Confidence Intervals and Hypothesis Testing – Solutions
STAT-UB.0001 – Statistics for Business Control

1. Of the Stern MBA students who filled out an online class survey, 42 reported their GMAT scores. The sample mean of the reported scores was 720, and the sample standard deviation was 35.

   (a) What is a reasonable population to associate with this sample?

   > **Solution:** The GMAT scores of all Stern MBA students.

   (b) What is the meaning of the "population mean"?

   > **Solution:** μ, the mean GMAT score of all Stern MBA students.

   (c) Find a 95% confidence interval for the population parameter.

   > **Solution:** We have $\bar{x} = 720$, $s = 35$, and $n = 42$. For a 95% confidence interval, $\alpha 0.050$ and $\alpha/2 = 0.025$; with $n = 42$, there are $n - 1 = 41$ degrees of freedom. Consulting the t table, we find
   >
   > $$t_{\alpha/2,n-1} = t_{0.025,41} \approx 2.021.$$
   >
   > (Note that $\mathrm{df} = 41$ is not in the table, so we look at $\mathrm{df} = 40$ instead.) The 95% confidence interval for μ is
   >
   > $$\begin{aligned} \bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= 720 \pm 2.021 \cdot \frac{35}{\sqrt{42}} \\ &= 720 \pm 10.9 \\ &= (709.1, 730.9) \end{aligned}$$

   (d) Under what conditions is the confidence interval valid?

   > **Solution:** Since n is large ($n \geq 30$), the interval is valid if we have a simple random sample.

2. Use the following sample means and sample standard deviations of other questions from that class survey to form 95% confidence intervals for the population mean of each variable.

   (a) Dinners per month: $\bar{x} = 9.0$, $s = 4.6$, $n = 47$.

   Solution:

   $$t_{\alpha/2,n-1} = t_{0.025,46} \approx 2.021$$

   $$\begin{aligned} \bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= 9.0 \pm 2.021 \cdot \frac{4.6}{\sqrt{47}} \\ &= 9.0 \pm 1.36 \\ &= (7.64, 10.36) \end{aligned}$$

   (b) Age (years): $\bar{x} = 26.7$, $s = 6.1$, $n = 47$.

   Solution:

   $$t_{\alpha/2,n-1} = t_{0.025,46} \approx 2.021$$

   $$\begin{aligned} \bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= 26.7 \pm 2.021 \cdot \frac{6.1}{\sqrt{47}} \\ &= 26.7 \pm 1.80 \\ &= (24.90, 28.50) \end{aligned}$$

   (c) Time planned for studying per week (hours): $\bar{x} = 15.75$, $s = 10.5$, $n = 46$.

   Solution:

   $$t_{\alpha/2,n-1} = t_{0.025,45} \approx 2.021$$

   $$\begin{aligned} \bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} &= 15.75 \pm 2.021 \cdot \frac{10.5}{\sqrt{46}} \\ &= 15.75 \pm 3.13 \\ &= (12.62, 18.88) \end{aligned}$$

3. In Problem 2, what assumptions do we need for the confidence intervals to be valid? How could we check these assumptions?

Solution: We need that the sample is a simple random sample. In this case, that is equivalent to the sample being unbiased: every member of the population has an equal chance of being selected. If the sample is *not* unbiased, then we need the bias to be unrelated to the quantities of interest. In particular, we need that "Dinners per Month," "Age," and "Study Time" are unrelated to the student's choice of statistics class.

We do *not* need that the population is normal, because $n \geq 30$ in each of these examples.

4. In each of the following situations, find α and $t_{\alpha/2,n-1}$.

   (a) An 80% confidence interval with $n = 10$.

   > **Solution:**
   >
   > $$\alpha = .20, \quad n-1 = 9 \text{ degrees of freedom}, \quad t_{.100,9} = 1.383.$$

   (b) A 99% confidence interval with $n = 25$.

   > **Solution:**
   >
   > $$\alpha = .01, \quad n-1 = 24 \text{ degrees of freedom}, \quad t_{.005,24} = 2.797$$

   (c) A 90% confidence interval with $n = 30$.

   > **Solution:**
   >
   > $$\alpha = .10, \quad n-1 = 29 \text{ degrees of freedom}, \quad t_{.050,29} \approx 1.701.$$
   >
   > Note that df $= 29$ is not in the table, so we approximate the value of $t_{.050,29}$ by using either df $= 28$ or df $= 30$ (either answer is acceptable).

5. A random sample of 36 measurements was selected from a population with unknown mean μ. The sample mean is $\bar{x} = 12$ and the sample standard deviation is $s = 18$. Calculate an approximate 95% confidence interval for μ. Use the approximation $t_{\alpha/2,n-1} = t_{0.025,35} \approx 2$.

