Exploiting the Confidence Interval-Hypothesis Test Equivalence in Basic Statistics Classes

Ken Constantine
Current affiliation: Eastern Nazarene College
Fall affiliation: Taylor University

Abstract:
An emphasis is offered for the inference portion of an elementary Statistics course: the equivalence between confidence intervals and tests of hypotheses. This equivalence is rarely mentioned in basic texts but seems helpful to students. Student reference sheets which employ this equivalence are available on-line.

Outline:
We begin with an example of the inconsistency which can befall the beginning student who conducts a 1-sided test of hypothesis and constructs a 2-sided confidence interval. We then develop the (standard) 1-sided confidence interval and discuss the general equivalence between confidence intervals and tests of hypothesis. After demonstrating how a 1-sided interval rescues consistency in our example, we discuss a few pedagogical issues relating to classroom implementation. We finish with some reference sheets which offer students a format which emphasizes the interval-test equivalence.

EXAMPLE:
The population for our example is from the *Chronicle of Philanthropy*, 1 May 2003, page 12. There are 100 U. S. metropolitan areas and for each is given: the number of itemized tax returns filed, the average discretionary income for those returns, and the average charitable donation amount for those returns. See the URL

http://alpha2.enc.edu/~constant/library/statistics/data/charity/charity.xls

for an Excel spreadsheet providing this information for the full population.

From this population a sample of 10 cities was selected by simple random sampling. Those cities and their data are as follows:

city	**income**	**donation**
Charlotte	47,262	3,747
Cincinnati	44,229	3,163
Cleveland	46,425	3,141
Denver	60,326	6,094
Jacksonville	59,444	4,356
LA	74,960	5,169
Memphis	71,335	6,464
Milwaukee	44,396	3,749
Research Triangle	48,783	3,383
San Diego	39,086	2,680

standard methods

We begin with standard methods for testing the hypotheses

H₀: μ = \$5,000 vs.

Hₐ: μ < \$5,000,

where μ denotes the mean donation per tax return. (It must be confessed in passing that the value of \$5000 was chosen because it serves the purpose of this paper, not necessarily because of its intrinsic interest.)

The classic test statistic for this problem is the one-sample T-Statistic

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{\bar{x} - 5000}{s/\sqrt{10}}, \text{ where}$$

$\bar{x}$ and s denote the mean and standard deviation of the sample, respectively.

The value of the test statistic for our data is

$$T = \frac{4195 - 5000}{412} = -1.96.$$

The corresponding p-value is computed as the area to the left of t= −1.96 in a T-distribution with 9 degrees of freedom; it equals .041. At a 5% level of significance, our conclusion is to reject the null hypothesis and infer that *the mean donation per itemized tax return is less than \$5000.*

While checking assumptions is not the issue at hand, we remark in passing that we would want students to be aware of the assumptions upon which their methods are built (i.e. normality of the population and random sampling), know how to check those assumptions (e.g. via a histogram of the data), and to appreciate the sensitivity of the methods to those assumptions (not sensitive to normality here).

At this point we would want our students to ask if "statistical significance" actually implies "practical consequence" in the specific application. We would hope that they address this question by constructing a confidence interval for the parameter of interest: μ in this example.

The standard (2-sided) interval for μ which has level 95% is

$$\left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \ , \ \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right] = \left[\bar{x} - 2.261\frac{s}{\sqrt{10}} \ , \ \bar{x} + 2.261\frac{s}{\sqrt{10}}\right].$$

For our data, this interval estimate equals [3263 , 5126] and so we infer that the mean donation per return is between \$3263 and \$5126.

Inconsistency

The glaring inconsistency here is that our test conclusion was that the mean is less than \$5000 while our confidence interval includes the possible value of \$5000 for this same mean.

If elementary texts are followed slavishly, this is the sort of potential inconsistency with which students are left.

Consistency Rescued

The cure for this problem is a general equivalence which, in itself, is potent and practical for basic students:

A test of the null hypothesis **H₀**:$\theta=\theta_0$ *will* <u>*not*</u> *reject* **H₀** *at level* α

if and only if

θ_0 *is INside the **appropriate** confidence interval for* θ *which has coefficient* $1-\alpha$.

As a consequence of this equivalence, 1-sided confidence intervals are needed for consistency with 1-sided tests. We illustrate this need with the following derivation for the methods used in our example.

Our test of **H$_0$**: $\mu = \mu_0$ vs. **H$_a$**: $\mu < \mu_0$ will not reject **H$_0$** if and only if $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} \geq -t_\alpha$. If **H$_0$**:$\mu=\mu_0$ is true, this has probability $P\left(\frac{\bar{x}-\mu_0}{s/\sqrt{n}} \geq -t_\alpha\right) = 1-\alpha$. Rewriting this expression yields $1-\alpha = P\left(\bar{x}+t_\alpha \frac{s}{\sqrt{n}} \geq \mu\right)$. Thus we have a 1-sided confidence interval for μ: $\left(-\infty, \bar{x}+t_\alpha \frac{s}{\sqrt{n}}\right]$

While this derivation is probably inappropriate for most basic Statistics classes, such an interval can be heuristically justified: *If μ is less than μ_0, the estimation question is how large μ might be.* The above interval answers precisely that question.

Example revisited

For our data, the 95%, 1-sided confidence interval for the the mean μ of all U.S. cities' donation per return is

$$\mu \leq \bar{x}+t_\alpha \frac{s}{\sqrt{n}}, \text{ i.e.}$$

$$\mu \leq \bar{x}+1.833\frac{s}{\sqrt{10}}.$$

For our data, the interval estimate is $\mu <$ \$4950.

This interval **is** consistent with the test result because it EXcludes the null value of μ=\$5000 which the test rejected.

Pedagogical Implications

We advocate that, on the basis of the previous discussion, a basic Statistics course ought to include 1-sided confidence intervals. Furthermore, we claim that the equivalence between confidence intervals and tests ought to be emphasized at every opportunity. The result should be removal of a potential problem and emphasis of an important connection.

Classroom Experience

It has been our experience that students fare well with the addition of 1-sided confidence intervals. To put this more modestly and in perspective, students still struggle with the bigger issues (e.g. the meaning of "confidence level" and the meaning of "p-values") but no more so than they would without 1-sided confidence intervals.

Depending on the coverage of a basic course, students may encounter settings where the equivalence is a bit subtle (e.g. chi-squared tests where the parameter in question is a non-centrality parameter).

Resources

Reference sheets to support the preceding approach are available at the following URL:

http://alpha2.enc.edu/~constant/library/statistics/TofC.htm

Each sheet shows the three types of alternative hypothesis available and the corresponding confidence interval. The settings covered include

- 1-sample
  - mean
  - variance
  - proportion
- 2-sample
  - difference of means
  - ratio of variances
  - difference of proportions