Confidence Interval Applications Solutions

Lt Col Ken Horton   Lt Col Kris Pruitt   Professor Bradley Warner

25 October, 2020

Exercises

1. Chronic illness

In 2013, the Pew Research Foundation reported that "45% of U.S. adults report that they live with one or more chronic conditions".[1] However, this value was based on a sample, so it may not be a perfect estimate for the population parameter of interest on its own. The study reported a standard error of about 1.2%, and a normal model may reasonably be used in this setting.

   a. Create a 95% confidence interval for the proportion of U.S. adults who live with one or more chronic conditions. Also interpret the confidence interval in the context of the study.

$0.45 \pm 1.96 \times 0.012 = (0.426, 0.474)$ We are 95% confident that 42.6% to 47.4% of U.S. adults live with one or more chronic conditions.

   b. Create a 99% confidence interval for the proportion of U.S. adults who live with one or more chronic conditions. Also interpret the confidence interval in the context of the study.

```
qnorm(.995)
```

```
## [1] 2.575829
```

```
0.45 +c(-1,1)*qnorm(.995)*0.012
```

```
## [1] 0.41909 0.48091
```

We are 99% confident that 41.9% to 48.1% of U.S. adults live with one or more chronic conditions.

   c. Identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

   - We can say with certainty that the confidence interval from part a contains the true percentage of U.S. adults who suffer from a chronic illness.

False, we're only 95% confident.

---

[1]http://pewinternet.org/Reports/2013/The-Diagnosis-Difference.aspx The Diagnosis Difference. November 26, 2013. Pew Research.

- If we repeated this study 1,000 times and constructed a 95% confidence interval for each study, then approximately 950 of those confidence intervals would contain the true fraction of U.S. adults who suffer from chronic illnesses.

True, this is the definition of the confidence level.

- The poll provides statistically significant evidence (at the $\alpha = 0.05$ level) that the percentage of U.S. adults who suffer from chronic illnesses is not 50%.

True, the equivalent significance level of a two-sided hypothesis test for a 95% confidence interval is indeed 5%, and since the interval lies below 50% this statement is correct.

- Since the standard error is 1.2%, only 1.2% of people in the study communicated uncertainty about their answer.

False, the 1.2% measures the uncertainty associated with the sample proportion (the point estimate) not the uncertainty of individual observations, uncertainty in the sense of not being sure of one's answer to a survey question.

- Suppose the researchers had formed a one-sided hypothesis, they believed that the true proportion is less than 50%. We could find an equivalent one-sided 95% confidence interval by taking the upper bound of our two-sided 95% confidence interval.

False. In constructing a one-sided confidence interval we need α to be in one tail. Only taking the upper value of a two-sided 95% confidence interval leads to a 97.5% one-sided confidence interval, a more conservative value.

```r
qnorm(.95)
```

```
## [1] 1.644854
```

```r
0.45 +qnorm(.95)*0.012
```

```
## [1] 0.4697382
```

Notice that 46.9 is smaller than 47.4.

2. Vegetarian college students

Suppose that 8% of college students are vegetarians. Determine if the following statements are true or false, and explain your reasoning.

a. The distribution of the sample proportions of vegetarians in random samples of size 60 is approximately normal since $n \geq 30$.

FALSE. For the distribution of $\hat{p}$ to be approximately normal, we need to have at least 10 successes and 10 failures in our sample.

b. The distribution of the sample proportions of vegetarian college students in random samples of size 50 is right skewed.

TRUE. The success-failure condition is not satisfied

$$np = 50 \times 0.08 = 4$$

and

$$n(1 - p) = 50 \times 0.92 = 46$$

therefore we know that the distribution of $\hat{p}$ is not approximately normal. In most samples we would expect $\hat{p}$ to be close to 0.08, the true population proportion. While $\hat{p}$ can be as high as 1 (though we would expect this to effectively never happen), it can only go as low as 0. Therefore the distribution would probably take on a right-skewed shape. Plotting the sampling distribution would confirm this suspicion.

c. A random sample of 125 college students where 12% are vegetarians would be considered unusual.

