# Biostatistics [SBE304] (Fall 2019)
# Tutorial 5

Statistical Intervals Estimations

Lecturer: Prof. Ayman M. Eldeib TA: Asem Alaa

Tuesday 29th October, 2019

## 1 Tutorial facts

The problems in this tutorial comprises:

(A) Programming Works:

- **1** programming in-class demos.
- **2** programming homework.
- **0** self-practicing programming works.

(B) Problem Set:

- **13** problems to be solved in-class.
- **6** problems homework.
- **0** self-practicing problems.

Join this GitHub assignment page to create a repository for your submissions: `https://classroom.github.com/a/P_MPVz-4`

## 2 Statistical Intervals Estimations

### 2.1 Pre-class reading

1. Lecture notes of "Confidence Intervals" by Prof. Ayman M. Eldeib
2. From Chapter 8 of Montgomery's textbook, read (pp. 170-188)

### 2.2 Chapter overview: Confidence Intervals

#### 2.2.1 When measuring $n$ random variables $Y_i \sim i.i.d.$

**Hypotheses about the population mean** $E[Y_i]$
**Z-test** (when $n \geq 30$ or if normality with known variances could be assumed)

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

**CI for Normal Population:** A $100(1-\alpha)\%$ CI for the mean $\mu$ of a population when $\sigma$ is known is

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{1}$$

T -test (normality must be assured; for large n this is the same as the z-test). When $\bar{X}$ is the sample mean of a SRS of size n from a Normal($\mu, \sigma^2$) population then the RV

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has a probability distribution-t with n-1 degrees of freedom. Note: the density of $t_\nu$ is symmetric around 0. $t_\nu$ is more spread out than a normal, indeed the few dof the more spread. When dof is large ($\geq 40$), the t and normal curve are close. In addition we have that

$$P(|\frac{\bar{X} - \mu}{S/\sqrt{n}}| \leq t_{\alpha/2,n-1}) = 1 - \alpha$$

As a result, the $(1-\alpha)100\%$ CI for the population mean $\mu$ under the normal model is

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}$$

Note that here we make the assumption that the observations are realizations of a SRS from a Normal distribution with unknown mean and variance.
**Large Sample Test** for the population proportion (proportions are just means; only valid for $np_0 \geq 10$ and $n(1-p_0) \geq 10$. The $(1-\alpha)$ confidence interval for a population mean $\mu$ is

$$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}}$$

For a population proportion

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \qquad \hat{p} = \bar{X}$$

**Hypotheses about the population variance** $V[X_i]$
The $(1-\alpha)100\%$ CI for the variance $\sigma^2$ of a normal population has a lower limit:

$$(n-1)s^2/\chi^2_{\alpha/2,n-1}$$

and Upper limit:

$$(n-1)s^2/\chi^2_{1-\alpha/2,n-1}$$

A confidence interval for $\sigma$ has lower and upper limits that are the square roots of the corresponding limits in the interval for $\sigma^2$. An upper or a lower confidence bound results from replacing $\alpha/2$ with $\alpha$ in the corresponding limit of the CI.

## 2.3 Problem Set

### 2.3.1 Confidence Interval on the Mean of a Normal Distribution, Variance Known

**1. Problem**

For a normal population with known variance $\sigma^2$, answer the following questions:

a. What is the confidence level for the interval $\bar{x} - 2.14\sigma/\sqrt{n} \leq \mu \leq \bar{x} + 2.14\sigma/\sqrt{n}$

b. What is the confidence level for the interval $\bar{x} - 2.49\sigma/\sqrt{n} \leq \mu \leq \bar{x} + 2.49\sigma/\sqrt{n}$

c. What is the confidence level for the interval $\bar{x} - 1.85\sigma/\sqrt{n} \leq \mu \leq \bar{x} + 1.85\sigma/\sqrt{n}$

d. What is the confidence level for the interval $\mu \leq \bar{x} + 2.00\sigma/\sqrt{n}$

e. What is the confidence level for the interval $\bar{x} - 1.96\sigma/\sqrt{n} \leq \mu$

**1. Solution**

**2. Problem**

Consider the one-sided confidence interval expressions for a mean of a normal population.

a. What value of $z_\alpha$ would result in a 90% CI?

b. What value of $z_\alpha$ would result in a 95% CI?

c. What value of $z_\alpha$ would result in a 99% CI?

**2. Solution**

## 3. Problem

A random sample has been taken from a normal distribution and the following confidence intervals constructed using the same data: (38.02, 61.98) and (39.95, 60.05)

a. What is the value of the sample mean?

b. One of these intervals is a 95% CI and the other is a 90% CI. Which one is the 95% CI and why?

