For these explorations (11, 12), we drew a total of 1000
random samples–each with 9 observations–from our popula-
tion, a standard normal distribution with mean $\mu = 0$ and
standard deviation $\sigma = 1$ (see Ref. 12, Fig. 2). These were the
observations–the data–for *samples 1*, *2*, and *1000*:

```
> # Sample Observations

  [1]   0.422   1.103   1.006   1.034   0.285  -0.647   1.235   0.912   1.825
  [2]   0.154  -0.654  -0.147   1.715   0.720   0.804   0.256   1.155   0.646
   :
[1000]  0.560  -1.138   0.485  -0.864  -0.277   2.198   0.050   0.500   0.587
```

Each time we drew a random sample, we calculated the sample
statistics listed in Table 1. These were the statistics for *samples
1*, *2*, and *1000*:

```
> # Sample    Mean     SD      SE      t       LCI     UCI
     [.1]     [.2]    [.3]    [.4]    [.5]    [.6]    [.7]
       1     0.797   0.702   0.234   3.407   0.362   1.232
       2     0.517   0.707   0.236   2.193   0.079   0.955
       :
     1000    0.233   0.975   0.325   0.718  -0.371   0.838
```

The commands in *lines 35–63* of Advances_Statistics_Code_
CI.R generate the observations and compute the sample
statistics. These commands are identical to those in the first
two scripts (11, 12).

With these 1000 sets of sample statistics, we are ready to
explore confidence intervals.

Confidence Intervals

When we began these explorations, we wanted in part to
estimate $\mu = 0$, the mean of our population. In the first
iteration of our simulation, the sample mean $\bar{y} = 0.797$ esti-

---

Address for reprint requests and other correspondence: D. Curran-Everett,
Div. of Biostatistics and Bioinformatics, M222, National Jewish Health, 1400
Jackson St., Denver, CO 80206 (e-mail: EverettD@NJHealth.org).

[1] A parameter is a numerical constant: for example, the population mean.
[2] Snedecor published the early editions of *Statistical Methods* in 1937, 1938,
and 1940. William Cochran contributed a chapter to the 1956 edition and
helped author the 1967, 1980, and 1989 editions.

[3] This file is available through the Supplemental Material link for this article
at the *Advances in Physiology Education* website.

Table 1. *Sample statistics calculated for each random sample*

Column	Heading	Sample Statistic
1	Sample	Sample number
2	Mean	Mean $\bar{y}$
3	SD	Standard deviation s
4	SE	Standard error of the mean SE$\{\bar{y}\} = s/\sqrt{n}$
5	t	Observed value of $t = \bar{y}/\text{SE}\{\bar{y}\}$
6	LCI	Lower confidence interval bound
7	UCI	Upper confidence interval bound

[Reprinted from Ref. 12.]

mated the population mean. In the second iteration, the sample mean 0.517 estimated the population mean. All told, we have 1000 sample estimates of the population mean: 900 of them are between -0.523 and $+0.552$ (Fig. 1).[4] We can generalize from this empirical distribution of sample means to the theoretical distribution of the sample mean, a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ (12, 15), where n is the number of observations in the sample.

In the theoretical distribution of the sample mean (Fig. 2), $100(1 - \alpha)\%$ of the possible sample means are covered by the interval

$$[\mu - a, \mu + a] \quad,$$

where the allowance a is

$$a = z_{\alpha/2} \cdot \text{SD}\{\bar{y}\} \quad. \tag{1}$$

---

[4] The command in *line 75* of Advances_Statistics_Code_CI.R returns these values. Your values will differ slightly.
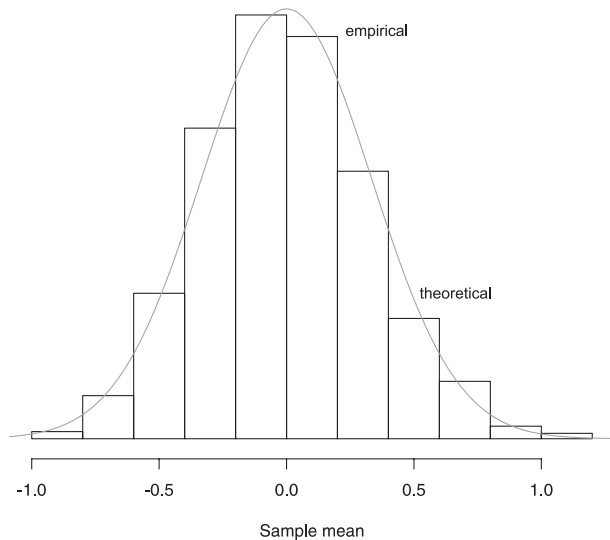

