ADM 2304 - Assignment 1

Renad Gharz

16/05/2022

Question 1 - Real Estate

The dataset RealEstate contains information on the listing of 1,047 real estate properties in a certain region.

1.1.a

Treating the data in column **Living Area [sq ft]** as the population, use software to find the population mean and the population standard deviation. Is the population data reasonably normal? Examine a boxplot and a histogram of the data in column **Living Area [sq ft]** to justify your answer. From here on, assume that the population standard deviation is not known.

1.1.b

Suppose that for a family of four, the ideal property size is between 2,400 and 2,800 square feet. Use software to code the data in column **Living Area [sq ft]** and put the result in column **Coded Data**. Code living areas between 2,400 and 2,800 square feet as "1" and code living areas outside of this interval as "0". Now calculate the population proportion using software and report the result.

1.1.c

The data in column **Sample 1** is a simple random sample drawn from the data in column **Living Area [sq ft]**. Calculate manually a 95% confidence interval for the population mean based on this sample and confirm your calculations using software.

1.1.d

Code the data in column **Sample 1** using the same instructions as in part b) above and put the result in column **Sample_p**. Calculate manually a 90% confidence interval for the population proportion of properties with living areas between 2,400 and 2,800 square feet and confirm your calculations using software. Assume that the required conditions are met, and you can use the normal approximation.

1.1.e

Now use software to randomly draw 19 additional samples of size n = 40 from column **Living Area [sq ft]**. The procedure must be repeated 19 times. Put these 19 samples in columns **Sample 2, Sample 3, . . . , Sample 19, Sample 20**. For each of these additional samples, use software to calculate the 95% confidence interval for the population mean.

1.1.f

Now count the number of confidence intervals, obtained from all the 20 samples, that contain the true value of the population mean from part a). Is this what you might expect? Explain your answer.

1.2.a

Using the data in column **Sample 1**, manually test the hypothesis that the population mean is not equal to 2,000 square feet. Use a 5% significance level and the critical value approach. Confirm your results using software. Is your conclusion supported by the confidence interval from part c)? Explain your answer.

1.2.b

Using the data in column Sample_p, manually test the hypothesis that the population proportion of properties that are ideal for a family of four is less than 20%. Use a 10% significance level. Calculate the p-value manually (i.e., using a normal distribution table) and explain how it confirms the conclusion reached by using the critical value approach. Assume that the normal approximation is reasonable in this case. Check your results using software.

1.3.a

Suppose you want to estimate the average living area of the real estate properties in the region. If you want to obtain a 95% confidence interval with a margin of error of $\pm$ 50 square feet, what sample size would you recommend? Assume for this exercise that the population standard deviation is 641 square feet.

1.3.b

Assume that you now would like to know what proportion of the real estate properties in the region are ideal for a family of four. This population proportion is not known. To estimate this population proportion with a margin of error of $\pm$ 0.02, what sample size would you recommend? Consider a 90% confidence level.

2. Package Subscribers

Bell provides cable, phone, and internet services to customers, some of whom subscribe to *packages* consisting of multiple services. Suppose that in Ontario 25% of Bell customers are package subscribers. A local Bell representative in Ottawa wonders if the proportion of package subscribers in the city is larger than the provincial proportion. After sending a survey to 100 customers from his subscriber list at random, only 25 of them responded, and of those, 11 are package subscribers. Does this constitute sufficient evidence that the true proportion of package subscribers in the Ottawa is more than the provincial proportion? Consider a 5% significance level and clearly explain the reasoning behind your answer.