Selected Problems

Solutions

Problems

AMS-UCSC

June 5th, 2015
Topics

We will talk about...

1. Solutions to Selected Problems
Problem 1

A simple random sample of size 400 was taken from the population of all manufacturing establishments in a certain state. The results are that 16 establishments had 250 employees or more.

1. Estimate the percentage of manufacturing establishments with 250 employee or more.
   4%

2. Attach a standard error to the estimate.

\[
\frac{\sqrt{.04 \times .96}}{\sqrt{400}} \approx .01
\]
Problem 2

A simple random sample of 1,000 persons is taken to estimate the percentage of Democrats in a large population. It turns out that 543 of the people in the sample are Democrats.

1. Calculate the sample percentage and SE for the sample percentage.
   The sample percentage is given by: \((543/1,000) \times 100\% = 54.3\%\)
   The SE for the sample percentage is:
   \(\sqrt{(0.543) \times (0.457)}/\sqrt{1,000} \times 100\% = 1.6\%\)

2. Find an approximate 95% confidence interval for the percentage of Democrats in the population.
   Two SE’s corresponds to 3.2%. Thus a 95% confidence interval is given by (51.1%, 57.5%)
The speed of light is measured 25 times by a new procedure. The 25 measurements are recorded and show no trend or pattern. The average of the measurements is 299,789.2 kilometers per second and the SD is 12 kilometers per second. Find an approximate 95% confidence interval for the speed of light.

1. Calculate the SE of the average.
   The SE is given by $\frac{12}{\sqrt{25}} = 12/5 = 2.4$.

2. Find an approximate 95% confidence interval for the speed of light.
   Two SE’s correspond to 4.8 km per second. Thus a 95% confidence interval is given by $(299,784.4, 299,794)$. 
Problem 4

Find the area under a Student’s $t$ curve with 3 degrees of freedom in the following cases:

1. To the right of 2.35.
   5%

2. To the left of -2.35.
   5%

3. Between -2.35 and 2.35.
   90%

4. Are these values higher or lower than the ones that correspond to the standard normal curve?
   a) and b) are smaller for the normal, as a consequence, c) is larger.
Problem 5

Looking at data and making sense of them is the first step of a statistical analysis.
The scatter diagram below shows the ages of 1,000 husbands and wives in a town in California. Explore the plot. Is there anything wrong with the data?

The range of $x$ does not correspond to the usual range of married men. In particular, there is a 5 years old man married to a 20 years old woman.
True or false:

1. To make a $t$ test with 4 measurements use a Student’s $t$ curve with 4 degrees of freedom.  
   F

2. For a given experiment the null hypothesis is that the average is equal to 231 units. The alternative hypothesis is that the average is above 231 units. You compute a $z$-test and the corresponding value $P$-value is 2.5%. The conclusion is that the probability that the average is equal to 231 units is 2.5%.  
   F

3. The R.M.S. error for a regression line of $y$ on $x$ is less than or equal to the SD of $y$.  
   T
True or false

1. The correlation between the daily minimum temperatures of L.A. and San Francisco is higher when measured in Fahrenheit than when it is measured in Celsius.
   F

2. The correlation between two variables is -.92, this implies that there is a strong negative linear association between the variables.
   T
Problem 7

A newspaper article says that on average, college freshmen spend 7.5 hours a week going to parties. One administrator does not believe that these figures apply at her college, which has nearly 3,000 freshmen. She takes a simple random sample of 100 freshmen, and interviews them. On average, they report 6.6 hours a week going to parties, and the SD is 9 hours.

1. Formulate the null and the alternative hypothesis.
   \( H_0 : \) The average number of hours a week that college freshmen go to parties is 7.5
   \( H_1 : \) The average number of hours a week that college freshmen go to parties is less that 7.5

2. Is the difference between 6.6 and 7.5 real?

   \[
   \frac{6.6 - 7.5}{9/\sqrt{100}} = -1
   \]

3. What is your conclusion? There is not enough evidence in the data to reject the null hypothesis since the \( P \)-value is close to 0.16.
Problem 8

A statistical analysis is made of the midterm and final scores in a large class. The results are

average midterm score \( \approx 60, \) SD \( \approx 15 \)
average final score \( \approx 65, \) SD \( \approx 20, \) r \( \approx 0.50 \)

1. Using the normal approximation, about what percentage of the students scored over 80 on the midterm?
80 points on the midterm corresponds to

\[
\frac{80 - 60}{15} = 1.33
\]

standard units. Using the normal we obtain that approximately 9% of the students scored over 80 on the midterm.

2. What is the R.M.S. error?

\[
\sqrt{1 - .5^2} \times 20 = 17.32
\]
Problem 8 (Cont.)

1. What is the slope of the regression line?
   \[ \frac{0.5 \times 20}{15} = 0.67 \]

2. What is the predicted final score for a student who scored 80 in the midterm?
   80 points on the midterm is 1.33 SD units above average. This corresponds to \[1.33 \times 0.5 = 0.67\] SD above average on the final. That corresponds to \[0.67 \times 20 = 13.4\] points over average on the final, so the students that scored 80 on the midterm, scored, on average, \[65 + 13.4 = 78.4\] on the final.

3. Of the students who scored 80 on the midterm, about what percentage scored over 80 on the final? In standard units we have
   \[ \frac{80 - 78.4}{17.32} = 0.09 \]
   and there is an area of about 46% to the right of this value under the normal curve.
Good luck in your final exam!!!!