Fig. 1. Empirical (black) and theoretical (gray) distributions of the sample mean for 9 observations. The empirical distribution is composed of 1000 sample means. The empirical standard deviation of the 1000 sample means, 0.326, is near the theoretical value of 1/3. [Reprinted from Ref. 12.]
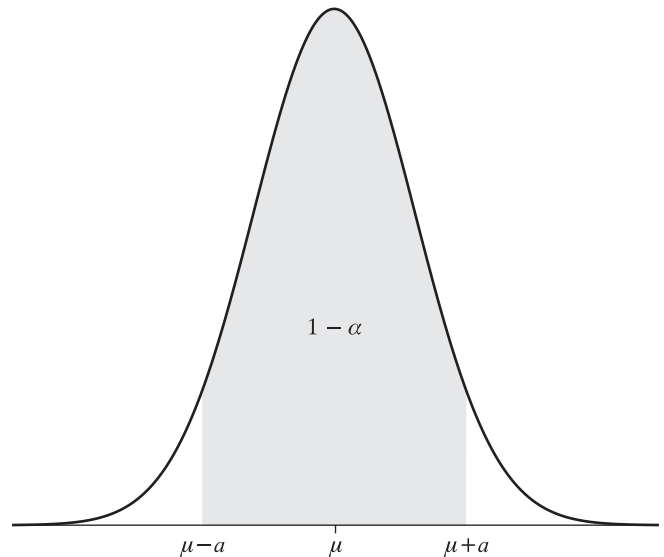

Fig. 2. Theoretical distribution of the sample mean for n observations. The interval $[\mu - a, \mu + a]$ covers $100(1 - \alpha)\%$ of the possible sample means (see *Eq. 1*). Compared to the distribution of population values, the theoretical distribution of the sample mean is narrower by a factor of $1/\sqrt{n}$ and taller by factor of $\sqrt{n}$ (see Fig. 1 in Ref. 13).

In *Eq. 1*, $z_{\alpha/2}$ is the $100[1 - (\alpha/2)]$th percentile from the standard normal distribution, and SD$\{\bar{y}\}$ is the standard deviation of the sample means, $\sigma/\sqrt{n}$. The standard deviation of the distribution of the sample mean is also called the standard error of the sample mean SE$\{\bar{y}\}$.

Suppose we want the interval $[\mu - a, \mu + a]$ to cover 90% of the possible sample means for 9 observations. In this situation, $\mu = 0$, $\alpha = 0.10$, and $z_{\alpha/2} = 1.645$. Because we defined the population standard deviation σ to be 1 (see Ref. 12),

$$\text{SD}\{\bar{y}\} = \sigma/\sqrt{n} = 1/\sqrt{9} = 1/3 \quad,$$

and the resulting allowance a is

$$a = z_{\alpha/2} \cdot \text{SD}\{\bar{y}\} = 1.645 \cdot 1/3 = 0.548 \quad.$$

Therefore, the interval $[-0.548, +0.548]$ covers 90% of the sample means for 9 observations. This theoretical interval agrees with the empirical interval of $[-0.523, +0.552]$.

When we calculate the interval $[\mu - a, \mu + a]$, we use the population mean μ to learn about possible values of the sample mean $\bar{y}$. Is this what we do when we calculate a confidence interval for the mean of some population? Sadly, no. When we calculate a confidence interval, we use the sample mean $\bar{y}$ to learn about possible values of the population mean μ. Happily, we can use the interval $[\mu - a, \mu + a]$ to derive a confidence interval.