> **Solution:** We compute a 95% confidence interval for μ via the formula $\bar{x} \pm t_{0.025,n-1}\frac{s}{\sqrt{n}}$. In this case, we get $12 \pm 2\frac{18}{\sqrt{36}}$ i.e., 12 ± 6.

6. Complete Problem 5, with a 99% confidence interval instead of a 95% confidence interval.

> **Solution:** For a $100(1-\alpha)\%$ confidence interval for μ, we use the formula $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. For a 99% confidence interval, we have $\alpha = .01$ and $z_{\alpha/2} = 2.576$. Thus, our confidence interval for μ is $12 \pm 2.576\frac{18}{\sqrt{36}}$ i.e., 12 ± 7.728.

7. Complete Problem 5, with an 80% confidence interval instead of a 95% confidence interval.

> **Solution:** For a $100(1-\alpha)\%$ confidence interval for μ, we use the formula $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. For a 80% confidence interval, we have $\alpha = .20$ and $z_{\alpha/2} = 1.282$. Thus, our confidence interval for μ is $12 \pm 1.282\frac{18}{\sqrt{36}}$ i.e., 12 ± 3.846.

8. How reliable is the SoHo Halal Guy's Yelp rating? The SoHo Halal Guy at Broadway and Houston (`http://www.yelp.com/biz/soho-halal-guy-new-york`) currently has 53 Yelp reviews (4 1-star; 1 2-star; 6 3-star; 17 4-star; and 25 5-star). The average star rating is 4.1 and the sample standard deviation of the star ratings is 1.2. How much should we trust the number "4.1"? We will use a confidence interval to quantify the uncertainty associated with this number.

   (a) What is a reasonable population to associate with this sample?

   Solution: All ratings of the Halal Cart (past and future).

   (b) What is the meaning of the population mean, μ?

   Solution: The parameter of interest is μ, the mean start rating of all people who ever review the Halal Cart. Equivalently, the μ is equal to expected star rating of a random Halal Cart reviewer.

   (c) Find a 95% confidence interval for the population mean, μ.

   Solution:
   For a 95% confidence interval, we have $\alpha = 0.05$ and $\alpha/2 = 0.025$. The sample size is $n = 53$. There are $n - 1 = 52$ degrees of freedom. Thus, using the t table, we have

   $$t_{\alpha/2,n-1} = t_{0.025,52} \approx 2.009.$$

   The 95% confidence interval for the population mean, μ, is

   $$\begin{aligned} \bar{x} \pm 2.009 \frac{s}{\sqrt{n}} &= 4.1 \pm 2.009 \frac{1.2}{\sqrt{53}} \\ &= 4.1 \pm 0.33 \\ &= (3.77, 4.43). \end{aligned}$$

   (d) Under what conditions is the confidence interval valid?

   Solution: For a confidence interval for a mean to be valid, we need that (i) the observed sample is a simple random sample from the population, and (ii) $n \geq 30$ or the population is normal. Clearly, assumption (ii) holds. Here, it is reasonable to assume (i) as long as the Halal Cart and its customer base do not change in the future.

9. La Colombe at Lafayette and 4th St (`http://www.yelp.com/biz/la-colombe-new-york-2/`) currently has 612 Yelp reviews (16 1-star; 24 2-star; 50 3-star; 185 4-star; and 337 5-star). The average star rating is 4.31 and the sample standard deviation of the star ratings is 0.96. Find a 95% confidence interval for the expected rating of a random La Colombe Yelp reviewer.

Solution: Since $n \geq 30$, we can approximate $t_{0.025,n-1} \approx 2$. (A more accurate approximation would be $t_{0.025,n-1} \approx 1.960$. An approximate 95% confidence interval for the population mean is

$$\begin{aligned}
\bar{x} \pm 2\frac{s}{\sqrt{n}} &= 4.31 \pm 2\frac{0.96}{\sqrt{612}} \\
&= 4.31 \pm 0.08 \\
&= (4.23, 4.39)
\end{aligned}$$

Confidence Interval for Proportion

10. A CNN/ORC post-debate poll surveyed 547 voters who watched the third presidential debate on October 19, 2016. The results are at `http://www.cnn.com/2016/10/19/politics/hillary-clinton-wins-third-presidential-debate-according-to-cnn-orc-poll`. Of the respondents, 52% thought that Clinton did the best job, while 39% thought that Trump did.

   (a) What is a reasonable population to associate with this sample?

   Solution: The opinions of all voters who watched the debate.

