SUPPLEMENTARY EXERCISES, CHAPTER 5

INTRODUCTION TO INference: ESTIMATION AND PREDICTION

S5.1. Distortion product otoacoustic emission (DPOAE) is a method for testing human hearing by measuring emissions produced by the cochlea (a part of the ear) in response to an input signal. For one set of 20 subjects randomly-selected from adults with normal hearing, the signal-noise ratio of the response to a 5 kHz input signal had a mean of 21.11 and a variance of 61.03. The data showed no evidence of nonnormality or outliers.

a. What is the target population to which inference will be made?
b. Compute a 99% confidence interval for the true mean signal-noise ratio of this population.
c. Explain what 99% confidence means for the interval in b.
d. A new subject has her hearing tested with a 5 kHz input signal, and a signal-noise ratio of 5.19 is obtained. Construct an appropriate level 0.90 interval based on the given data to decide if her hearing is normal.
e. Obtain a range of signal-noise ratios that you are 99% certain contain 95% of all signal-noise ratios for adults with normal hearing. (NOTE: negative values of the signal-noise ratio are legitimate since the ratio is in logged units. A negative value merely means that the noise is larger than the signal.)

S5.2. It seems as if many, if not most, retail prices end in a 9 (e.g., $29.99). To see if this is true, students in a business course took a random sample of 100 retail prices, and found that 19 ended in a 9. Obtain a 99% approximate score confidence interval for the proportion \( p \) of all retail prices that end in a 9. Is your result consistent with what you found in exercise? Justify your conclusion.

S5.3. The Rockwell hardness of a random sample of 15 shearing pins taken from a large production lot has mean 49.12 and standard deviation 1.29. The data appear stationary and normal.

a. Obtain a point estimate for the mean Rockwell hardness of the production lot.
b. Obtain a 95% confidence interval for the mean Rockwell hardness of the production lot. Tell what "95% confidence" means here.
c. Obtain an interval you are 95% confident will contain the Rockwell hardness measurement of the next shearing pin sampled from the lot.
d. The quality supervisor is interested in an interval that she can be 90% confident will contain at least 99% of the Rockwell hardness measurements of all shearing pins in the lot. Construct such an interval.
e. Assuming you know the population standard deviation to be 1.3, how large a sample would be required to estimate the mean Rockwell hardness to within 0.1 with 95% confidence?

S5.4. Suppose we have a normal population with known variance \( \sigma^2 \), and that we want to estimate its mean \( \mu \) to within one-tenth of a standard deviation (i.e., 0.1\( \sigma \)) with confidence 0.90. What sample size is needed?

S5.5. Piston rings for an automobile engine are produced by a forging process. To monitor quality, one piston ring is selected each hour from the production line for evaluation. Measurements of the inside diameters of 15 consecutively-selected rings have a mean of 74.998 mm, and a standard deviation of 0.007 mm. The data appear stationary and normal.

a. Obtain a point estimate for the process mean inner diameter.
b. Obtain a 95% confidence interval for the mean. Tell what "95% confidence" means here.
c. Obtain an interval you are 95% confident will contain the inner diameter measurement of the next piston ring produced.
d. The quality supervisor is interested in an interval that she can be 90% confident will contain at least 99% of the population of all possible measurements. Construct such an interval.
S5.6. In an experiment to test the effectiveness of nicotine patches in curbing cigarette smoking, a group of 200 volunteers was randomly divided into two groups of 100 each. Subjects in the first group were given a six week program of counseling, while those in the second group were given a six week program that included use of nicotine patches in addition to the counseling. Subjects were followed for one year after the program ended. Of the nicotine patch group, 26 remained non-smoking for the entire year, while of the group who received counseling alone, 86 resumed smoking within the year.

a. Let \( p_1 \) denote the population proportion of smokers who remain non-smoking for at least one year after the counseling and nicotine patch treatment, and \( p_2 \) the population proportion of smokers who remain non-smoking for at least one year after the counseling-only treatment. Obtain a point estimate of \( p_1 - p_2 \).

b. Construct an approximate 99% score confidence interval for \( p_1 - p_2 \). Make sure the necessary assumptions are satisfied to ensure a good approximation.

c. Does the interval you constructed provide convincing evidence the nicotine patch/counseling treatment is more effective than counseling treatment alone? Support your answer.