FALSE.

$$SE_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$= \sqrt{\frac{.08(.92)}{125}} = 0.0243$$

A $\hat{p}$ of 0.12 is only $\frac{0.12 - 0.08}{0.0243} = 1.65$ standard errors away from the mean, which would not be considered unusual.

The p-value is:

```
2*(1-pnorm(1.65))
```

```
## [1] 0.09894294
```

d. A random sample of 250 college students where 12% are vegetarians would be considered unusual.

TRUE.

$$SE_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$= \sqrt{\frac{.08(.92)}{250}} = 0.0172$$

Notice that doubling the sample size only reduced the standard error by $\sqrt{2}$.

A $\hat{p}$ of 0.12 is $\frac{0.12 - 0.08}{0.0172} = 2.32$ standard errors away from the mean, which might be considered unusual.

The p-value is:

```
2*(1-pnorm(2.32))
```

```
## [1] 0.02034088
```

  e. The standard error would be reduced by one-half if we increased the sample size from 125 to 250.

FALSE. Since n appears under the square root sign in the formula for the standard error, increasing the sample size from 125 to 250 would decrease the standard error of the sample proportion only by a factor of $\sqrt{2}$.

  f. A 99% confidence will be wider than a 95% because to have a higher confidence level requires a wider interval.

TRUE. The width is a function of the margin of error. Keeping all else the same, the critical values for 95% and 99% are

```
temp<-qnorm(c(.985,.995))
names(temp)<-c("95%","99%")
```

```
temp
```

```
##      95%      99%
## 2.170090 2.575829
```

3. Orange tabbies

Suppose that 90% of orange tabby cats are male. Determine if the following statements are true or false, and explain your reasoning.
a. The distribution of sample proportions of random samples of size 30 is left skewed.

TRUE. The success-failure condition is not satisfied

$$n\hat{p} = 30 \times 0.90 = 27$$

and

$$n(1 - \hat{p}) = 30 \times 0.10 = 3;$$

therefore we know that the distribution of $\hat{p}$ is not nearly normal. In most samples we would expect $\hat{p}$ to be close to 0.90, the true population proportion. While $\hat{p}$ can be as low as 0 (though we would expect this to happen very rarely), it can only go as high as 1. Therefore the distribution would probably take on a left-skewed shape. Plotting the sampling distribution would confirm this suspicion.

  b. Using a sample size that is 4 times as large will reduce the standard error of the sample proportion by one-half.

TRUE. Since n appears in a square root for SE, using a sample size that is 4 times as large will reduce the SE by half.

  c. The distribution of sample proportions of random samples of size 140 is approximately normal.

TRUE. The success-failure condition is satisfied

$$n\hat{p} = 140 \times 0.90 = 126$$

and

$$n(1 - \hat{p}) = 140 \times 0.10 = 14;$$

therefore the distribution of $\hat{p}$ is nearly normal.

4. Working backwards

A 90% confidence interval for a population mean is (65,77). The population distribution is approximately normal and the population standard deviation is unknown. This confidence interval is based on a simple random sample of 25 observations. Calculate the sample mean, the margin of error, and the sample standard deviation.

The sample mean is the midpoint of the confidence interval:

$$\bar{x} = \frac{65 + 77}{2} = 71$$

The margin of error is half the width of the confidence interval:

$$ME = \frac{(77 - 65)}{2} = 6$$

Using df = 25 - 1 = 24 and the confidence level of 90% we can find the critical value from the t-table distribution.

```
qt(.95,24)
```

```
## [1] 1.710882
```

Lastly, using the margin of error and the critical value we can solve for s:

$$ME = t_{24} \times \frac{s}{\sqrt{n}}$$

$$6 = 1:71 \times \frac{s}{\sqrt{25}}$$

$$s = 17.54$$

5. Find the p-value

An independent random sample is selected from an approximately normal population with an unknown standard deviation. Find the p-value for the given set of hypotheses and T test statistic. Also determine if the null hypothesis would be rejected at $\alpha = 0.05$.

    a. $H_A : \mu > \mu_0$, $n = 11$, $T = 1.91$

```
1-pt(1.91,10)
```

```
## [1] 0.04260244
```

The p-value is less than 0.05, reject the null hypothesis.

