Stat 107: Introduction to Business and Financial Statistics Homework 2: Due Monday, Sept 19

Xiner Zhou

September 16, 2016

# (1)

Suppose monthly returns of AAPL stock are normally distributed with mean 2.48% and standard deviation 12.27%. There is a 1% chance that AAPL monthly returns will be below what value?

qnorm(0.01,mean=0.0248,sd=0.1227)

## [1] -0.2606429

# (2)

Suppose that annual stock returns for a particular company are normally distributed with a mean of 16% and a standard deviation of 10%. You are going to invest in this stock for one year.

##### (a)

Find that the probability that your one-year return will exceed 30%

1-pnorm(0.3, mean=0.16, sd=0.1)

## [1] 0.08075666

##### (b)

Find that probability that your one-year return will be negative.

pnorm(0, mean=0.16, sd=0.1)

## [1] 0.05479929

# (3)

The variance of Stock A is 0.0016, the variance of the market is 0.0049 and the covariance between the two is 0.0026. What is the correlation coefficient?

cat("The correlation coefficient between the stock and the market=", 0.0026/(sqrt(0.0016)\*sqrt(0.0049)), "\n")

## The correlation coefficient between the stock and the market= 0.9285714

# (4)

The distribution of annual returns on common stocks is roughly symmetric, but extreme observations are more frequent than in a normal distribution. Because the distribution is not strongly nonnormal, the mean return over even a moderate number of years is close to normal. Annual real returns on the Standard & Poor's 500-Stock Index over the period 1871 to 2004 have varied with mean 9.2% and standard deviation 20.6%. Andrew plans to retire in 45 years and is considering investing in stocks.

##### (a)

What is the probability (assuming that the past pattern of variation continues) that the mean annual return on common stocks over the next 45 years will exceed 15%?

1-pnorm(0.15, mean=0.092, sd=0.206/sqrt(2004-1871+1))

## [1] 0.0005586024

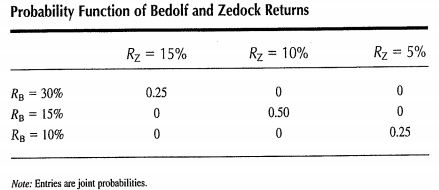
##### (b)

What is the probability that the mean return will be less than 5%?

pnorm(0.05, mean=0.092, sd=0.206/sqrt(2004-1871+1))

## [1] 0.009134461

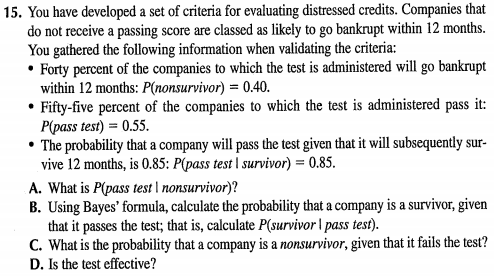
# (5) QIA Workbook: Chapter 4, numbers 14

. .

# Expected value of the returns on Bedolf Corporation is  
ERb<-0.3\*0.25+0.15\*0.5+0.1\*0.25  
# Expected value of the returns on Zedock Corporation is  
ERz<-0.15\*0.25+0.1\*0.5+0.05\*0.25  
  
# Covariance is  
cat("Covariance=",(0.3-ERb)\*(0.15-ERz)\*0.25+(0.15-ERb)\*(0.1-ERz)\*0.50+(0.1-ERb)\*(0.05-ERz)\*0.25,"\n")

## Covariance= 0.0025

# (6) QIA Workbook: Chapter 4, numbers 15

.

##### A:

By the Law of Total Probability, P(pass test)=P(pass test|nonsurvivor)P(nonsurvivor) + P(pass test|survivor)P(survivor), therefore, P(pass test|nonsurvivor)=(P(pass test)-P(pass test|survivor)P(survivor))/P(nonsurvivor)

cat("P(pass test|nonsurvivor)=",(0.55-0.85\*(1-0.4))/0.4,"\n")

## P(pass test|nonsurvivor)= 0.1

##### B:

By the Bayes' Rule, P(survivor|pass test)=P(survivor & pass test)/P(pass test)=P(pass test|survivor)P(survivor)/P(pass test)

cat("P(survivor|pass test)=",0.85\*(1-0.4)/0.55,"\n")

## P(survivor|pass test)= 0.9272727

##### C:

By the Bayes' Rule, P(nonsurvivor|not pass test)=P(nonsurvivor & not pass test)/P(not pass test)=P(not pass test|nonsurvivor)P(nonsurvivor)/P(not pass test)=(1-P(pass test|nonsurvivor))P(nonsurvivor)/(1-P(pass test))

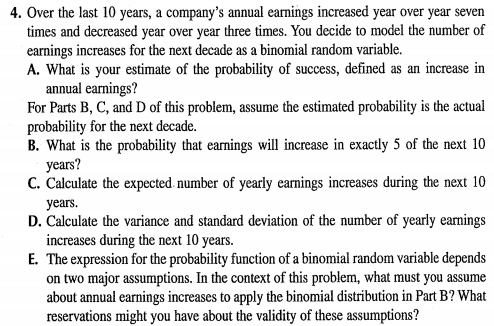
cat("P(nonsurvivor|not pass test)=",(1-0.1)\*0.4/(1-0.55),"\n")

## P(nonsurvivor|not pass test)= 0.8

##### D:

Answer: The purpose of the test is to detect companies likely to go bankrupcy, the null hypothesis is "The company goes bankrupcy in the next 12 months". The test is effective if the type I error <5%, which is: P(pass test|nonsurvivor)<0.05. However, from A) we know the type I error is 10% whihc is too high, therefore, the test is not effective, it bears too high risk of not detecting true nonsurvivors.