First, we write the interval $[\mu - a, \mu + a]$ as the probability expression

$$\Pr\{\mu - a \le \bar{y} \le \mu + a\} = 1 - \alpha \quad.$$

What does this expression mean in words? It means that the probability is $1 - \alpha$ that a sample mean is covered by–lies within–the interval $[\mu - a, \mu + a]$ (see Fig. 2). Then, we rearrange the joint inequality portion of the expression to get

$$\Pr\{\bar{y} - a \le \mu \le \bar{y} + a\} = 1 - \alpha \quad .$$

In this form, the interval

$$[\bar{y} - a, \bar{y} + a] \qquad (2)$$

is called the $100(1 - \alpha)\%$ confidence interval for the population mean μ. Now we have what we want.

In an actual experiment, we do not know the population standard deviation σ. Therefore, we use the sample standard deviation s to estimate the population standard deviation σ and $s/\sqrt{n}$ to estimate the standard error of the sample mean. In addition, when we calculate a $100(1 - \alpha)\%$ confidence interval for some population mean μ, we handle our uncertainty about the actual value of σ by replacing $z_{\alpha/2}$ (*Eq. 1*) with $t_{\alpha/2,v}$, the $100[1 - (\alpha/2)]$th percentile from a Student t distribution with $v = n - 1$ degrees of freedom. As a result, the allowance we apply to the sample mean to obtain the $100(1 - \alpha)\%$ confidence interval (*Eq. 2*) becomes

$$a = t_{\alpha/2,v} \cdot \mathrm{SE}\{\bar{y}\} \quad ,$$

where $\mathrm{SE}\{\bar{y}\} = s/\sqrt{n}$.[5] This allowance is bigger than the allowance in *Eq. 1*: we are more uncertain about the value of the population mean μ. This happens because if $v < \infty$, then $t_{\alpha/2,v} > z_{\alpha./2}$ for all values of α.

Suppose we want to calculate a confidence interval for the population mean $\mu = 0$ using the observations 0.422, 1.103, ..., 1.825 of the first sample. The mean and standard deviation of these 9 observations are $\bar{y} = 0.797$ and $s = 0.702$, and the estimated standard error of the mean is

$$\mathrm{SE}\{\bar{y}\} = s/\sqrt{n} = 0.702/\sqrt{9} = 0.234 \quad .$$

Because $n = 9$, there are $v = 8$ degrees of freedom. If we want a 90% confidence interval, then $\alpha = 0.10$, $t_{\alpha/2,v} = 1.860$, and the allowance $a = 1.860 \times 0.234 = 0.435$. Therefore, the 90% confidence interval is

$$[0.362, 1.232] \doteq [0.36, 1.23] \quad .$$

What does this expression mean in words? We can declare, with 90% confidence, that the population mean is included in the interval [0.36, 1.23]. Because 0 is outside this interval, we can state, with 90% confidence, that 0 is not a plausible value of the population mean. This inference is consistent with our second exploration in which we rejected the null hypothesis H_0: $\mu = 0$ and concluded that the sample observations were consistent with having come from a population that had a mean other than 0 (see Ref. 12).

But now we have a problem: a single confidence interval either does or does not include the true value of some population parameter. In a real experiment, we do not know which outcome has occurred. So the question is, where does the notion of confidence in a confidence interval come from? The answer: not from a single confidence interval but from a theoretical process of calculating a whole bunch of confidence intervals. For these explorations, we drew a total of 1000 random samples. Each time we drew a random sample, we calculated its mean and standard deviation. Because the population from which we drew the samples was distributed over a range of possible values, the sample means (see Fig. 1) and standard deviations (see Ref. 12, Fig. 3) varied among our 1000 samples. Therefore, we calculated 1000 different confidence intervals. We expect about $100(1 - \alpha)\%$ of these confidence intervals to include the actual value of the population mean (Fig. 3). This is the underlying concept of confidence in a confidence interval. The next question is, how do we use a confidence interval to help us make an inference about scientific importance?

In a manner similar to Ref. 14, suppose you find three articles in *Physiological Genomics* that investigated independently the impact of three different drugs on the expression of some gene. Suppose also that a fractional change of 0.25 (25%) results in an altered phenotype. Each study involved a sample of 9 subjects, and each reported a 90% confidence interval for the fractional change in expression of the gene. For each drug, these are the sample mean $\bar{y}$, sample standard deviation s, P value, and 90% confidence interval:

Drug	$\bar{y}$	s	P	Conf Int
A	0.797	0.702	0.005	0.36 to 1.23
B	0.008	0.007	0.005	0.004 to 0.01
C	0.797	2.106	0.14	−0.51 to + 2.10

How do you interpret these results?