## 3. Solution

## 4. Problem

A confidence interval estimate is desired for the gain in a circuit on a semiconductor device. Assume that gain is normally distributed with standard deviation $\sigma = 20$.

a. How large must n be if the length of the 95% CI is to be 40?

b. How large must n be if the length of the 99% CI is to be 40?

## 4. Solution

## 5. Problem

Suppose that n = 100 random samples of water from a freshwater lake were taken and the calcium concentration (milligrams per liter) measured. A 95% CI on the mean calcium concentration is $0.49 \leq \mu \leq 0.82$.

a. Would a 99% CI calculated from the same sample data be longer or shorter?

b. Consider the following statement: There is a 95% chance that $\mu$ is between 0.49 and 0.82. Is this statement correct? Explain your answer.

c. Consider the following statement: If n = 100 random samples of water from the lake were taken and the 95% CI on $\mu$ computed, and this process were repeated 1000 times, 950 of the CIs would contain the true value of $\mu$. Is this statement correct? Explain your answer.

## 5. Solution

## 6. Problem

The yield of a chemical process is being studied. From previous experience, yield is known to be normally distributed and $\sigma = 3$. The past 5 days of plant operation have resulted in the following percent yields: 91.6, 88.75, 90.8, 89.95, and 91.3. Find a 95% two-sided confidence interval on the true mean yield.

## 6. Solution

## 7. Problem

A manufacturer produces piston rings for an automobile engine. It is known that ring diameter is normally distributed with $\sigma = 0.001$ millimeters. A random sample of 15 rings has a mean diameter of $\bar{x} = 74.036$ millimeters.

a. Construct a 99% two-sided confidence interval on the mean piston ring diameter.

b. Construct a 99% lower-confidence bound on the mean piston ring diameter. Compare the lower bound of this confidence interval with the one in part (a).

### 7. Solution

## 8. Problem

We assume that the probability distribution of blood pressure, X, is $N(\mu, \sigma^2)$ distribution. Suppose we know that $\sigma = 6$. To estimate $\mu$, we randomly selected 9 people and measured their blood pressure. The sample mean is $\bar{x} = 110$.

a. Write down the sampling distribution of the sample mean $\bar{X}$ and find its standard deviation.

b. Find the 80% confidence interval estimation for $\mu$.

### 8. Solution

## 9. Problem

A civil engineer is analyzing the compressive strength of concrete. Compressive strength is normally distributed with $\sigma^2 = 1000(\text{psi})^2$. A random sample of 12 specimens has a mean compressive strength of $\bar{x} = 3250psi$.

a. Construct a 95% two-sided confidence interval on mean compressive strength.

b. Construct a 99% two-sided confidence interval on mean compressive strength. Compare the width of this confidence interval with the width of the one found in part (a).

### 9. Solution

## 10. Problem

If the sample size n is doubled, by how much is the length of the CI on $\mu$ in Equation (1) reduced? What happens to the length of the interval if the sample size is increased by a factor of four?

### 10. Solution

## 11. Problem

By how much must the sample size n be increased if the length of the CI on $\mu$ in Equation (1) is to be halved?

### 11. Solution

## 12. Problem

Ishikawa et al. ["Evaluation of Adhesiveness of Acinetobacter sp. Tol 5 to Abiotic Surfaces," Journal of Bioscience and Bioengineering (Vol. 113(6), pp. 719–725)] studied the adhesion of various biofilms to solid surfaces for possible use in environmental technologies. Adhesion assay is conducted by measuring absorbance at $A_{590}$. Suppose that for the bacterial strain Acinetobacter, five measurements gave readings of 2.69, 5.76, 2.67, 1.62, and 4.12 dyne-cm$^2$. Assume that the standard deviation is known to be 0.66 dyne-cm$^2$.

a. Find a 95% confidence interval for the mean adhesion.

b. If the scientists want the confidence interval to be no wider than 0.55 dyne-cm$^2$, how many observations should they take?

### 12. Solution

## 13. Problem

An article in the Journal of Agricultural Science ["The Use of Residual Maximum Likelihood to Model Grain Quality Characteristics of Wheat with Variety, Climatic and Nitrogen Fertilizer Effects" (1997, Vol. 128, pp. 135–142)] investigated means of wheat grain crude protein content (CP) and Hagberg falling number (HFN) surveyed in the United Kingdom. The analysis used a variety of nitrogen fertilizer applications (kg N/ha), temperature ($^\circ$ C), and total monthly rainfall (mm). The following data below describe temperatures for wheat grown at Harper Adams Agricultural College between 1982 and 1993. The temperatures measured in June were obtained as follows: 15.2 14.2 14.0 12.2 14.4 12.5 14.3 14.2 13.5 11.8 15.2

Assume that the standard deviation is known to be $\sigma = 0.5$.

a. Construct a 99% two-sided confidence interval on the mean temperature.

b. Construct a 95% lower-confidence bound on the mean temperature.

c. Suppose that you wanted to be 95% confident that the error in estimating the mean temperature is less than 2 degrees Celsius. What sample size should be used?

d. Suppose that you wanted the total width of the two-sided confidence interval on mean temperature to be 1.5 degrees Celsius at 95% confidence. What sample size should be used?