b. $H_A : \mu < \mu_0$, $n = 17$, $T = -3.45$

```
pt(-3.45,16)
```

```
## [1] 0.001646786
```

The p-value is less than 0.05, reject the null hypothesis.

c. $H_A : \mu \neq \mu_0$, $n = 7$, $T = 0.83$

```
2*(1-pt(0.83,6))
```

```
## [1] 0.4383084
```

The p-value is greater than 0.05, fail to reject the null hypothesis.

d. $H_A : \mu > \mu_0$, $n = 28$, $T = 2.13$

```
1-pt(2.13,27)
```

```
## [1] 0.02121769
```

The p-value is less than 0.05, reject the null hypothesis.

6. Sleep habits of New Yorkers

New York is known as "the city that never sleeps". A random sample of 25 New Yorkers were asked how much sleep they get per night. Statistical summaries of these data are shown below. Do these data provide strong evidence that New Yorkers sleep less than 8 hours a night on average?

n	$\bar{x}$	s	min	max
25	7.73	0.77	6.17	9.78

a. Write the hypotheses in symbols and in words.

$H_0 : \mu = 8$ New Yorkers sleep 8 hrs per night on average.

$H_A : \mu < 8$ New Yorkers sleep less than 8 hrs per night on average.

b. Check conditions, then calculate the test statistic, T, and the associated degrees of freedom.

1. Independence: The sample is random and 25 is less than 10% of all New Yorkers, so it seems reasonable that the observations are independent.
2. Sample size: Sample size is less than 30, therefore we use a t-test which implies the population must be normal.
3. Skew: We don't have the data to look at a qq plot so we will make some guesses. All observations are within three standard deviations of the mean. For now we will proceed while acknowledging that we are assuming the skew is perhaps moderate or less (moderate skew would be acceptable for this sample size).

The test statistic and degrees of freedom can be calculated as follows:

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

$$\frac{7.73 - 8}{\frac{0.77}{\sqrt{25}}} = -1.75$$

$df = 25 - 1 = 24$.

    c. Find and interpret the p-value in this context.

```
pt(-1.75,24)
```

```
## [1] 0.04644754
```

If in fact the true population mean of the amount New Yorkers sleep per night was 8 hours, the probability of getting a random sample of 25 New Yorkers where the average amount of sleep is 7.73 hrs per night or less is 0.046. This p-value is close to 0.05.

    d. What is the conclusion of the hypothesis test?

Since the p-value is less than 0.05, we reject the null hypothesis that New Yorkers sleep an average of 8 hours per night in favor of the alternative that they sleep less than 8 hours per night on average. However, the p-value is close to the significance level and we may want to run the study again and/or look at sample sizes to determine how big of a difference from 8 hours is important from a practical standpoint.

Let's look at the power of this test. Remember that power is the probability of rejecting the null when the alternative is true. Since the alternative specifies a range of values for the parameter, we must specify a value. This is done by subject matter experts. How much of a difference in average sleep is needed from a practical standpoint to say it is different? Let's say that a half an hour is important to detect. We will use the function `power.t.test()` to determine the power of our test.

```
power.t.test(n=25,delta=.5,sd=.77,alternative = "one.sided",type="one.sample")
```

```
##
##      One-sample t test power calculation
##
##              n = 25
##          delta = 0.5
##             sd = 0.77
##      sig.level = 0.05
##          power = 0.9342637
##    alternative = one.sided
```

This is a high level of power, the typical value used by researchers is 80%.

Let's see what the power is for a 15 minute difference.

```
power.t.test(n=25,delta=.25,sd=.77,alternative = "one.sided",type="one.sample")
```

```
##
##      One-sample t test power calculation
##
##              n = 25
##          delta = 0.25
##             sd = 0.77
##      sig.level = 0.05
##          power = 0.4731184
##    alternative = one.sided
```

This is a much lower power and thus not an effective test with that sample size for finding that small of a difference. We can turn the problem around and ask what sample size we need for 80% power?

```
power.t.test(power=.8,delta=.25,sd=.77,alternative = "one.sided",type="one.sample")
```

```
##
##      One-sample t test power calculation
##
##              n = 60.02642
##          delta = 0.25
##             sd = 0.77
##      sig.level = 0.05
##          power = 0.8
##    alternative = one.sided
```

We need 60 subjects in the study.

   e. Construct a 95% confidence interval that corresponded to this hypothesis test, would you expect 8
      hours to be in the interval?

