

**Confidence Intervals & Hypothesis Tests (Solutions)**  
STAT-UB.0003: Regression and Forecasting Models

## Confidence intervals

1. A random sample of 36 measurements was selected from a population with unknown mean  $\mu$  and known standard deviation  $\sigma = 18$ . The sample mean is  $\bar{x} = 12$ . Calculate a 95% confidence interval for  $\mu$ .

**Solution:** We compute a 95% confidence interval for  $\mu$  via the formula  $\bar{x} \pm 2 \frac{\sigma}{\sqrt{n}}$ . In this case, we get  $12 \pm 2 \frac{18}{\sqrt{36}}$  i.e.,  $12 \pm 6$ .

2. With respect to the previous problem, which of the following statements are true:
- (a) There is a 95% chance that  $\mu$  is between 6 and 18.
  - (b) The population mean  $\mu$  will be between 6 and 18 about 95% of the time.
  - (c) In 95% of all future samples, the sample mean will be between 6 and 18.
  - (d) The population mean  $\mu$  is between 6 and 18.
  - (e) None of the above.

**Solution:** The correct answer is (e). The numbers  $\mu$ , 6, and 18 are all nonrandom, so it makes no sense to talk about probability. Instead, we can say that we have 95% *confidence* that  $\mu$  is between 6 and 18. The term “confidence” denotes subjective belief, as opposed to “probability,” which is concerned with randomness.

3. Complete Problem 1, with a 99% confidence interval instead of a 95% confidence interval.

**Solution:** For a  $100(1 - \alpha)\%$  confidence interval for  $\mu$ , 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  $\mu$  is  $12 \pm 2.576 \frac{18}{\sqrt{36}}$  i.e.,  $12 \pm 7.728$ .

4. The SoHo Halal Guy at Broadway and Houston currently has 35 Yelp reviews<sup>1</sup> (1 one-star; 1 two-star; 5 three-star; 12 four-star; and 16 five-star). The average star rating is 4.17 and the sample standard deviation of the star ratings is 0.98.

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

**Solution:** All ratings of the SoHo Halal Guy (past and future).

(b) What is the meaning of the population mean,  $\mu$ ?

**Solution:** The population mean,  $\mu$ , the mean start rating of all people who ever review the SoHo Halal Guy. Equivalently, the  $\mu$  is equal to expected star rating of a randomly-chosen SoHo Halal Guy reviewer.

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

**Solution:**

For a 95% confidence interval, we have  $\alpha = 0.05$  and  $\alpha/2 = 0.025$ . The sample size is  $n = 35$ . There are  $n - 1 = 34$  degrees of freedom. Thus, using the t table, we have

$$t_{\alpha/2, n-1} = t_{0.025, 34} \approx t_{0.025, 30} = 2.042.$$

The 95% confidence interval for the population mean,  $\mu$ , is

$$\begin{aligned}\bar{x} \pm 2.042 \frac{s}{\sqrt{n}} &= 4.17 \pm 2.042 \frac{0.98}{\sqrt{35}} \\ &= 4.17 \pm 0.34 \\ &= (3.83, 4.51)\end{aligned}$$

(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. Here, it is reasonable to assume (i) as long as the SoHo Halal Guy and his customer base do not change in the future. Since rating can only be 1, 2, 3, 4 or 5, it is not reasonable to assume that the population is normal. However, since  $n \geq 30$ , assumption (ii) holds.

---

<sup>1</sup><http://www.yelp.com/biz/soho-halal-guy-new-york>

## Hypothesis tests

5. We collect a simple random sample of size  $n = 100$  from a population. The sample mean is  $\bar{x} = 12.4$  and the sample standard deviation is  $s = 8.0$ . Use this data to test the null hypothesis  $H_0 : \mu = 10.0$  against the alternative  $H_a : \mu \neq 10.0$ , where  $\mu$  denotes the population mean:
- (a) Compute the test statistic.