S5.7. The proper operation of a nuclear power plant depends on the reactor feedwater system. If the concentrations of metals, such as iron or copper, are too high, reactor performance can be adversely affected. In order to monitor metal concentrations, measurements are taken each day. A recent set of copper concentration measurements, one taken on each of 15 successive days, had a mean of 0.2631 ppb and a standard deviation of 0.0129 ppb. It is desired to perform statistical inference for the process that produced these measurements.

a. If it is desired to estimate the process mean, what assumptions about the process should be checked?

b. Assume you have checked the assumptions in (a), and there is no evidence that the assumptions do not hold.

i. Obtain a point estimate for the process mean copper concentration.

ii. Obtain a 95% confidence interval for the mean. Tell what "95% confidence" means here.

iii. Obtain an interval you are 95% confident will contain tomorrow's measurement.

iv. Management is interested in an interval that it can be 95% confident will contain at least 99% of the population of all possible measurements. Construct such an interval.

S5.8. A survey is to be done to estimate the proportion of a population with a certain characteristic. How large a sample is needed to be 99% confident that the estimate will be within 0.02 of the true proportion?

S5.9. Infection with an apparently harmless, newly recognized virus called hepatitis G, seems to interfere with HIV, slowing its progress and prolonging survival of AIDS patients, according to an Associated Press article in the September 6, 2001 edition of the Boston Globe. A study by researchers at the Iowa City Veterans Affairs Medical Center and the University of Iowa looked at 362 HIV-infected patients treated between 1988 and 1999. The researchers found that 144 of these patients were also infected with hepatitis G. Of the 144 who were infected with hepatitis G, 41 died during four years of follow-up compared with 123 of the 218 who were not infected with hepatitis G.

Assume the 144 patients with both HIV and hepatitis G form a random sample from the population of all such individuals, and that the second sample of 218 patients form a random sample from the population of all patients with HIV but no hepatitis G.

(a) Obtain a point estimate of the difference between the proportion of the population of patients infected with both HIV and hepatitis G who survive four years and the proportion of the population of patients infected with HIV but not with hepatitis G who survive four years.

(b) Obtain a 95% approximate score confidence interval for the difference estimated in (a).

(c) Explain what "95% confidence" means here.

(d) Explain in layman's terms what the interval you computed in (a) means in terms of the population proportions involved.

S5.10. A computer manufacturer randomly samples computer mice from the production line and subjects them to wear tests. In one such test, the left mouse button is pressed repeatedly by a testing machine until it fails. Data on the number of press-release cycles until failure in 12 test mice has mean 6709.6 and standard deviation 117.2. The process is stationary, and the data show no evidence of nonnormality.

(a) Construct a level 0.99 confidence interval for the population mean cycles to failure.
(b) Construct an interval based on the 12 values in the data set which you are 99 percent confident will contain the number of cycles to failure of the next mouse tested. If the next mouse fails after 6208 cycles, what will you conclude based on the interval you calculated?

(c) Construct an interval you are 99% confident will contain at least 95% of all cycles to failure values in the population. Interpret this interval in terms of “99% confidence”.

S5.11. Consider a level $L$ confidence interval for the mean $\mu$ in the C+E model where $\sigma$ is known. If I want to double the precision of the interval (i.e., make it half as wide), how much must I increase the sample size? How about if I want to triple the precision? Give a general formula for how much I must increase the sample size $n$ to make the interval $k$ times as precise.

S5.12. Market researchers want to estimate the proportion of the population that will respond to an advertising flyer. To do so, they obtain a random sample of size 50 and find that 5 respond to the flyer.

(a) Obtain a point estimate of the proportion of the population who will respond to the flyer.
(b) Obtain a 95% approximate score confidence interval for the population proportion estimated in (a).
(c) Explain what “95% confidence” means here.
(d) Explain in layman's terms what the interval you computed in (b) means in terms of the population proportions involved.
(e) Suppose the market researchers want to estimate the population proportion to within ±0.05 with 95% confidence. For planning purposes, they are willing to assume that the true population proportion $p$ equals 0.10. How large a sample should they plan?

S5.13. A manufacturer of concrete beams analyzes the quality of the product by sampling randomly from production and measuring various characteristics. One of these is porosity of the concrete. Porosity measurements on 10 recently tested units give a mean porosity of 0.149 with standard deviation 0.017.

