**Quantitative Methods**

**List of Exercises N. 4**

**Selected Exercises from McClave (2014) – Chapters 5 and 6**

* 1. **The Concept of a Sampling Distribution**

1. (3). Consider the population described by the probability distribution shown below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 |
| p(x) | ,2 | 0,3 | 0,2 | 0,2 | 0,1 |

The random variable x is observed twice. If these observations are independent, verify that the different samples of size 2 and their probabilities are as shown in the next column.

|  |  |  |  |
| --- | --- | --- | --- |
| Sample | Probability | Sample | Probability |
| 1 , 1 | 0,04 | 3 ,4 | 0,04 |
| 1 , 2 | 0,06 | 3, 5 | 0,02 |
| 1 , 3 | 0,04 | 4, 1 | 0,04 |
| 1 , 4 | 0,04 | 4, 2 | 0,06 |
| 1 , 5 | 0,02 | 4, 3 | 0,04 |
| 2 , 1 | 0,06 | 4, 4 | 0,04 |
| 2 , 2 | 0,09 | 4, 5 | 0,02 |
| 2 , 3 | 0,06 | 5, 1 | 0,02 |
| 2 , 4 | 0,06 | 5, 2 | 0,03 |
| 2 , 5 | 0,03 | 5, 3 | 0,02 |
| 3 , 1 | 0,04 | 5, 4 | 0,02 |
| 3 , 1 | 0,06 | 5, 5 | 0,01 |
| 3 ,3 | 0,04 |  |  |

1. Find the sampling distribution of the sample mean .
2. Construct a probability histogram for the sampling distribution .
3. What is the probability that is 4,5 or larger?
4. Would you expect to observe a value of . Equal to 4.5 or larger? Explain.
5. (5). Refer to exercise 1. Assume that a random sample of n = 2 measurements is randomly selected from the population.
6. List the different values that the sample median m may assume and find the probability of each. Then give the sampling distribution of the sample median.
7. Construct a probability histogram for the sampling distribution of the sample median and compare it with the probability histogram for the sample mean (Exercise 1, question b).
   1. **Properties of Sampling Distributions: Unbiasedness and Minimum Variance**
8. (9). Consider the following probability distribution:

|  |  |  |  |
| --- | --- | --- | --- |
| x | 2 | 4 | 9 |
| p(x) |  |  |  |

1. Calculate μ for this distribution.
2. Find the sampling distribution of the sample mean for a random sample of n = 3 measurements from this distribution, and show that is unbiased estimator of μ.
3. Find the sampling distribution of the sample median m for a random sample of n = 3 measurements from this distribution, and show that the median is a biased estimator of μ.
4. If you wanted to use a sample of 3 measurements from this population to estimate μ, which estimator would you use? Why?
5. (10). Consider the following probability distribution:

|  |  |  |  |
| --- | --- | --- | --- |
| x | 0 | 1 | 2 |
| p(x) |  |  |  |

1. Find μ.
2. Find the sampling distribution of the sample mean for a random sample of n = 3 observations.
3. Find the sampling distribution of the sample median of a sample of n = 3 observations from this population.
4. Refer to parts b and c, and show that both the mean and the median are unbiased estimators of μ for this population.
5. Find the variances of the sampling distributions of the sample mean and the sample median.
6. Which estimator would you use to estimate μ. Why?

**5.3. The sampling distribution of the sample mean and the Central limit theorem**

1. (33, RHISH). ***Phishing attacks to e-mail accounts***. *Phishing* describes an attempt to extract personal/financial information from unsuspecting people through fraudulent e-mail. Data from an actual phishing attack against an organization were presented in *Chance* (Summer 2007). The interarrival times, i.e., the time differences (in seconds), for 267 fraud box e-mail notifications, were recorded and are saved in the accompanying file. For this exercise, consider these interarrival times to represent the population of interest.

a) Construct a histogram for the interarrival times. Describe the shape of the population of interarrival times.

b) Find the mean and standard deviation of the population of interarrival times.

c) Now consider a random sample *n =* 40 interarrival times selected from the population. Describe the shape of the sampling distribution of , the sample mean. Theoretically, what are and ?

d) Find *P* ().

e) Use a random number generator to select a random sample of *n* = 40 interarrival times from the population, and calculate the value of .

f) Refer to part **e**. Obtain the values of computed by the students and combine them into a single data set. Form a histogram for these values of . Is the shape approximately normal?

g) Refer to part **f**. Find the mean and standard deviation of the -values. Do these values approximate and respectively?

* 1. **Confidence Interval for a Population Mean: Normal (z) Statistic**

1. (16, BLKFRI). ***Shopping on Black Fridays***. The day after Thanksgiving – called Black Friday – is one of the largest shopping days in the USA. Winthrop University researchers conducted interviews with a sample of 38 women shopping on Black Friday to gauge their shopping habits and reported the results in the International Journal of Retail and Distribution Management (Vol. 39, 2011). One question was “How many hours do you usually spend shopping on Black Friday?” Data for the 38 shoppers are listed in the accompanying table.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 6 | 6 | 4 | 4 | 3 | 16 | 4 | 4 | 5 | 6 | 6 | 5 | 5 | 4 |
| 6 | 5 | 6 | 4 | 5 | 4 | 4 | 4 | 7 | 12 | 5 | 8 | 6 | 10 |
| 5 | 8 | 8 | 3 | 3 | 8 | 5 | 6 | 10 | 11 |  |  |  |  |

