LabHypothesis3

Diego Castro

10/28/2021

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

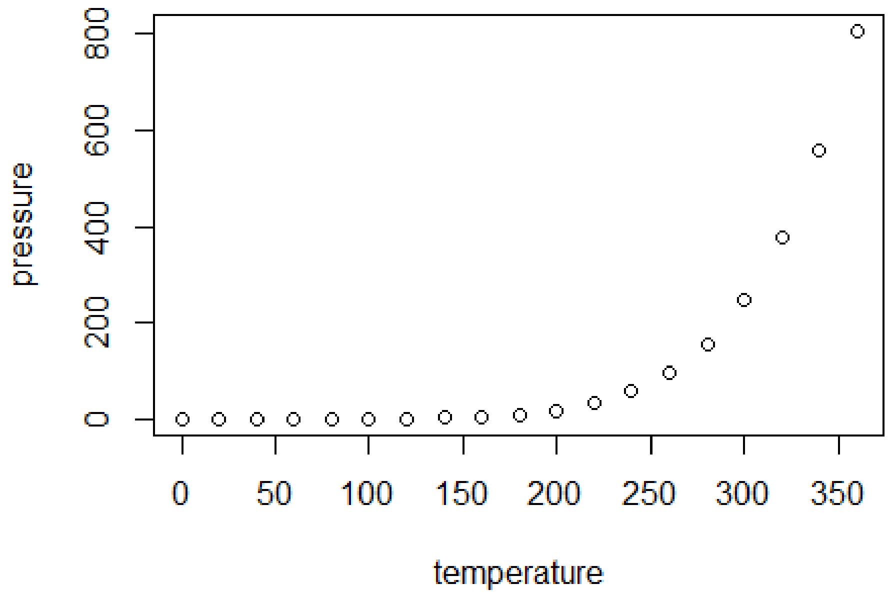
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

ECO-R002 - LAB 3 Hyp. Testing

```
#Example 1:
# BankX plans to launch a new financial product. A sample of 25 potential inv
estors, collected the following
#information on the P amount they wish to invest in the new product (normally
distributed, with variance 500):#i xi = 1000 and P#i(xi − ¯x)2 = 9600.

#Hypothesis Testing for a Population Mean with known variance


#  Null Hiphotesis : Is possible to state that the mean investmet on this pro
duct is 45

alpha <- .05
z.half.alfa <- qnorm(1-alpha/2)
z.half.alfa

## [1] 1.959964

c(-z.half.alfa, z.half.alfa)

## [1] -1.959964  1.959964
```

```r
xbar<- 1000/25   # sample mean
xbar

## [1] 40

mu0 <- 45         #hyphotesized value
sigma <- sqrt(500) # population Standar deviation
n <- 25
z = (xbar-mu0)/(sigma/sqrt(n))
z

## [1] -1.118034

# We can not reject the Null Hypothesis

#Alternatevely
pnorm(z, lower.tail = FALSE)  # upper tail

## [1] 0.8682238

pnorm(z, lower.tail = TRUE)   # lower tail

## [1] 0.1317762

pnorm(z)

## [1] 0.1317762

# For the two tail p-valued we choose the minimum of the values (lower or upp
er) and is the defualt in R

pval <- 2* pnorm(z)
pval  # two tail p value

## [1] 0.2635525

#Hypothesis Testing for a Population Mean with unknown variance

#How would you change your answer if you had doubts on the value of the varia
nce?

t.alpha <- .05   #  significance level

t.half.alpha <- qt(1-alpha/2, 25-1)
c(-t.half.alpha, t.half.alpha)     # Critical Values

## [1] -2.063899  2.063899

#Thus, the decision rule is that the null hypothesis should be rejected if t
<= -2.06 or Q >= 2.06

#Compute the test statistic
```

```r
xbar<- 1000/25   # sample mean
xbar

## [1] 40

mu0 <- 45             #hyphotesized value
s <- sqrt(500)  # sample Standar deviation
n <- 25
t = (xbar-mu0)/(s/sqrt(n))
t

## [1] -1.118034

#We are not able to reject the null hypothesis

#Alternatevely
pt(t, df=25-1, lower.tail = FALSE)

## [1] 0.8626903

pt(t, df=25-1, lower.tail = TRUE)

## [1] 0.1373097

pval <- 2* pt(t, df=25-1)
pval    # two tail p valued

## [1] 0.2746193

# Hypothesis Testing for a Population Variance

# Are there reasons to doubt the value of the variance?

q.alpha = .05 # significance level
q.half.alpha.up = qchisq(1-alpha/2, 25-1) # critical values
q.half.alpha.up

## [1] 39.36408

q.half.alpha.low = qchisq(alpha/2, 25-1) # critical values
q.half.alpha.low

## [1] 12.40115

#Thus, the decision rule is that the null hypothesis should be rejected if Q
<= 12.4 or Q >= 39.4.