(a) Construct a level 0.90 confidence interval for the population mean porosity.
(b) Construct an interval based on the 10 values in the data set which you are 90 percent confident will contain the porosity of the next unit tested. If the next sample tested has porosity 0.201, what will you conclude based on the interval you calculated?
(c) Construct an interval you are 90% confident will contain at least 99% of the porosities of all units in the population. Interpret this interval in terms of “90% confidence”.
(d) For all these intervals, what is being assumed about the population distribution of porosity measurements?

S5.14. A government study claims that 7% of individual federal tax returns under-report taxes due by more than $500. In a recent audit, 15 of a random sample of 100 individual federal tax returns under-reported taxes due by more than $500. Is this convincing evidence that the 7% rate claimed by the government study is too low? Use a level 0.95 confidence interval to justify your conclusion.

S5.15. Should researchers rely on the accuracy of self-reported measurements? To investigate this question, a study was conducted. Fifteen randomly-selected people were asked to report their weights on a questionnaire, and then were weighed on an accurate scale. For each subject, the difference between the measured weight and the self-reported weight was computed. These differences were found to have a mean of 3.2 and a standard deviation of 3.6 pounds. The differences showed no evidence of outliers or nonnormality.

(a) Compute a level 0.90 confidence interval for the population mean difference.
(b) What does the interval you computed in (a) suggest about the mean difference between measured and reported weights? Justify your response.
(c) A new, randomly-selected individual is going to join a second phase of the study. Predict the difference between her actual and reported weight with 90% confidence.
(d) Based on the interval from (c), do you predict that the new individual will report a lower weight than her actual weight? Justify your response.
(e) Obtain an interval that with 90% confidence will contain at least 99% of all population differences between measured and reported weights.
S5.16. In order to study whether the proportion of left handers among musicians is different than the proportion among non-musicians, two random samples are taken: one of 25 musicians and one of 25 non-musicians. In the samples, 5 of the musicians and 2 of the non-musicians are left handed.

(a) Construct a level 0.99 approximate score confidence interval for the difference between the population proportion of left-handers among musicians and non-musicians.

(b) Does the interval allow you to conclude that there is a difference in the population proportions? Justify your answer.

S5.17. As part of a biological research project, researchers need to quantify the density of a certain type of malignant cell in blood. In order to assure the accuracy of measurement, two experienced researchers each make 10 separate counts of the number of such cells in the same blood sample. The counts of the first researcher have a mean of 141.1 and a standard deviation of 16.8, while those of the second researcher have a mean of 132.7 and a standard deviation of 15.9.

(a) Based on these results is it reasonable to assume that the means of all possible readings for the two researchers are the same? Compute two confidence intervals to help you decide: one assuming equal population variances, and the other assuming unequal population variances. Use a confidence level of 0.99 for both intervals.

(b) Compare the standard errors $\hat{\sigma}(\bar{Y}_1 - \bar{Y}_2)$ and $\hat{\sigma}(\bar{Y}_1 - \bar{Y}_2)$ from part (a). What do you find? Can you explain this phenomenon?

S5.18. As part of its quality assurance program, an audio speaker manufacturer tests the pull-apart strength of speaker cabinets. Tests done on a random sample of 11 cabinets taken from production show a mean pull-apart strength of 121.70 pounds with a standard deviation of 27.90 pounds.

(a) The quality supervisor is interested in obtaining an interval that he is 95% confident will contain the pull-apart strength of the next cabinet tested. Construct such an interval. Interpret what 95% confidence means for this interval.

(b) The quality division head is interested in estimating the mean pull-apart strength of all production speaker cabinets of this type. She asks for an interval she can be 99% confident will contain that mean. Construct such an interval. Interpret what 99% confidence means for this interval.

(c) The assistant vice president for quality is interested in an interval that he is 90% confident will contain the pull-apart strengths of at least 95% of all production cabinets. Construct such an interval. Interpret what 90% confidence means for this interval.

(d) What assumptions about the population distribution should be checked before constructing these intervals?

S5.19. A gubernatorial candidate’s pollsters conduct a poll to compare the candidate’s support among likely male and female voters. Of 200 randomly-sampled likely female voters, 127 voice support for the candidate, compared with 158 of 300 randomly-sampled likely male voters.

(a) Obtain a point estimate of the difference, $p_f - p_m$, where $p_f$ is the population proportion of likely female voters who support the candidate, and $p_m$ is the population proportion of likely male voters who support the candidate.

(b) Compute a level 0.99 approximate score interval for $p_f - p_m$.