Drug A increased expression by 80%, a change that differed from 0 ($P = 0.005$). The confidence interval suggests the true impact of *drug A* is probably a 36–123% increase in expression, a change that is scientifically meaningful. *Drug A* produced a convincing change of scientific importance.
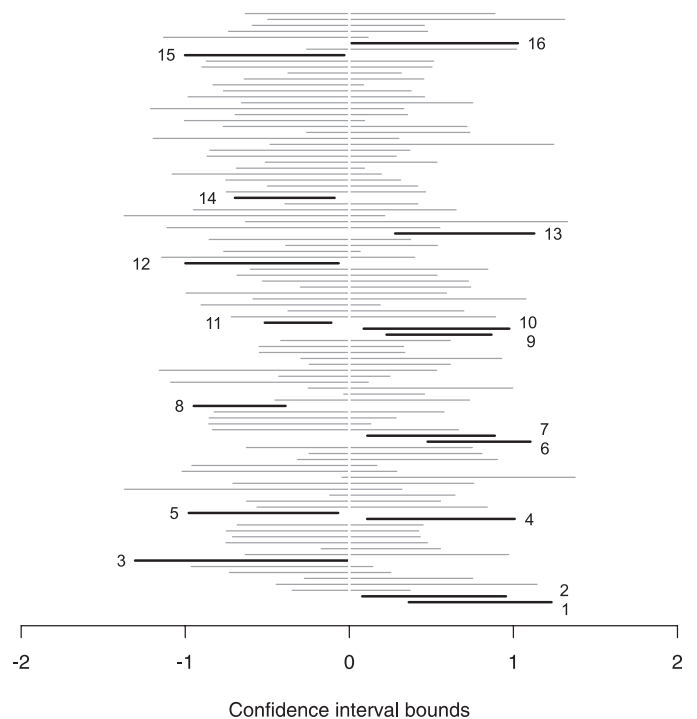

Fig. 3. Confidence intervals for the initial 100 samples of 9 observations. It is because of random sampling that the position and length of the confidence intervals vary from sample to sample. About 90 of these intervals–the actual number will vary–are expected to cover the population mean of 0. In this simulation, 84 of the confidence intervals cover 0; the 16 exceptions are highlighted (numbered black lines). To generate this data graphic, highlight and submit the lines of code from Figure 3: first line to Figure 3: last line.

---

[5] The standard error of the sample mean $\mathrm{SE}\{\bar{y}\}$ is identical to the standard deviation of the theoretical distribution of the sample mean $\mathrm{SD}\{\bar{y}\}$ in *Eq. 1*.

Drug B increased expression by 1%, a change that also differed from 0 (P = 0.005). The confidence interval suggests the true impact of *drug B* is probably a 0.4–1% increase in expression, a change that is scientifically trivial but quite precise. *Drug B* produced a convincing change of no scientific importance.

Drug C increased expression by 80%, a change consistent with 0 (P = 0.14). The confidence interval suggests the true impact of *drug C* could range from a 51% decrease to a 210% increase in expression. Either would be scientifically meaningful. Because it is relatively long, the confidence interval for *drug C* is an imprecise estimate of the true impact of *drug C* on expression of the gene. *Drug C* bears further study using a larger sample size.

Note that the scientific importance of the upper and lower bounds of a confidence interval depends on scientific context.

Summary

As this exploration has demonstrated, a confidence interval is a range that we expect, with some level of confidence, to include the true value of a population parameter such as the mean. For example, when we construct a confidence interval for the mean of some population, we assign numerical limits to the expected discrepancy between the sample mean $\bar{y}$ and the population mean μ. A confidence interval is useful because it focuses our attention away from a singularly statistical *P* value and toward the scientific importance of some experimental result.

In the next installment of this series, we will explore bootstrapping, a statistical technique even more recent than confidence intervals. Bootstrapping gives us an approach we can use to assess whether the inferences we make from hypothesis tests and confidence intervals are justified.

