LAB 03 - Hypothesis Testing (Examples)

AML

28/10/2020

Example 1:

```
alpha = .05 # significance level
z.half.alpha = qnorm(1-alpha/2)
c(-z.half.alpha, z.half.alpha) # critical values
```

```
## [1] -1.959964  1.959964
```

```
xbar = 1000/25 # sample mean
mu0 = 45 # hypothesized value
sigma = sqrt(500) # population standard deviation
n = 25 # sample size
z = (xbar-mu0)/(sigma/sqrt(n)) # test statistic
z
```

```
## [1] -1.118034
```

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

```
## [1] 0.8682238
```

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

```
## [1] 0.1317762
```

```
pval = 2 * pnorm(z) # lower tail
pval # two-tailed p-value
```

```
## [1] 0.2635525
```

```
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
```

```
xbar = 1000/25 # sample mean
mu0 = 45 # hypothesized value
s = sqrt(400) # sample standard deviation
n = 25 # sample size
t = (xbar-mu0)/(s/sqrt(n)) # test statistic
t
```

```
## [1] -1.25
```

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

```
## [1] 0.8883243
```

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

```
## [1] 0.1116757
```

```
pval = 2 * pt(t, df=25-1) # lower tail
pval # two-tailed p-value
```

```
## [1] 0.2233515
```

```
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
```

```
sigma_sqr_0 = 500 # hypothesized value
s_sqr = 400 # sample standard deviation
n = 25 # sample size
3
```

```
## [1] 3
```

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

```
## [1] 19.2
```

Example 2:

```
alpha = .05 # significance level
z.alpha = qnorm(1-alpha)
z.alpha # critical value
```

```
## [1] 1.644854
```

```
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
```

Example 3:

```
nX = 21
nY = 25
sX = 1.30
sY = 1.16
4
```

```
## [1] 4
```

```
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
```

```
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
```

```
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 # two-tailed p-value
```

```
## [1] 0.04740744
```

Example 4

```
library(data.table)
setwd("/Users/andrea/Desktop/UEA/Classes/Econometrics/Data")
# change the file's path to your own
dt.stocks <- data.table(read.csv("data_r.csv"))
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
```

Confidence Intervals

Calculate the 95% confidence interval for the stocks' means. Example:

```
xbar <- dt.stocks[, mean(idjcomp, na.rm=TRUE)]
s <- dt.stocks[, sd(idjcomp, na.rm=TRUE)]
n <- dt.stocks[, length(which(!is.na(idjcomp)))]
error <- qnorm(0.975)*s/sqrt(n)
left <- xbar-error
right <- xbar+error
left
```

```
## [1] 0.2391836
```

```
right
```

```
## [1] 1.156545
```

Inference for the population mean.

Let's look at some examples using our stock data. For instance, if we wanted to check whether the mean of S&P500 is equal to zero we would write:

```
dt.stocks[, t.test(isp500)]
```

```
##
##  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
```

If we wanted to test whether it is greater than some specific value (say 0.5) and use a 99% confidence interval, we would write:

```
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

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
```

Impact of crisis on stock indices

Create an indicator variable that takes the value of 1 if it is post-2008 (year of the economic crisis).

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

You can use the indicator variable to look at the mean value of the stock variation before and after the crisis.

```
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 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 variables.

```
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
```