**Solution:** Since the population standard deviation ( $\sigma$ ) is unknown, we use a t-statistic:

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{(12.4) - (10.0)}{(8.0)/\sqrt{(100)}} \\ &= 3.00 \end{aligned}$$

- (b) Use a z-table to compute an approximate p-value.

**Solution:** The t-statistic has  $n - 1 = 99$  degrees of freedom, which is relatively large ( $\geq 30$ ). So, using Table II from Appendix D, an approximate p-value is

$$\begin{aligned} p &\approx P(|Z| \geq 3.00) \\ &= 1 - 2(.4987) \\ &= .0026. \end{aligned}$$

The problem does not ask for it, but we can use Minitab or a calculator to compute the exact p-value:

$$\begin{aligned} p &= P(|T_{99}| \geq 3.00) \\ &= .0034. \end{aligned}$$

(Here,  $T_{99}$  denotes a random t-statistic with 99 degrees of freedom.)

- (c) What is the meaning of the p-value? Give a one-sentence description.

**Solution:** If the population mean were equal to 10.0, then the chance of seeing data at least as extreme as observed would be  $p \approx .0026$ .

A more precise sentence is the following: if the population mean were equal to 10.0 and we were to repeat the experiment—collecting a new sample—then the chance of seeing a test statistic for the new sample at least as extreme as the test statistic for the observed sample would be  $p \approx .0026$ .

(d) Using a significance level ( $\alpha$ ) of 5%, what is the result of the hypothesis test?

**Solution:** Since  $p < .05$ , we would reject the null hypothesis at significance level 5%.

6. National Public Radio's *Planet Money* podcast performed an experiment to measure the "wisdom of the crowd" with regard to estimating the weight of a cow<sup>2</sup>. After being shown a picture of a cow, respondents were asked to guess its weight, in pounds. The mean of the 17,109 guesses was 1282 pounds, and the standard deviation was 534. The true weight of the cow was 1355 pounds. So, the crowd of 17,109 respondents under-estimated the weight of the cow by 73 pounds. It's possible that a larger crowd could do better. Given the data available, is this plausible? That is, is it plausible that with a large enough crowd, the estimation error could be made arbitrarily small? We will answer this by performing a hypothesis test.

(a) What is the sample? What is the population? What is the interpretation of the population mean,  $\mu$ ?

**Solution:**  
 The sample is the reported  $n = 17109$  guesses. One possible population is the (hypothetical) guesses of all respondents, if the experiment were allowed to run forever, accumulating more and more respondents. That is, the population is a hypothetical "infinite crowd". The population mean is the average guess of all respondents in the population.

(b) What are the null and alternative hypotheses, in terms of  $\mu$ ?

**Solution:**

$H_0 : \mu = 1355$  (the "infinite crowd" would get the weight exactly right)  
 $H_a : \mu \neq 1355$  (the "infinite crowd" would err in estimating the weight)

(c) Compute the test statistic.

**Solution:**

$$\begin{aligned}
 t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\
 &= \frac{(1282) - (1355)}{(534)/\sqrt{(17109)}} \\
 &= -17.88
 \end{aligned}$$

(d) Use a z-table to compute an approximate p-value; if the test statistic falls outside the z-table, report  $p < .0001$ .

**Solution:**

<sup>2</sup><http://www.npr.org/sections/money/2015/08/07/430372183/episode-644-how-much-does-this-cow-weigh>

An approximate p-value is

$$p \approx P(|Z| > 17.88) \\ < .0001$$

(e) In the context of the problem, what is the interpretation of the p-value?

**Solution:** If the “infinite crowd” were completely accurate, then the chance of seeing data like observed would be less than .0001 (less than 0.01%). This is very compelling evidence that the infinite crowd is not completely accurate.

(f) Is it plausible that with a large enough crowd, the estimation error could be made arbitrarily small?

**Solution:** As part (e) notes, this is very implausible.  
Note: it is impossible to assign a probability to the statement. In particular, the p-value does *not* give the probability that the statement (the null hypothesis) is true.