**Research Project**

**Kenny Vo**

    Find a large data set.  Then following through each section of the  
    text, write a word problem using your data to form the question  
    and answer the question.  
    Try to turn this into a report that tells some sort of story.  
    Give me an idea what your results mean and how you could use it.

I've selected the "California Housing Prices" dataset, which provides details on each Californian unit community's median housing costs, population, and median income. The "California Housing Prices" dataset was obtained from Carnegie Mellon University's StatLib repository. It was published in the 1997 article "Sparse Spatial Autoregressions" in the journal Statistics and Probability Letters. There are 20,640 records in this collection.

Question 1: What is the California housing market's median price range as a whole?

Answer: The median price of a home in California ranges from $14,999 to $500,001.

Question 2: What are the median housing prices in California's mean and standard deviation?

Answer: In California, the median price of a home is $207,300 on average, with a $116,001 standard deviation.

Question 3: What is the likelihood that a randomly selected Californian unit group has a median property price above $300,000?

Answer: The chance is roughly 0.0668 or 6.68% using the normal distribution approximation with the mean and standard deviation determined.

Question: What is the likelihood that a randomly selected Californian unit group has a population of more than 10,000?

Answer: Using the statistics, it is estimated that 62.77% of unit groups in California have a population of more than 10,000.

Question 4: Using a sample of 500 unit groups, what is the 95% confidence interval for the population mean of California's median housing prices?

Answer: The 95% confidence interval is (196,487, 223,913) using the sample mean and standard deviation, a t-distribution with 499 degrees of freedom, and a confidence level of 95%.

Overall, I can see that the median house prices in California have a wide range, with a high mean and standard deviation. Additionally, there is a slim possibility of discovering a neighborhood with a median housing cost of more than $300,000. The population mean of median living prices can, however, be estimated with a high degree of confidence using a sample size of 500 block groups, with a 95% confidence interval that excludes the general minimum and maximum values. Real estate investors, developers, or governmental organizations could utilize these findings to inform their choices about housing development, cost, and policies in California.