



**Professional Masters in  
Applied Statistics and Data Science  
(PM-ASDS)**

**An Assignment on  
Statistical Inference**

**Submitted to:**

**Dr. Tapati Basak**

Associate Professor

Department of Statistics, JU

**Submitted by:**

Mohammad Saiduzzaman Sayed

ID: 20215063

Batch: 5<sup>th</sup>

Sec: A

**Course Name: Statistical Inference**

**Course Code: PM-ASDS03**

## **Statistical Inference**

### **Dataset:**

- i) Data1: PlantGrowth
- ii) Data2: Puromycin

### **Solution of problem: 1**

#### **Computer Codes:**

```
# Import built-in dataset
Data1<-(PlantGrowth)
Data1

# Selecting First Sample
S1<-sample(Data1$weight, size = 40,replace=TRUE)
S1

# Point Estimate for S1
mean(S1)

# Selecting Second Sample
S2<-sample(Data1$group, size = 40,replace=TRUE)
S2

# Point Estimate for S2 ctrl category
table<-table(S2)
table
prop <- prop.table(table)
cat_ctrl = prop["ctrl"]
cat_ctrl

# Required packages to calculate Confidence Interval
install.packages("Rmisc")
library(Rmisc)
#For S1 and 90% confidence levels
CI(S1, ci=0.9)
```

```

#For S1 and 99% confidence levels
  CI(S1, ci=0.99)

# To get x argument
  n=length(Data1[,1])
  x= cat_ctrl*n
#For S2 and 90% confidence levels
  prop.test(x=x,n=length(Data1$group),conf.level= 0.9)
#For S2 and 99% confidence levels
  prop.test(x=x,n=length(Data1$group),conf.level = 0.99)

```

### **Result:**

#### **S1:**

5.17, 5.87, 4.50, 3.59, 5.54, 6.15, 4.17, 5.29, 5.80, 6.15, 5.50, 5.18, 4.50, 4.81, 6.11, 5.37, 4.81, 6.03, 5.18, 5.14, 5.87, 5.17, 5.14, 5.29, 5.54, 6.11, 6.15, 5.14, 6.11, 4.41, 5.37, 4.89, 5.50, 5.14, 4.17, 6.15, 3.83, 5.33, 5.12, 5.54

#### **S2:**

trt2, trt1, trt1, ctrl, trt1, trt2, ctrl, trt1, trt2, ctrl, trt1, ctrl, trt2, trt2, trt1, ctrl, trt1, trt2, ctrl, ctrl, ctrl, ctrl, trt2, trt1, ctrl, trt1, trt1, ctrl, trt1, ctrl, ctrl, trt2, trt2, trt1, trt2, trt2, ctrl, trt1, trt1, trt1

Levels: ctrl, trt1, trt2

### **Sample estimate:**

mean(S1): 5.27075

proportion(S2): ctrl = 0.35

#### **CI:**

For S1 and 90% confidence levels,

Upper: 5.447034

Mean: 5.270750

Lower: 5.094466

For S1 and 99% confidence levels,

Upper: 5.554071

Mean: 5.270750

Lower: 4.987429

For S2 and 90% confidence levels,

Upper: 0.5167364

Proportion: 0.35

Lower: 0.2108369

For S2 and 99% confidence levels,

Upper: 0.5973104

Proportion: 0.35

Lower: 0.1606450

### **Interpretations:**

- (i) In our "PlantGrowth" dataset, there are 30 rows which are observations, and 2 Columns, which are variables. Variables are "weight" and "group". The "weight" variable is quantitative, and the "group" is a categorical variable.

Then, we select two samples (S1, S2) of sizes 40 for each of the variables with replacement because the sample size is greater than the population observations. For the categorical variable, we select the "ctrl" category.

We assume that the samples are normally distributed because of sample size is large enough. In a normal distribution, the mean is considered a more efficient point estimator than the median for the quantitative variables. So, Mean is the appropriate point estimator for S1.

For categorical variable S2, the proportion of successes in our population. This may consist of estimating a single parameter. So, Proportion is the appropriate point estimator for S2.

- (ii) For S1 and 90% confidence levels, we say that we are 90% confident that the population mean is between 5.447034 and 5.094466 and 99% confidence levels, we say that we are 99% confident that the population mean is between 5.554071 and 4.987429.

For S2 and 90% confidence levels, we say that we are 90% confident that the population proportion is between 0.21 and 0.52 and 99% confidence levels, we say that we are 99% confident that the population proportion is between 0.16 and 0.597.