sigma_sqr_0 = 500 # hypothesized value
s_sqr = sqrt (400) # sample standard deviation
n = 25 # sample size
```

```r
Q = ((n-1)*s_sqr)/sigma_sqr_0 # test statistic
Q

## [1] 0.96

#Example 2
#A particular type of cancer therapy has a 60% success rate. A group of resea
rchers developed a new type of
#treatment and its effectiveness is to be tested. In 61 cases, 47 were succes
sfully treated.
#Is there enough empirical evidence that allows us to conclude that the new t
reatment is better than the old one?

#The null hypothesis is that p = 0.60
#z is the appropiate test statistic

alpha <- .05
z.aplha <- qnorm(1-alpha)
z.aplha

## [1] 1.644854

#Thus, the decision rule is that the null hypothesis should be rejected if z
> 1.65.

p0 = 0.60
fn = 47/61 # sample proportion
n = 61 # sample size
z = (fn-p0)/sqrt((0.6*(1-0.6))/n) # test statistic
z

## [1] 2.718084

# Reject the hyphotesis that there are the same

#Example 3

#Hypothesis Testing for a Difference in Population Means - Independent Sample
s and Variance Unknown
#Is there a difference between average dividends of the stocks in Dow Jones a
nd the ones in Eurostoxx,
#knowing they have equal variances and normally distributed? The data of two
independent samples is the
#following

nX = 21
nY = 25
sX = 1.30
sY = 1.16
```

```r
Sp_sqr = ((nX - 1)*(sX^2) + (nY-1)*(sY^2))/(nX + nY - 2)
Sp_sqr

## [1] 1.502145

df <- nX + nY-2
df

## [1] 44

#The null hypothesis is that μ1 − μ2 = 0.

t.alpha <- .05   #  significance level

t.half.alpha <- qt(1-alpha/2, 44)
c(-t.half.alpha, t.half.alpha)     # Critical Values

## [1] -2.015368  2.015368

xbar = 3.27
ybar = 2.53
t = ((xbar-ybar)-0)/sqrt((Sp_sqr/nX)+(Sp_sqr/nY))
t

## [1] 2.039748

# We reject the null

pt(t, df=44, lower.tail=FALSE) # upper tail

## [1] 0.02370372

pt(t, df=44, lower.tail=TRUE) # lower tail

## [1] 0.9762963

pval = 2 * pt(t, df=44, lower.tail=FALSE)
pval

## [1] 0.04740744

#Example 4
library(data.table)

## Warning: package 'data.table' was built under R version 4.1.1

setwd ("C:/Users/usuario/OneDrive - University of East Anglia/PhD/First Semes
tre/Econometrics/Laboratories/Lab3")
list.files()

##  [1] "data_r.csv"
##  [2] "ECO-R002-lab03-hypothesis-testing-examples (1).pdf"
##  [3] "ECO-R002-lab03-hypothesis-testing-examples.pdf"
```

```
##  [4] "ECO-R002-lab03_SipleRegression.pdf"
##  [5] "Lab3.docx"
##  [6] "Lab3.pdf"
##  [7] "Lab3.Rmd"
##  [8] "Lab3LinearRegresion.R"
##  [9] "LabMatrix.R"
## [10] "sales-data.csv"
```

```
dt.stocks <- data.table(read.csv("data_r.csv"))
head(dt.stocks)
```

```
##    serial Year Month  DJComp   DJInd DJUtil  DJTran NASDAQ  SP500  SP100
## 1:      1 1990   Jan  959.54 2590.54 223.65 1045.87  415.8 329.08 307.88
## 2:      2 1990   Feb  986.07 2627.25 220.38 1129.09  425.8 331.89 312.48
## 3:      3 1990   Mar 1012.10 2707.21 214.66 1183.14  435.5 339.94 320.03
## 4:      4 1990   Apr  979.70 2656.76 203.09 1129.98  420.1 330.80 314.23
## 5:      5 1990   May 1040.16 2876.66 211.39 1171.53  459.0 361.23 342.66
## 6:      6 1990   Jun 1031.07 2880.69 210.01 1142.70  462.3 358.02 339.80
##    Treas3m IDJComp IDJInd IDJUtil IDJTran INASDAQ ISP500 ISP100 ITreas3m
## 1:    7.90      NA     NA      NA      NA      NA     NA     NA       NA
## 2:    8.00    2.76   1.42   -1.46    7.96    2.41   0.85   1.49     0.64
## 3:    8.17    2.64   3.04   -2.60    4.79    2.28   2.43   2.42     0.66
## 4:    8.04   -3.20  -1.86   -5.39   -4.49   -3.54  -2.69  -1.81     0.65
## 5:    8.01    6.17   8.28    4.09    3.68    9.26   9.20   9.05     0.64
## 6:    7.99   -0.87   0.14   -0.65   -2.46    0.72  -0.89  -0.83     0.64
```