REFERENCES

1. **Altman DG.** Statistics in medical journals: developments in the 1980s. *Stat Med* 10: 1897–1913, 1991.
2. **Altman DG.** Statistics in medical journals: some recent trends. *Stat Med* 19: 3275–3289, 2000.
3. **Altman DG, Machin D, Bryant TN, Gardner MJ.** *Statistics with Confidence*. Bristol: BMJ Books, 2000.
4. **Altman DG, Gore SM, Gardner MJ, Pocock SJ.** Statistical guidelines for contributors to medical journals. *Br Med J* 286: 1489–1493, 1983.
5. **Bailar JC III, Mosteller F.** Guidelines for statistical reporting in articles for medical journals. *Ann Intern Med* 108: 266–273, 1988.
6. **Bancroft TA.** George W. Snedecor, pioneer statistician, 1881–1974. *Am Stat* 28: 108–109, 1974.
7. **Belia S, Fidler F, Williams J, Cumming G.** Researchers misunderstand confidence intervals and standard error bars. *Psychol Methods* 10: 389–396, 2005.
8. **Cochran WG.** George W Snedecor. *J R Stat Soc Ser A* 137: 456–457, 1974.
9. **Council of Science Editors, Style Manual Subcommittee.** *Scientific Style and Format: the CSE Manual for Authors, Editors, and Publishers* (7th ed.). Reston, VA: Rockefeller Univ. Press, 2006.
10. **Cox GM, Homeyer PG.** Professional and personal glimpses of George W. Snedecor. *Biometrics* 31: 265–301, 1975.
11. **Curran-Everett D.** Explorations in statistics: hypothesis tests and *P* values. *Adv Physiol Educ*; doi:10.1152/advan.90218.2008.
12. **Curran-Everett D.** Explorations in statistics: standard deviations and standard errors. *Adv Physiol Educ* 32: 203–208, 2008.
13. **Curran-Everett D, Benos DJ.** Guidelines for reporting statistics in journals published by the American Physiological Society. *Am J Physiol Endocrinol Metab* 287: E189–E191, 2004.
14. **Curran-Everett D, Benos DJ.** Last Word on Perspectives "Guidelines for reporting statistics in journals published by the American Physiological Society: the sequel". *Adv Physiol Educ* 31: 306–307, 2007.
15. **Curran-Everett D, Taylor S, Kafadar K.** Fundamental concepts in statistics: elucidation and illustration. *J Appl Physiol* 85: 775–786, 1998.
16. **Fidler F, Thomason N, Cumming G, Finch S, Leeman J.** Editors can lead researchers to confidence intervals, but can't make them think: statistical reform lessons from medicine. *Psychol Sci* 15: 119–126, 2004.
17. **Gardner MJ, Altman DG.** Confidence intervals rather than *P* values: estimation rather than hypothesis testing. *Br Med J* 292: 746–750, 1986.
18. **International Committee of Medical Journal Editors.** Uniform requirements for manuscripts submitted to biomedical journals. *Ann Intern Med* 108: 258–265, 1988.
19. **Kempthorne O.** George W. Snedecor, 1881–1974. *Int Stat Rev* 42: 319–323, 1974.
20. **Neyman J.** On the two different aspects of the representative method. *J R Stat Soc Ser A* 97: 558–625, 1934.
21. **Neyman J.** On the problem of confidence intervals. *Ann Math Stat* 6: 111–116, 1935.
22. **Neyman J.** Outline of a theory of statistical estimation based on the classical theory of probability. *Philos Trans R Soc Lond A* 236: 333–380, 1937.
23. **R Development Core Team.** *R: a Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2008; http://www.R-project.org.
24. **Snedecor GW.** *Statistical Methods Applied to Experiments in Agriculture and Biology* (4th ed.). Ames, IA: Iowa State College Press, 1946.
25. **Stigler SM.** *Statistics on the Table: the History of Statistical Concepts and Methods*. Cambridge, MA: Harvard Univ. Press, 1999.
26. **Yates F, Mather K.** Ronald Aylmer Fisher, 1890–1962. *Biogr Mem Fellows R Soc Lond* 9: 91–120, 1963.