

Andrew Moore, 10/10/2021

MATH-471, Homework 4

1a.

We have a sample of data $\underline{x} = 6, 4, 8, 4, 4, 4, 2, 5, 5, 7$. The sample size is 10.

We are told that the data represent the nightly count of animals caught in Alex's trap. Based on the description, we will assume the data comes from a *Poisson* distribution. The probability mass function for the Poisson distribution is

$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ for } x = 0, 1, 2, \dots \text{ and } \lambda > 0.$$

1b.

For the Poisson distribution, both the mean and variance are controlled by the same parameter, λ . To estimate the mean and variance, we will maximize the likelihood function, defined as

$$L(\lambda|x_1, x_2, \dots, x_{10}) = \prod_{i=1}^{10} f(x_i|\lambda) = \prod_{i=1}^{10} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}.$$

Taking the natural log of L , we have

$$\begin{aligned} \ln L(\lambda|x_1, x_2, \dots, x_{10}) &= \sum_{i=1}^{10} \ln \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \\ &= \sum_{i=1}^{10} [-\lambda + x_i \ln \lambda - \ln(x_i!)] \\ &= -\sum_{i=1}^{10} \lambda + \sum_{i=1}^{10} x_i \ln \lambda - \ln(x_i!) \\ &= -10\lambda + \sum_{i=1}^{10} x_i \ln \lambda - \ln(x_i!). \end{aligned}$$

Differentiating $\ln L$ results in

$$\frac{\partial}{\partial \lambda} [\ln L(\lambda|x_1, x_2, \dots, x_{10})] = \frac{\partial}{\partial \lambda} [-10\lambda + \sum_{i=1}^{10} x_i \ln \lambda - \ln(x_i!)] = -10 + \frac{\sum_{i=1}^{10} x_i}{\lambda}.$$

Setting the equation to 0, and solving for $\hat{\lambda}$ we have

$$\begin{aligned}
0 &= -10 + \frac{\sum_{i=1}^{10} x_i}{\hat{\lambda}} \\
10 &= \frac{\sum_{i=1}^{10} x_i}{\hat{\lambda}} \\
10\hat{\lambda} &= \sum_{i=1}^{10} x_i \\
\hat{\lambda} &= \frac{\sum_{i=1}^{10} x_i}{10} \\
\hat{\lambda} &= \frac{49}{10} = 4.9.
\end{aligned}$$

Thus, our estimate for the parameter is $\hat{\lambda} = 4.9$, i.e., $\mu = 4.9$ and $\sigma^2 = 4.9$.

We can test whether this point is the global maximum by using the 2nd-derivative test. Taking the second derivative of $\ln L$ we have

$$\frac{\partial^2}{\partial^2 \lambda} [\ln(L(\lambda|x_1, x_2, \dots, x_{10}))] = \frac{\partial}{\partial \lambda} \left[-10 + \frac{\sum_{i=1}^{10} x_i}{\lambda} \right] = \frac{-\sum_{i=1}^{10} x_i}{\lambda^2}$$

This result is negative, and will be negative for all λ (because, by definition, $\lambda > 0$). This means that we have found the maximum.

5.1

The county government in a city that is dominated by a large state university is concerned that a small subset of its population has been overutilized in the selection of residents to serve on county court juries. The county decides to determine the mean number of times that an adult resident of the county has been selected for jury duty during the past 5 years. They will then compare the mean jury participation for full-time students to that of nonstudents.

- a. Identify the populations of interest to the county officials.
- b. How might you select a sample of voters to gather this information?

5.5

A company that manufactures coffee for use in commercial machines monitors the caffeine content in its coffee. The company selects 50 samples of coffee every hour from its production line and determines the caffeine content. From historical data, the caffeine content (in milligrams, mg) is known to have a normal distribution with $\hat{\sigma} = 7.1$ mg. During a 1-hour time period, the 50 samples yielded a mean caffeine content of $\bar{y} = 110$ mg.

- a. Identify the population about which inferences can be made from the sample data.
- b. Calculate a 95% confidence interval for the mean caffeine content μ of the coffee produced during the hour in which the 50 samples were selected.
- c. Explain to the CEO of the company in nonstatistical language the interpretation of the constructed confidence interval.

5.6

Refer to Exercise 5.5. The engineer in charge of the coffee manufacturing process examines the confidence intervals for the mean caffeine content calculated over the past several weeks and is concerned that the intervals are too wide to be of any practical use. That is, they are not providing a very precise estimate of μ .

- a. What would happen to the width of the confidence intervals if the level of confidence of each interval is increased from 95% to 99%?
- b. What would happen to the width of the confidence intervals if the number of samples per hour was increased from 50 to 100?

5.8

As part of the recruitment of new businesses, the city's economic development department wants to estimate the gross profit margin of small businesses (under \$1 million in sales) currently residing in the city. A random sample of the previous years annual reports of 15 small businesses shows the mean net profit margin to be 7.2% (of sales) with a standard deviation of 12.5%.

- a. Construct a 99% confidence interval for the mean gross profit margin of μ of all small businesses in the city.
- b. The city manager reads the report and states that the confidence interval for μ constructed in part (a) is not valid because the data are obviously not normally distributed and thus the sample size is too small. Based on just knowing the mean and standard deviation of the sample of 15 businesses, do you think the city manager is valid in his conclusion about the data? Explain your answer.

5.10

The susceptibility of the root stocks of a variety of orange tree to a specific larva is investigated by a group of researchers. Forty orange trees are exposed to the larva and then examined by the researchers 6 months after exposure. The number of larvae per gram is recorded on each root stock. The mean and standard deviation of the logarithm of the counts are recorded to be 9.02 and 1.12, respectively.

- a. Use the sample information to construct a 90% confidence interval on the mean of the logarithm of the larvae counts.
- b. Identify the population for which this confidence interval could be used to assess the susceptibility of the orange trees to the larva.

5.14

The housing department in a large city monitors the rent for rent-controlled apartments in the city. The mayor wants an estimate of the average rent. The housing department must determine the number of apartments to include in a survey in order to be able to estimate the average rent to within \$100 using a 95% confidence interval. From past surveys, the monthly charge for rent-controlled apartments ranged from \$1,000 to \$3,500. How many renters must be included in the survey to meet the requirements?

5.15

Refer to Exercise 5.14. Suppose the mayor's staff reviews the proposed survey and decides that in order for the survey to be taken seriously the requirements need to be increased.

- a. If the level of confidence is increased to 99% with the average rent estimated within \$50, how many apartments need to be included in the survey?
- b. Suppose the budget for the survey will not support increasing the level of confidence to 99%. Provide an explanation to the mayor, who has never taken a statistics course, of the impact on the accuracy of the estimate of the average rent of not raising the level of confidence from 95% to 99%.