Week 6 Homework

Steven Simonsen

2024-02-23

5.2 Identify the parameter, Part II: For each of the following situations, state whether the parameter of interest is a mean or a proportion.

(a) A poll shows that 64% of Americans personally worry a great deal about federal spending and the budget deficit.

Proportion - Each American in the poll reports whether or not they worry about federal spending and the budget deficit (yes/no), so this is a categorical variable.

(b) A survey reports that local TV news has shown a 17% increase in revenue within a two year period while newspaper revenues decreased by 6.4% during this time period.

Mean - The local TV news reported a percentage increase and decrease over a time period, and we can average over these percentages.

(c) In a survey, high school and college students are asked whether or not they use geolocation services on their smart phones.

Proportion - Each student answers whether or not they use geolocation services on their smart phones (yes/no), so this is a categorical variable.

(d) In a survey, smart phone users are asked whether or not they use a web-based taxi service.

Proportion - Each smart phone user answers whether or not they use a web-based taxi service (yes/no), so this is a categorical variable.

(e) In a survey, smart phone users are asked how many times they used a web-based taxi service over the last year.

Mean - Each smart phone user reports a numerical value: how many times they used a web-based taxi service.

5.4 Unexpected expense. In a random sample 765 adults in the United States, 322 say they could not cover a $400 unexpected expense without borrowing money or going into debt.

(a) What population is under consideration in the data set?

Adults in the United States.

(b) What parameter is being estimated?

The fraction of adults in the United States who could not cover a $400 unexpected expense without going into debt or borrowing money.

(c) What is the point estimate for the parameter?

322/765 p^ = .421 or 42.1%

(d) What is the name of the statistic we use to measure the uncertainty of the point estimate?

Standard Error

(e) Compute the value from part (d) for this context.

sqrt((1-.421)/765) SE=.0275 or 2.8%

Note: Using p hat (p^) for this estimate, as the population proportion (p), is unknown.

(f) A cable news pundit thinks the value is actually 50%. Should she be surprised by the data?

50-42.1=7.9% difference in p^ 7.9/2.8=2.82 standard deviations difference

The standard error (.0275) is the standard deviation of p^. 50% is almost 3 standard deviations away from the observed value of p^ (42.1%) since we know the Standard Error, or standard deviation of p^ to be 2.8%. Using the general rule of thumb that an observed value beyond 2 standard deviations is unusual, this would be surprising.

(g) Suppose the true population value was found to be 40%. If we use this proportion to recompute the value in part (e) using p = 0:4 instead of ^p, does the resulting value change much?

sqrt((1-.4)/765) SE = .0280 or 2.8%. The resulting value does not change much, due to n being a fairly large sample size.

5.10 Twitter users and news, Part II: A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter, and the standard error for this estimate was 2.4%. Identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

(a) The data provide statistically significant evidence that more than half of U.S. adult Twitter users get some news through Twitter. Use a significance level of alpha = 0.01. (This part uses concepts from Section 5.3 and will be corrected in a future edition.)

False.

print(round(qnorm(1-.01/2),2))

z = 2.58

52+2.58x2.4=58.19%

52-2.58x2.4=45.81%

99% Confidence interval (45.81%, 58.19%). Due to this finding, the answer to this question is false. We are 99% confident that the proportion of U.S. adult Twitter users who get at least some news from Twitter falls between 45.81% and 58.19%. Because 50% is included in the confidence interval, the data does not provide statistically significant evidence, as stated above. This means that in a hypothesis test, we would accept the null hypothesis, or status quo, that 50% of U.S adult Twitter users get some news through Twitter.

(b) Since the standard error is 2.4%, we can conclude that 97.6% of all U.S. adult Twitter users were included in the study.

False - The standard error describes the uncertainty in the overall estimate from natural fluctuations due to randomness, not the percentage of participants included in the study.

(c) If we want to reduce the standard error of the estimate, we should collect less data.

False - The standard error is equal to the square root of (1-p)/n, where p is the population parameter and n is the sample size. If less data is collected, n is smaller, which in turn increases the size of the standard error.

(d) If we construct a 90% confidence interval for the percentage of U.S. adults Twitter users who get some news through Twitter, this confidence interval will be wider than a corresponding 99% confidence interval.

False. In a lesser confidence interval, the margin of error (z*x SE) is less than in a larger confidence interval. In other words, as our level of confidence decreases, the range of our confidence interval becomes narrower. Below is the calculation of a 90% confidence interval to prove this theory, assuming the same p value and standard error.

print(round(qnorm(1-.1/2),2))

z=1.64

52-1.64x2.4=48.06

52+1.64x2.4=55.94

5.22 Getting enough sleep: 400 students were randomly sampled from a large university, and 289 said they did not get enough sleep. Conduct a hypothesis test to check whether this represents a statistically significant difference from 50%, and use a significance level of 0.01.

H0: p=0.5, HA: p not=0.5. Alpha (significance level) = 0.01.

289/400 p hat = .723

Check intdependence: simple random sampling gets us independence. Success/failure conditions are satisfied since 400*.723=289.2 and 400(1-.723)=110.8, which are both greater than or equal to 10.

Next, calculate standard error: sqrt((1-.723)/400) Standard Error = .027

Next. calculate z: round(qnorm(1-.01/2),2) z = 2.58

calculate p-value: 2*pnorm(-abs(2.58)) p-value=.001

Next, calculate 99% confidence interval: .723 - 2.58 x .027=.6533

.723 + 2.58 x .027=.7927

99% Confidence Interval(65.33%, 79.27%)

The p-value is less than alpha=.05. Due to this, and based on confirmation of independence and the success/failure condition, we would reject the null hypothesis. Additionally, 50% falls outside of our 99% confidence interval. The null hypothesis states that 50% of students from a large university did not get enough sleep. In other words, we accept the alternative hypothesis, which states there is a significant difference from 50% percent of the students from a large university stating they do not get enough sleep. We can state with 99% confidence that between 65.33% and 79.27% of large university students do not get enough sleep. Since 50% is outside of the confidence interval, the result is statistically significant.