```
dt.stocks <- setnames(dt.stocks, tolower(names(dt.stocks)))
head(dt.stocks)
```

```
##    serial year month  djcomp   djind djutil  djtran nasdaq  sp500  sp100
## 1:      1 1990   Jan  959.54 2590.54 223.65 1045.87  415.8 329.08 307.88
## 2:      2 1990   Feb  986.07 2627.25 220.38 1129.09  425.8 331.89 312.48
## 3:      3 1990   Mar 1012.10 2707.21 214.66 1183.14  435.5 339.94 320.03
## 4:      4 1990   Apr  979.70 2656.76 203.09 1129.98  420.1 330.80 314.23
## 5:      5 1990   May 1040.16 2876.66 211.39 1171.53  459.0 361.23 342.66
## 6:      6 1990   Jun 1031.07 2880.69 210.01 1142.70  462.3 358.02 339.80
##    treas3m idjcomp idjind idjutil idjtran inasdaq isp500 isp100 itreas3m
## 1:    7.90      NA     NA      NA      NA      NA     NA     NA       NA
## 2:    8.00    2.76   1.42   -1.46    7.96    2.41   0.85   1.49     0.64
## 3:    8.17    2.64   3.04   -2.60    4.79    2.28   2.43   2.42     0.66
## 4:    8.04   -3.20  -1.86   -5.39   -4.49   -3.54  -2.69  -1.81     0.65
## 5:    8.01    6.17   8.28    4.09    3.68    9.26   9.20   9.05     0.64
## 6:    7.99   -0.87   0.14   -0.65   -2.46    0.72  -0.89  -0.83     0.64
```

```
#Calculate the 95% confidence interval for the stocks' means.
summary(dt.stocks)
```

```
##      serial            year          month               djcomp
##  Min.   :  1.00   Min.   :1990   Length:296         Min.   : 858.1
##  1st Qu.: 74.75   1st Qu.:1996   Class :character   1st Qu.:1771.0
##  Median :148.50   Median :2002   Mode  :character   Median :2987.7
```

```
##  Mean   :148.50   Mean   :2002                             Mean   :2910.0
##  3rd Qu.:222.25   3rd Qu.:2008                             3rd Qu.:3832.7
##  Max.   :296.00   Max.   :2014                             Max.   :6040.9
##
##      djind           djutil          djtran          nasdaq
##  Min.   : 2442   Min.   :177.2   Min.   : 822.3   Min.   : 329.8
##  1st Qu.: 5518   1st Qu.:224.6   1st Qu.:2048.0   1st Qu.:1095.2
##  Median : 9923   Median :303.2   Median :2947.3   Median :1950.0
##  Mean   : 8896   Mean   :326.5   Mean   :3283.5   Mean   :1924.9
##  3rd Qu.:11155   3rd Qu.:412.3   3rd Qu.:4553.0   3rd Qu.:2511.2
##  Max.   :17068   Max.   :554.3   Max.   :8294.7   Max.   :4696.7
##
##      sp500            sp100          treas3m          idjcomp
##  Min.   : 304.0   Min.   :289.8   Min.   :0.010   Min.   :-14.1700
##  1st Qu.: 640.3   1st Qu.:449.8   1st Qu.:0.935   1st Qu.: -1.1950
##  Median :1107.0   Median :561.9   Median :3.340   Median :  1.0600
##  Mean   :1019.3   Mean   :561.6   Mean   :3.157   Mean   :  0.6979
##  3rd Qu.:1320.3   3rd Qu.:646.4   3rd Qu.:5.112   3rd Qu.:  3.1750
##  Max.   :2003.4   Max.   :929.9   Max.   :8.170   Max.   : 10.3200
##                                                   NA's   :1
##      idjind             idjutil             idjtran             inasdaq
##  Min.   :-15.1300   Min.   :-13.3900   Min.   :-21.9900   Min.   :-22.900
##  1st Qu.: -1.4950   1st Qu.: -2.1000   1st Qu.: -2.3500   1st Qu.: -2.460
##  Median :  1.0900   Median :  0.9500   Median :  1.6100   Median :  1.780
##  Mean   :  0.7178   Mean   :  0.3937   Mean   :  0.8702   Mean   :  1.036
##  3rd Qu.:  3.4000   3rd Qu.:  2.9050   3rd Qu.:  4.5700   3rd Qu.:  4.645
##  Max.   : 10.6000   Max.   : 11.7600   Max.   : 17.4500   Max.   : 21.980
##  NA's   :1          NA's   :1          NA's   :1          NA's   :1
##      isp500             isp100             itreas3m
##  Min.   :-16.9400   Min.   :-47.4400   Min.   :0.0000
##  1st Qu.: -1.7750   1st Qu.: -1.7600   1st Qu.:0.0800
##  Median :  1.1100   Median :  1.1300   Median :0.2700
##  Mean   :  0.7046   Mean   :  0.5106   Mean   :0.2558
##  3rd Qu.:  3.4500   3rd Qu.:  3.2400   3rd Qu.:0.4200
##  Max.   : 11.1600   Max.   : 10.7900   Max.   :0.6600
##  NA's   :1          NA's   :1          NA's   :1
```

```
dt.stocks[, t.test(idjcomp, conf.level = 0.95)]
```

```
##
##  One Sample t-test
##
## data:  idjcomp
## t = 2.982, df = 294, p-value = 0.003103
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.2372876 1.1584412
## sample estimates:
## mean of x
## 0.6978644
```

```
dt.stocks[, t.test(djind, conf.level = 0.95)]

##
##  One Sample t-test
##
## data:  djind
## t = 40.733, df = 295, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  8465.754 9325.342
## sample estimates:
## mean of x
##  8895.548

dt.stocks[, t.test(djutil, conf.level = 0.95)]

##
##  One Sample t-test
##
## data:  djutil
## t = 52.414, df = 295, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  314.2069 338.7228
## sample estimates:
## mean of x
##  326.4648

dt.stocks[, t.test(djtran, conf.level = 0.95)]

##
##  One Sample t-test
##
## data:  djtran
## t = 35.157, df = 295, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3099.652 3467.257
## sample estimates:
## mean of x
##  3283.454

dt.stocks[, t.test(nasdaq, conf.level = 0.95)]

##
##  One Sample t-test
##
## data:  nasdaq
## t = 32.739, df = 295, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
```

```
##  1809.169 2040.590
## sample estimates:
## mean of x
##  1924.879
```

```
dt.stocks[, t.test(sp500, conf.level = 0.95)]
```

```
##
##  One Sample t-test
##
## data:  sp500
## t = 42.529, df = 295, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   972.1652 1066.5052
## sample estimates:
## mean of x
##  1019.335
```

```
dt.stocks[, t.test(sp100, conf.level = 0.95)]
```

```
##
##  One Sample t-test
##
## data:  sp100
## t = 69.823, df = 295, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  545.7608 577.4188
## sample estimates:
## mean of x
##  561.5898
```

```
dt.stocks[, t.test(treas3m, conf.level = 0.95)]
```

```
##
##  One Sample t-test
##
## data:  treas3m
## t = 23.44, df = 295, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.891990 3.422131
## sample estimates:
## mean of x
##  3.157061
```

```
dt.stocks[, t.test(idjcomp, conf.level = 0.95)]
```

```
##
##  One Sample t-test
##
```

```
## data:  idjcomp
## t = 2.982, df = 294, p-value = 0.003103
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.2372876 1.1584412
## sample estimates:
## mean of x
## 0.6978644
```

```
dt.stocks[, t.test(idjind, conf.level = 0.95)]
```

```
##
##  One Sample t-test
##
## data:  idjind
## t = 2.9669, df = 294, p-value = 0.003255
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.2416394 1.1938860
## sample estimates:
## mean of x
## 0.7177627
```

```
dt.stocks[, t.test(idjutil, conf.level = 0.95)]
```

```
##
##  One Sample t-test
##
## data:  idjutil
## t = 1.5767, df = 294, p-value = 0.1159
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.09770209  0.88502413
## sample estimates:
## mean of x
##  0.393661
```

```
dt.stocks[, t.test(idjtran, conf.level = 0.95)]
```

```