### **Solution of problem: 2**

#### **Computer Codes:**

```
# Import built-in dataset
Data2<-(Puromycin)
Data2

# Selecting First Sample
S21<-sample(Data2$conc,size=10,replace=FALSE)
S21

# Selecting Second Sample
S22<-sample(Data2$rate,size=15,replace=FALSE)
S22

# Selecting Third Sample
S23<-sample(Data2$state,size=30,replace=TRUE)
S23

# Point Estimate for S21
mean(S21)
# Point Estimate for S22
mean(S22)
# Point Estimate for S23 treated category
table<-table(S23)
prop<-prop.table(table)
cat_treated<-prop["treated"]
cat_treated

# Required packages to calculate Confidence Interval
install.packages("Rmisc")
library(Rmisc)

#For S21 and 95% confidence levels
```

```

      CI(S21, ci=0.95)
# For S22 and 95% confidence levels
      CI(S22, ci=0.95)

# To get x argument
      n<-length(Data2$state)
      x<-cat_treated*n
      x
#For S23 and 95% confidence levels
      prop.test(x=x,n=length(Data2$state),conf.level = 0.95)

# Required packages to calculate Confidence Interval
      install.packages("BSDA")
      library(BSDA)

# When the population variances are known, for S21 and S22 at 95% CI
      z.test(S21,S22,alternative="two.sided",mu=0,sigma.x=sqrt(var(S21)),sigma.
      y=sqrt(var(S22)),conf.level=0.95)

# When the population variances are unknown but equal, for S21 and S22 at 95%
CI
      t.test(S21,S22,alternative = "two.sided", mu=0,var.equal = TRUE,conf.level =
      0.95)

# When the population variances are unknown but unequal, for S21 and S22 at 95%
CI
      t.test(S21,S22,alternative = "two.sided", mu=0,var.equal = FALSE,conf.level
      = 0.95)

```

## Results:

### S21:

0.11, 0.56, 0.56, 0.22, 1.10, 0.02, 1.10, 0.06, 0.11, 0.02

### S22:

201, 86, 144, 131, 139, 76, 200, 124, 47, 67, 107, 51, 158, 98, 123

**S23:**

Untreated, treated, untreated, treated, treated, treated, treated, treated, untreated, treated, treated, untreated, treated, untreated, treated, treated, treated, untreated, treated, untreated, untreated, treated, treated, untreated, treated

Levels: treated, untreated

**Sample estimate:**

mean(S21): 0.386

mean(S22): 124.7333

proportion(S23): treated = 0.6

**CI:**

For S21 and 95% confidence levels,

Upper: 0.69083122

Mean: 0.386

Lower: 0.08116878

For S22 and 95% confidence levels,

Upper: 153.98542

Mean: 124.73333

Lower: 95.48124

For S23 and 95% confidence levels,

Upper: 0.7886031

Proportion: 0.6

Lower: 0.3799026

When the population variances are known, for S21 and S22

95 percent confidence interval:

-151.07999, -97.61468

When the population variances are unknown but equal, for S21 and S22

95 percent confidence interval:

-159.15225, -89.54242

When the population variances are unknown but unequal, for S21 and S22

95 percent confidence interval:

-153.60031, -95.09435

### **Interpretation:**

- (i) In our "Puromycin" dataset, there are 23 rows which are observations, and 3 Columns, which are variables. Variables are "conc", "rate" and "state". The "conc" and "rate" variable is quantitative, and the "state" is a categorical variable.

Then, we select two samples (S21, S22) of sizes (10, 15) for conc and rate variables without replacement because the sample size is less than the population observations. For category variable rate, we select sample (S23) of size 30 with replacement because the sample size is greater than the population observations. For instance, we select the "treated" category.

We assume that the samples are normally distributed. In a normal distribution, the mean is considered a more efficient point estimator than the median for the quantitative variables. So, Mean is the appropriate point estimator for S21 and S22.

- (ii) For categorical variable S23, the proportion of successes in our population. This may consist of estimating a single parameter. So, Proportion is the appropriate point estimator for S23.



(iii) For S21 and 95% confidence levels, we say that we are 95% confident that the population mean is between 0.69083122 and 0.08116878.

For S22 and 95% confidence levels, we say that we are 95% confident that the population mean is between 153.98542 and 95.48124.

For S23 and 95% confidence levels, we say that we are 95% confident that the population proportion is between 0.79 and 0.38.

(iv) For the two independent samples S21 and S22, the 95% confidence intervals for two the differences between two population means,

When the population variances are known, we use the z-test because we know population variances. And we use the t-test when our variances are unknown but equal or not.

THE END