

t-test Tutorial

FMU Biology Department

t-test in R

This documents includes code for creating a data object in R, displaying summary statistics, creating an informative plot, and running a *t*-test.

Background

You have completed a habitat restoration project in a western South Carolina forest to expand the population of the federally threatened Bigfoot *Gigantopithecus humongous*. To determine the success of your habitat restoration project you surveyed Bigfoot populations before the restoration project and 20 years after the restoration project. The 20 year time-lag was needed to allow natural reproduction to respond to the habitat restoration. Your survey methods consist of camera traps deployed at 10 sites and cameras recorded sightings for thirty days. Catch per unit effort (CPUE) is calculated as number of Bigfoot observed per camera day at each site.

Hypothesis: If Bigfoot habitat restoration was successful, then the number of Bigfoot observed will increase after 20 years

Null Hypothesis: There is no difference in Bigfoot CPUE between sample years.

Data

The data frame being created below contain 20 observations (rows) and 2 variables (columns). Each row represents one individual observation. The first column is the CPUE of Bigfoot observed at each site per day and the second column is the year classification (A = before habitat restoration, B = 20 after habitat restoration).

Important points to note. The function is “data.frame” which creates a data frame object that is being assigned the name “dat2”. Everything after “data.frame” are the arguments being passed to the function. The arguments include the column names (cpue and group) and the data in each column. Quotes (“”) are used to encapsulate text. This is important so R knows that this column is a categorical variable. The categorical variable, group, will be used as the predictor of cpue.

Enter the following code in R:

```
dat2 = data.frame(cpue = c(8.9,11.9,11.3,13.2,10.2,9.8,6.4,12.9,12.9,17.1,
                          21.7,22.3,21.2,20.9,18.8,20.9,17.3,18.4,22.8,21.3),
                  group = c("A","A","A","A","A","A","A","A","A","A",
                           "B","B","B","B","B","B","B","B","B","B"))
```

The next line of code will print the data contained in the dat2 object

```
dat2
```

```
##      cpue group
## 1    8.9      A
## 2   11.9      A
## 3   11.3      A
## 4   13.2      A
## 5   10.2      A
## 6    9.8      A
## 7    6.4      A
## 8   12.9      A
## 9   12.9      A
## 10  17.1      A
## 11  21.7      B
## 12  22.3      B
## 13  21.2      B
## 14  20.9      B
## 15  18.8      B
## 16  20.9      B
## 17  17.3      B
## 18  18.4      B
## 19  22.8      B
## 20  21.3      B
```

Display summary data

Important things to note: “dat2\$cpue” and “dat2\$group” references your data frame name (dat2) and the column name (e.g., group)

```
#Average CPUE per group
tapply(X = dat2$cpue, INDEX = dat2$group, FUN = mean)
```

```
##      A      B
## 11.46 20.56
```

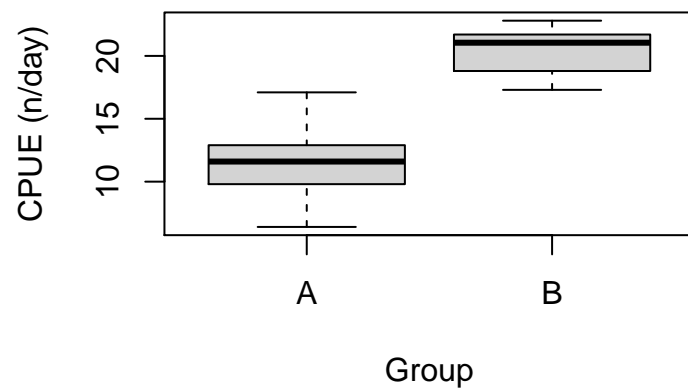
```
#Standard deviation of CPUE per group
tapply(X = dat2$cpue, INDEX = dat2$group, FUN = sd)
```

```
##      A      B
## 2.900268 1.792701
```

Plot data

The next code chunk will display your data in a boxplot with custom x- and y-axis labels.

```
boxplot(dat2$cpue~dat2$group,
        ylab = "CPUE (n/day)", xlab = "Group")
```



Run *t*-test

This code chunk performs a *t*-test and displays summary data.

```
res2 = t.test(dat2$cpue~dat2$group, var.equal=TRUE)
res2

##
## Two Sample t-test
##
## data: dat2$cpue by dat2$group
## t = -8.4399, df = 18, p-value = 1.134e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.365232 -6.834768
## sample estimates:
## mean in group A mean in group B
##          11.46          20.56
```

Full code block

```
#Create data frame
dat2 = data.frame(cpue = c(8.9,11.9,11.3,13.2,10.2,9.8,6.4,12.9,12.9,17.1,
                          21.7,22.3,21.2,20.9,18.8,20.9,17.3,18.4,22.8,21.3),
                  group = c("A","A","A","A","A","A","A","A","A","A",
                            "B","B","B","B","B","B","B","B","B","B"))

#Print data frame
dat2

#Summary statistics
#Average CPUE per group
tapply(X = dat2$cpue, INDEX = dat2$group, FUN = mean)

#Standard deviation of CPUE per group
tapply(X = dat2$cpue, INDEX = dat2$group, FUN = sd)

#Create Boxplot
boxplot(dat2$cpue~dat2$group,
        ylab = "CPUE (n/day)", xlab = "Group")

#t test
res2 = t.test(dat2$cpue~dat2$group, var.equal=TRUE)
#print results
res2
```