(c) Based on the interval you computed, can you conclude that there is a difference between the population proportions of likely female and male voters who support the candidate? Justify your answer.

S5.20. Researchers are testing a new cholesterol-lowering drug. The total cholesterol level of each of a sample of patients is measured prior to the start of the treatment. Treatment is administered for a month and each patient’s lifestyle is monitored to make sure there are no substantial changes that might affect the observed cholesterol levels. After a month on the treatment, each patient’s cholesterol level is measured again. Here are the results:

<table>
<thead>
<tr>
<th>Cholesterol Level</th>
<th>Patient</th>
<th>Before Treatment</th>
<th>After Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>240</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>257</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>215</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>289</td>
<td>251</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>236</td>
<td>238</td>
</tr>
</tbody>
</table>
(a) How should the sample be chosen so that the results will be applicable to a population of interest?

(b) Assume that the results are applicable to a population of interest, and that \( \mu_b \) is the population mean cholesterol level without taking the drug, and \( \mu_a \) is the population mean cholesterol level after taking the drug for a month. Construct a level 0.90 confidence interval for \( \mu_a - \mu_b \).

(c) Based on the interval in b, can the researchers conclude that the drug lowers cholesterol? Justify your answer.

S5.21. A company has a machine that makes steel bolts. Due to variations in the manufacturing process the diameter of these bolts vary slightly. It is widely known that this variation follows a Normal Distribution. A random sample of 30 bolts is drawn from the (infinite) population. The Sample Mean is 2.1cm and the Sample Standard Deviation is 0.15cm. The population variance is unknown.

(a) Construct a 95% Confidence interval for the Population Mean.

(b) Ideally, the company would like the mean diameter to be 2.1cm. From your answer to part (a), is there any evidence to suggest that this is not the case?

(c) Construct a 99% prediction interval for the diameter of the next bolt to be produced.

(d) Construct an interval that with 90% confidence contains 95% of the bolts.

S5.22. A company which produces laundry detergent wants to find out what proportion of the population of Worcester uses their product. The company has designed a sampling plan to obtain a sample from the population. Before the main sample is taken, a pilot study is done, in which 20 households were asked which laundry detergent they use. 8 responded that they used this company's detergent. In this question, suppose that \( L = 0.95 \) and that the population of Worcester can be assumed to be infinite.

(a) Calculate \( \hat{p} \), the point estimate for \( p \) using the approximate score method.

(b) Once the main sample has been conducted, the company would like a confidence interval for \( p \) that has width 0.08. Use your result from part (a) to help you find \( n \), the sample size required such that the confidence interval for \( p \) has width 0.08. Remember, \( n \) must be a whole number.

S5.23. A company produces metal washers. Due to variations in the manufacturing process the diameter of these washers vary slightly. It is widely known that this variation follows a Normal Distribution. A random sample of 36 washers is drawn from the (infinite) population. The Sample Mean is 1.06cm and the Sample Standard Deviation is 0.15cm.

(a) Construct a 95% Confidence interval for the Population Mean.

(b) Ideally, the company would like the mean diameter to be 1cm. From your answer to part (a), is there any evidence to suggest that this is not the case?

(c) Construct a 99% prediction interval for the diameter of the next washer to be produced.

(d) Construct an interval that with 90% confidence contains 95% of the washers.

S5.24. It is commonly thought that high school students who attend single-gender schools perform better than those who attend mixed-gender high schools. To test this theory, a study was performed in which a sample of graduating seniors was taken from each type of school and the number of students that were accepted to a Top 50 College were recorded. In the single-gender schools, 52 out of the 100 students sampled were accepted to a Top 50 College, whilst in the mixed-gender schools, 67 out of the 150 sampled were accepted.

(a) Construct a 95% approximate score confidence interval for the difference in the proportion of students from single-gender high schools attending a Top 50 college and the proportion of students from mixed-gender high schools doing likewise.

(b) From part (a), is there any evidence, using this measure of performance, that students from single-gender high schools perform better than students from mixed-gender high schools.

S5.25. A population can be described by a normal distribution model with \( \sigma^2 = 16 \), but the mean is unknown.

(a) Suppose I want to construct a 95% confidence interval for \( \mu \) such that the TOTAL WIDTH of the interval is no bigger than 1. Find the necessary sample size to accomplish this.

(b) Now suppose that we know that \( \mu = 100 \). Find the proportion of the population that have values bigger between 84 and 116.