   (b) There are a few population parameters of interest. Choose one.

   Solution: p, the proportion of all debate-watching voters who thought that Clinton won. (Alternatively, the proportion of all debate-watching voters who voters who thought the candidates tied, or that Trump won.)

   (c) Find a 95% confidence interval for the population parameter.

   Solution: The sample proportion is $\hat{p} = 0.52$. The sample size is $n = 547$. For a 95% confidence interval, we have $\alpha = 0.05$ and $\alpha/2 = 0.025$. Thus,

   $$z_{\alpha/2} = z_{0.050} = 1.960$$

   (use the df $= \infty$ section of the t-table.) Thus, the 95% confidence interval for p is

   $$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.52 \pm 1.96\sqrt{\frac{(0.52)(1-0.52)}{547}} \\ &= 0.52 \pm 0.04 \\ &= (0.48, 0.56) \end{aligned}$$

   For the proportion of debate-watching voters who thought that Trump won, the confidence interval is

   $$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.39 \pm 1.96\sqrt{\frac{(0.39)(1-0.39)}{547}} \\ &= 0.39 \pm 0.04 \\ &= (0.35, 0.43) \end{aligned}$$

   (d) Under what conditions is the confidence interval valid?

   Solution: We need a simple random sample, and we need to have expected at least 15 successes and 15 failures ($np \geq 15$ and $n(1-p) \geq 15$). The latter condition is almost certainly satisfied since we had $n\hat{p} = 284$ successes and $n(1-\hat{p}) = 263$ failures in the sample. For the former condition, we need the sample to be unbiased.

11. Use the following data from a Stern MBA class survey to estimate the relevant population proportions. Give 95% confidence intervals for these proportions.

(a) Gender: 17 Female, 30 Male.

Solution: If we let p be the proportion of Female in the population, then we have $\hat{p} = \frac{17}{17+30} = 0.36$ and $n = 17 + 30 = 47$. The 95% confidence interval for p is

$$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.36 \pm 1.96\sqrt{\frac{(0.36)(1-0.36)}{47}} \\ &= 0.36 \pm 0.14 \\ &= (0.22, 0.50) \end{aligned}$$

(b) Drinks at least one cup of coffee on a typical day: 37 Yes, 10 No.

Solution: If we let p be the proportion of coffee drinkers in the population, then we have $\hat{p} = \frac{37}{37+10} = 0.79$ and $n = 37 + 10 = 47$. The 95% confidence interval for p is

$$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.79 \pm 1.96\sqrt{\frac{(0.79)(1-0.79)}{47}} \\ &= 0.79 \pm 0.12 \\ &= (0.67, 0.91) \end{aligned}$$

(c) Political affiliation: 36 Democrat, 6 Republican, 5 Other. (For this problem there are three different choices for the population parameter; choose one of them.)

Solution: If we let p be the proportion of the population that are Democrats. Then, we have $\hat{p} = \frac{36}{36+6+5} = 0.77$ and $n = 36 + 6 + 5 = 47$. The 95% confidence interval for p is

$$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.77 \pm 1.96\sqrt{\frac{(0.77)(1-0.77)}{47}} \\ &= 0.77 \pm 0.12 \\ &= (0.65, 0.99) \end{aligned}$$

If instead we let p be the proportion of the population that are Republicans, then we have $\hat{p} = \frac{6}{36+6+5} = 0.13$ and $n = 36 + 6 + 5 = 47$. The 95% confidence interval for p is

$$\begin{aligned} \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.13 \pm 1.96\sqrt{\frac{(0.13)(1-0.13)}{47}} \\ &= 0.13 \pm 0.10 \\ &= (0.03, 0.23) \end{aligned}$$

12. In Problem 11, what are the relevant populations?

Solution: All Stern MBA students: their genders, whether they drink coffee, and their political affiliations.

13. In Problem 11, what assumptions do we need for the confidence intervals to be valid?

Solution: We need a simple random sample, and we need to have expected at least 15 successes and 15 failures in each sample. Since we do not have at least 15 successes and 15 failures in examples (b) and (c), these intervals are only approximate, and the true confidence level may be below 95%.

Hypothesis Test: Introduction

14. An analyst claims to have a reliable model for Yahoo's quarterly revenues. His model predicted that the most recent quarterly revenues could be described as a normal random variable with mean \$1.5B and standard deviation \$0.1B. In actuality, the revenues were \$1.0B. Is there evidence of a problem with the analyst's model? Why or why not?