##
##  One Sample t-test
##
## data:  idjtran
## t = 2.5428, df = 294, p-value = 0.01151
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1966751 1.5437317
## sample estimates:
## mean of x
## 0.8702034
```

```r
dt.stocks[, t.test(inasdaq, conf.level = 0.95)]
```

```
## 
##  One Sample t-test
## 
## data:  inasdaq
## t = 2.7047, df = 294, p-value = 0.007234
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.2822519 1.7903583
## sample estimates:
## mean of x 
##  1.036305
```

```r
dt.stocks[, t.test(isp500, conf.level = 0.95)]
```

```
## 
##  One Sample t-test
## 
## data:  isp500
## t = 2.8631, df = 294, p-value = 0.004497
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.220257 1.188896
## sample estimates:
## mean of x 
## 0.7045763
```

```r
dt.stocks[, t.test(itreas3m, conf.level = 0.95)]
```

```
## 
##  One Sample t-test
## 
## data:  itreas3m
## t = 23.493, df = 294, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.2343988 0.2772622
## sample estimates:
## mean of x 
## 0.2558305
```

```r
#Inference for the population mean.

dt.stocks[, t.test(isp500, alternative = c("greater"), mu=0.5, conf.level = 0
.99)]
```

```
## 
##  One Sample t-test
## 
## data:  isp500
```

```
## t = 0.83131, df = 294, p-value = 0.2032
## alternative hypothesis: true mean is greater than 0.5
## 99 percent confidence interval:
##  0.1289499       Inf
## sample estimates:
## mean of x
## 0.7045763

#Inference for difference of population means - paired samples

#Paired t-test: t.test(y1,y2,paired=TRUE) where y1 & y2 are numeric.
#Say we want to compare between IDJCOMP and INASDAQ.

dt.stocks[, t.test(idjcomp, inasdaq, paired=TRUE)]

##
##  Paired t-test
##
## data:  idjcomp and inasdaq
## t = -1.178, df = 294, p-value = 0.2397
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9038613  0.2269799
## sample estimates:
## mean of the differences
##              -0.3384407

#Inference for difference of population means - independent samples

#Impact of crisis on stock indices

dt.stocks[, postcrisis:=ifelse(year>2008,1,0)]

dt.stocks[postcrisis==0, mean(idjcomp, na.rm=TRUE)]

## [1] 0.5946696

dt.stocks[postcrisis==1, mean(idjcomp, na.rm=TRUE)]

## [1] 1.042353

#Then you can use the t.test to check whether the difference in means before
and after the crisis is statistically significant.

dt.stocks[, t.test(idjcomp ~ postcrisis)]

##
##  Welch Two Sample t-test
##
## data:  idjcomp by postcrisis
## t = -0.77236, df = 103.8, p-value = 0.4417
```

```
## alternative hypothesis: true difference in means between group 0 and group
1 is not equal to 0
## 95 percent confidence interval:
##  -1.5971454  0.7017787
## sample estimates:
## mean in group 0 mean in group 1
##       0.5946696       1.0423529

#You can also use the t-test to compare between the means of two different va
riables. Independent 2-group t-test: t.test(y1,y2) where y1 and y2 are numeri
c.
dt.stocks[, t.test(idjcomp, inasdaq, var.equal=TRUE)]

##
##  Two Sample t-test
##
## data:  idjcomp and inasdaq
## t = -0.75383, df = 588, p-value = 0.4513
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.2202060  0.5433246
## sample estimates:
## mean of x mean of y
## 0.6978644 1.0363051

dt.stocks[, t.test(idjcomp, inasdaq, var.equal=FALSE)]

##
##  Welch Two Sample t-test
##
## data:  idjcomp and inasdaq
## t = -0.75383, df = 486.57, p-value = 0.4513
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.2205853  0.5437039
## sample estimates:
## mean of x mean of y
## 0.6978644 1.0363051
```