**13. Solution**

### 2.3.2 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

**14. Problem**

An article in Obesity Research ["Impaired Pressure Natriuresis in Obese Youths" (2003, Vol. 11, pp. 745–751)] described a study in which all meals were provided for 14 lean boys for three days followed by one stress test (with a video-game task). The average systolic blood pressure (SBP) during the test was 118.3 mm HG with a standard deviation of 9.9 mm HG. Construct a 99% one-sided upper confidence interval for mean SBP.

**14. Solution**

**15. Problem**

A research engineer for a tire manufacturer is investigating tire life for a new rubber compound and has built 16 tires and tested them to end-of-life in a road test. The sample mean and standard deviation are 60,139.7 and 3645.94 kilometers. Find a 95% confidence interval on mean tire life

**15. Solution**

**16. Problem**

Determine the t-percentile that is required to construct each of the following one-sided confidence intervals:

a. Confidence level = 95%, degrees of freedom = 14

b. Confidence level = 99%, degrees of freedom = 19

c. Confidence level = 99.9%, degrees of freedom = 24

## 16. Solution

## 17. Problem

A healthcare provider monitors the number of CAT scans performed each month in each of its clinics. The most recent year of data for a particular clinic follows (the reported variable is the number of CAT scans each month expressed as the number of CAT scans per thousand members of the health plan): 2.31, 2.09, 2.36, 1.95, 1.98, 2.25, 2.16, 2.07, 1.88, 1.94, 1.97, 2.02

a. a. Find a 95% two-sided CI on the mean number of CAT scans performed each month at this clinic.

b. Historically, the mean number of scans performed by all clinics in the system has been 1.95. Is there any evidence that this particular clinic performs more CAT scans on average than the overall system average?

## 17. Solution

## 18. Problem

An article in Medicine and Science in Sports and Exercise ["Maximal Leg-Strength Training Improves Cycling Economy in Previously Untrained Men" (2005, Vol. 37, pp. 131–136)] studied cycling performance before and after 8 weeks of leg-strength training. Seven previously untrained males performed leg-strength training 3 days per week for 8 weeks (with four sets of five replications at 85% of one repetition maximum). Peak power during incremental cycling increased to a mean of 315 watts with a standard deviation of 16 watts. Construct a 95% confidence interval for the mean peak power after training

## 18. Solution

## 19. Problem

The brightness of a television picture tube can be evaluated by measuring the amount of current required to achieve a particular brightness level. A sample of 10 tubes results in $\bar{x} = 317.2$ and $s = 15.7$. Find (in microamps) a 99% confidence interval on mean current required. State any necessary assumptions about the underlying distribution of the data.

## 19. Solution

# 3 R Programming Language

## 3.1 R Programming Problems

**1. Problem**

A class demo in `week5/CIs.R` in the demos repository [1] to simulate the Confidence Intervals.

**2. Problem**

Using the `birthwt` dataset from `MASS` library (for description of this data run `?birthwt` in the R-console; make sure to load the library first by `library(MASS)` and `data = birthwt` to load the data in a new variable):

a. find the point estimate and the 85% confidence interval estimate for the **population proportion** of babies with low birthweight and the population proportion of mothers who have hypertension history (guide: read section 8.4 "Large-Sample Confidence Interval for a Population Proportion" of Montgomery's book, pp. 185-186).

b. find the point estimate and the 90% confidence interval estimate for the population mean of the number of physician visits during the first trimester.

c. Suppose that we want to estimate the population mean of birthweight. The acceptable margin error at 0.9 confidence level is 0.5 pounds (CI spans 1 pounds around the point estimate). If the range of birthweight is from 2 pounds to 11 pounds (discard all values above 11 and below 2 in a new list), what is the required sample size?

**3. Problem**

Using the `BodyTemperature.txt` data set (inside `R_problems` folder in the assignment repository), find the point estimate and the 80% confidence interval estimate for the population means of heart rate and normal body temperature.

[1] https://github.com/sbme-tutorials/biostatistics-sbe304.git