> **Solution:** The observed revenues were $(1.0 - 1.5)/(0.1) = 5$ standard deviations away from the expected value predicted by the analyst's model. If the analyst's model were correct, the chance of observing a deviation this large or larger would be extremely small (0.00005733%). This is evidence that there is a problem with the analyst's model.

15. David has a coin, which he claims to be fair (50% chance of "heads," and 50% chance of "tails"). He flips the coin 10 times, and gets "heads" all 10 times. Do you believe him that the coin is fair? Why or why not?

> **Solution:** If the coin were fair, there would be a 0.2% chance of getting the same outcome on all 10 flips. That is, the observed string of 10 heads in a row would be extremely unlikely. This is evidence that the coin is not fair.

Test on a Population Mean

16. (Adapted from Stine and Foster, 4M 16.2). Does stock in IBM return a different amount on average than T-Bills? We will attempt to answer this question by using a dataset of the 264 monthly returns from IBM between 1990 and 2011. Over this period, the mean of the monthly IBM returns was 1.26% and the standard deviation was 8.27%. We will take as given that the expected monthly returns from investing in T-Bills is 0.3%.

    (a) What is the sample? What are the sample mean and standard deviation?

    Solution: The $n = 264$ monthly IBM returns from 1990 to 2011. The sample mean and standard deviation (in %) are

    $$\bar{x} = 1.26$$
    $$s = 8.27$$

    (b) What is the relevant population? What are the interpretations of population mean and standard deviation?

    Solution: All monthly IBM returns (past, present, and future). The population mean, μ represents the expected return for a month in the future. The population standard deviation, σ, represents the standard deviation of the monthly returns for all months (past, present, and future).

    (c) What are the null and alternative hypotheses for testing whether or not IBM gives a different expected return from T-Bills (0.3%)?

    Solution:

    $$H_0 : \mu = 0.3$$
    $$H_a : \mu \neq 0.3$$

(d) Use an appropriate test statistic to summarize the evidence against the null hypothesis.

Solution: If the null hypothesis were true ($\mu = 0.3$), then the sample mean would have been a normal random variable with mean $\mu_{\bar{X}} = 0.3$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. The test statistic

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

would follow a t distribution with $n - 1 = 263$ degrees of freedom. The observed value of this statistic is

$$t = \frac{1.26 - 0.3}{8.27/\sqrt{264}} = 1.886$$

(e) If the null hypothesis were true (there were no difference in expected monthly returns between IBM and T-Bills) what would be the chance of observing data at least as extreme as observed?

Solution: If we approximate the distribution of the test statistic under H_0 as a standard normal random variable, then the chance of observing data at least as extreme as observed would be

$$p = P(|Z| > 1.886) \approx 0.05743.$$

(f) Is there compelling evidence (at significance level 5%) of a difference in expected monthly returns between IBM and T-Bills?

Solution: No, since $p \geq 0.05$, there is not compelling evidence.

(g) What assumptions do you need for the test to be valid? Are these assumptions plausible?

Solution: Since $n \geq 30$, we do not need to assume that the population is normal. We need that the observed sample is a simple random sample from the population; this might not hold if the period under observation (1990–2011) is not typical for IBM.

Test Statistic and Observed Significance Level (p-value)

17. In each of the following examples, for the hypothesis test with

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

find the test statistic (t) and the p-value.

(a) $\mu_0 = 5$; $\bar{x} = 7$; $s = 10$; $n = 36$.

Solution:

$$\begin{aligned} t &= \frac{7-5}{10/\sqrt{36}} \\ &= 1.2 \\ p &\approx P(|Z| > 1.2) \\ &= 0.2301 \end{aligned}$$

(b) $\mu_0 = 90$; $\bar{x} = 50$; $s = 200$; $n = 64$.

Solution:

$$\begin{aligned} t &= \frac{50-90}{200/\sqrt{64}} \\ &= -1.6 \\ p &\approx P(|Z| > 1.6) \\ &= 0.1096 \end{aligned}$$

(c) $\mu_0 = 50$; $\bar{x} = 49.4$; $s = 2$; $n = 100$.

Solution:

$$\begin{aligned} t &= \frac{49.4-50}{2/\sqrt{100}} \\ &= -3 \\ p &\approx P(|Z| > 3) \\ &= 0.002700 \end{aligned}$$

18. For each example from problem 17:

(a) Indicate whether a level 5% test would reject H_0.

Solution: We reject H_0 if $p < 0.05$: (a) do not reject H_0; (b) do not reject H_0; (c) reject H_0.

(b) Indicate whether a level 1% test would reject H_0.

Solution: We reject H_0 if $p < 0.01$: (a) do not reject H_0; (b) do not reject H_0; (c) reject H_0.