1. Describe he population of interest to the researchers.
2. What is the quantitative variable of interest to the researchers?
3. Use the information in the table to estimate the population mean number of hours spent shopping on Black Friday with a 95% confidence interval.
4. Give a practical interpretation of the interval.
5. A retail store advertises that the true mean number of hours spent shopping on Black Friday is 5.5 hours. Can the store be sued for false advertisement? Explain.
6. (20). ***Facial structure of CEOs.*** In Psychological Science (Vol. 22, 2011), researchers reported that a chief executive officer’s facial structure can be used to predict a firm’s financial performance. The study involved measuring the facial width-to-height ratio (WHR) for each in the sample of 55 CEOs at publicly traded Fortune 500 firms. These WHR values (determined by a computer analyzing a photo of the CEO’s face) had a mean of = 1,96 and a standard deviation of s = 0,15.
7. Find and interpret a 95% confidence interval for μ, the mean facial WHR for all CEOs at publicly traded Fortune 500 firms.
8. The researchers found that CEOs with wider faces (relative to height) tended to be associated with firms that had greater financial performance. They based their inference on an equation that uses facial WHR to predict financial performance. Suppose an analyst wants to predict the financial performance of a Fortune 500 firm based on the value of the true mean facial WHR of CEOs. The analyst wants to use the value of μ = 2,2. Do you recommend he use this value?
   1. **Confidence interval for a Population Mean: Student’s t-Statistic**
9. (31). ***Assessing the bending strength of a wooden roof.*** The white wood material used for the roof of an ancient Japanese temple is imported from Northern Europe. The wooden roof must withstand as much as 100 centimeters of snow in the winter. Architects at Tohoku University (Japan) conducted a study to estimate the mean bending strength of the white wood roof (Journal of the International Association for Shell and Spatial Structures, Aug. 2004). A sample of 25 pieces of the imported wood was tested and yielded the following statistics on breaking strength (MPa): = 74,5, s = 10,9. Estimate the true mean breaking strength of the white wood with a 90% confidence interval. Interpret the result.
10. (38, OVERBK). ***Overbooking policies for major airlines***. Airlines overbook flights in order to reduce the odds of flying with unused seats. An article in Transportation Research (Vol. 38, 2002) investigated the optimal overbooking policies for major airlines. One of the variables measured for each airline was the compensation (in dollars) per bumped passenger required to maximize future revenue. Consider the threshold levels of compensation for a random sample of 10 major airlines shown in the table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 825 | 850 | 1210 | 1370 | 1415 |
| 1500 | 1560 | 1625 | 2155 | 2220 |

Estimate the true mean threshold compensation level for all major worldwide airlines using a 90% confidence interval. Interpret the result practically.

**6.4 Large-sample confidence interval for a population proportion**

1. A random sample of 50 consumers taste-tested a new snack food. Their responses were coded (=: do not like; 1: like; 2: indifferent) and recorded as follows:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 0 | 0 | 1 | 2 | 0 | 1 | 1 | 0 | 0 |
| 0 | 1 | 0 | 2 | 0 | 2 | 2 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 |
| 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 0 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |

a) Use an 80% confidence interval to estimate the proportion of consumers who like the snack food.

b) Provide a statistical interpretation for the confidence interval you constructed in part **a**.

* 1. **Determining the Sample Size**

1. (72). ***Shopping on Black Fridays***. Refer to the International Journal of Retail and Distribution Management (Vol. 39, 2011) survey of Black Friday shoppers. One question was “How many hours do you usually spend shopping on Black Friday?”
2. How many Black Friday shoppers should be included in a sample designed to estimate the average number of hours spent shopping on Black Friday if you want the estimate to deviate no more than 0.5 hour from the true mean (use a confidence level of 95% and suppose the sample standard deviation “s” is equal to 2.755)?
3. Devise a sampling plan for collecting the data that will likely result in a representative sample.

**6.7 Confidence interval for a population variance**

12. (104, HCOUGH) ***Is honey a cough remedy?*** Archives of pediatrics and adolescent medicine (Dec. 2007) did a study of honey as a remedy for coughing. The 105 children ill in the sample were randomly divided into groups. One group received a dosage of an over-the-counter cough medicine (DM); another group received a dosage of honey (H). The coughing improvement scores (as determined by the children’s parents) for the patients in the two groups are reproduced in the accompanying table. The pediatric researchers desire information on the variation in coughing improvement scores for each of the two groups.

a) Find a 90% confidence interval for the standard deviation in improvement scores for the honey dosage group.

b) Repeat part **a** for the DM dosage group.

c) Based on the results, parts **a** and **b**, what conclusions can the pediatric researchers draw about which group has the smaller variation in improvement scores?

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Honey dosage: | 12 | 11 | 15 | 11 | 10 | 13 | 10 | 4 | 15 | 16 | 9 | 14 | 10 | 6 | 10 | 8 | 11 | 12 | 12 | 8 |
| 9 | 11 | 15 | 10 | 15 | 9 | 13 | 8 | 12 | 10 | 8 | 9 | 5 | 12 | 12 |  |  |  |  |  |
| DM dosage: | 4 | 6 | 9 | 4 | 7 | 7 | 7 | 9 | 12 | 10 | 11 | 6 | 3 | 4 | 9 | 12 | 7 | 6 | 8 | 12 |
| 12 | 4 | 12 | 13 | 7 | 10 | 13 | 9 | 4 | 4 | 10 | 15 | 9 |  |  |  |  |  |  |  |