We need an 95% upper bound on the confidence interval.

$$\bar{x} + t_{24,0.95}\frac{s}{\sqrt{n}} =$$

The critical value is

```
qt(.95,24)
```

```
## [1] 1.710882
```

So the upper confidence bound is

```
7.73+qt(.95,24)*0.77/sqrt(25)
```

```
## [1] 7.993476
```

The value of 8 is not included in the interval, but it is close. We need more data or another study to be confident in our results.

7. Vegetarian college students II

From problem 2 part c, suppose that it has been reported that 8% of college students are vegetarians. We think USAFA is not typical because of their fitness and health awareness, we think there are more vegetarians. We collect a random sample of 125 cadets and find 12% claimed they are vegetarians. Is there enough evidence to claim that USAFA cadets are different?

    a. Use `binom.test()` to conduct the hypothesis test and find a confidence interval.

```
binom.test(x=15,n=125,p=.08,alternative = "greater")
```

```
##
##
##
## data:  15 out of 125
## number of successes = 15, number of trials = 125, p-value = 0.07483
## alternative hypothesis: true probability of success is greater than 0.08
## 95 percent confidence interval:
##  0.07544411 1.00000000
## sample estimates:
## probability of success
##                   0.12
```

We fail to reject. Notice 0.08 is in the interval.

    b. Use `prop.test()` with `correct=FALSE` to conduct the hypothesis test and find a confidence interval.

```
prop.test(x=15,n=125,p=.08,alternative = "greater",correct=FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  15 out of 125
## X-squared = 2.7174, df = 1, p-value = 0.04963
## alternative hypothesis: true p is greater than 0.08
## 95 percent confidence interval:
##  0.08007111 1.00000000
## sample estimates:
##    p
## 0.12
```

We reject, what is going on?

    c. Use `prop.test()` with `correction=TRUE` to conduct the hypothesis test and find a confidence interval.

```
prop.test(x=15,n=125,p=.08,alternative = "greater",correct=TRUE)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  15 out of 125
```

```
## X-squared = 2.2011, df = 1, p-value = 0.06896
## alternative hypothesis: true p is greater than 0.08
## 95 percent confidence interval:
##  0.07682087 1.00000000
## sample estimates:
##    p
## 0.12
```

We fail to reject.

    d. Which test should you use?

Go into the help for `binom.test()` in R and it explains these intervals. The way to compare these is to look at coverage. By this we mean for a 95% confidence interval, does the interval include the true parameter 95% of the time? This can be checked by simulation, but mathematics are really needed for a more definitive answer.

The first test is an exact test and as stated in the help, it guarantees that the coverage rate is at least 95%. This means the interval tends to be too large. This is the largest interval and tends to be conservative.

The second test is the **Score** test, also called the **Wilson**. It is found by inverting the p-value, beyond the scope of this class. This interval is a nice compromise in its coverage.

The third is still a score but in the calculation of the p-value, a continuity correction is applied. This correction takes the limits on the binomial and extends it 0.5 in each direction. This is done to give the discrete binomial a better approximation to the continuous normal in the CLT. This interval is the default in R via `prop.test()`.

None of these are our simple confidence interval based on the normal approximation. Here is the code for it:

```r
binom.test(x=15,n=125,p=.08,alternative = "greater",ci.method = "Wald")
```

```
##
##  Exact binomial test (Wald CI)
##
## data:  15 out of 125
## number of successes = 15, number of trials = 125, p-value = 0.07483
## alternative hypothesis: true probability of success is greater than 0.08
## 95 percent confidence interval:
##  0.0721916 1.0000000
## sample estimates:
## probability of success
##                   0.12
```

```r
15/125-qnorm(.95)*sqrt(.12*.88/125)
```

```
## [1] 0.0721916
```

This interval is not very good. The coverage is poor especially for small sample sizes. To learn more, read *Approximate is better then exact for interval estimation of binomial proportions.* A. Agresti and B. A. Coull, American Statistician 52, 1998, 119-126.