# (7) QIA Workbook; Chapter 5, numbers 4

.

#### A:

The estimate of the probability of success,, defined as an increase in annual earnings, is 0.7

##### B:

dbinom(5, size=10, prob = 0.7)

## [1] 0.1029193

##### C:

cat("Expected number of yearly earnings increases during the next 10 years=",5\*0.7,"\n")

## Expected number of yearly earnings increases during the next 10 years= 3.5

##### D:

cat("Variance of the number of yearly earnings increases during the next 10 years=",5\*0.7\*(1-0.7),"\n")

## Variance of the number of yearly earnings increases during the next 10 years= 1.05

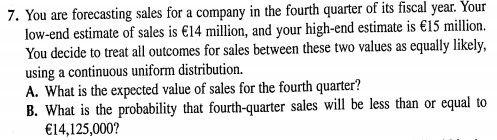
cat("Standard deviation of the number of yearly earnings increases during the next 10 years=",sqrt(5\*0.7\*(1-0.7)),"\n")

## Standard deviation of the number of yearly earnings increases during the next 10 years= 1.024695

##### E:

Answer: The assumptions we make are I.I.D. (Indepedent Identical Distribution), that is, the actual earning increases in each year in the next 10 years is 1) independent 2) identically distributed as Bernoulli distribution with probability of success 0.7. The validity of these assumptions are violated if, the actual earnings have serial correlation, or the probability of increase changes over time.

# (8) QIA Workbook; Chapter 5, numbers 7

.

##### A:

cat("The expected value of sales for the fourth quarter=",(14+15)/2,"million \n")

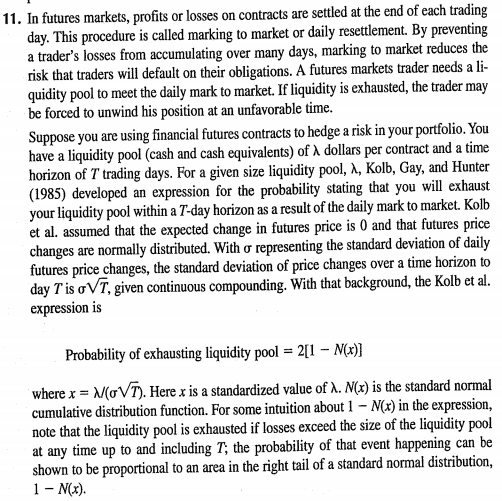
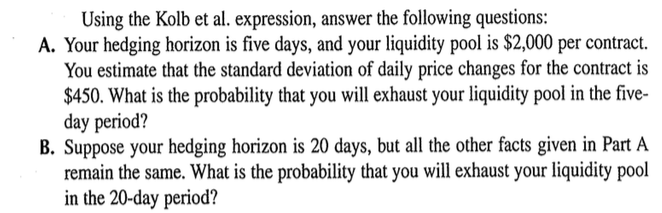
## The expected value of sales for the fourth quarter= 14.5 million

##### B:

cat("The probability that fourth-quarter sales will be less than or equal to 14,125,000=",(14.125-14)/(15-14),"\n")

## The probability that fourth-quarter sales will be less than or equal to 14,125,000= 0.125

# (9) QIA Workbook; Chapter 5, numbers 11

. .

##### A:

cat("Probability of exhausting liquidity pool=",2\*(1-pnorm(2000/(450\*sqrt(5)))),'\n')

## Probability of exhausting liquidity pool= 0.04685418

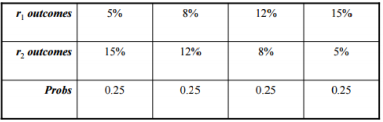
##### B:

cat("Probability of exhausting liquidity pool=",2\*(1-pnorm(2000/(450\*sqrt(20)))),'\n')

## Probability of exhausting liquidity pool= 0.3203164

# (7)

Suppose an investor can choose between two independent assets 1 and 2. The probability distributions for the rates of return for each asset are provided below.



##### a)

Compute the expected value of each asset

Er1<-0.05\*0.25+0.08\*0.25+0.12\*0.25+0.15\*0.25  
Er2<-0.15\*0.25+0.12\*0.25+0.08\*0.25+0.05\*0.25  
cat("Expected value of asset 1=", Er1, "\n")

## Expected value of asset 1= 0.1

cat("Expected value of asset 2=", Er2, "\n")

## Expected value of asset 2= 0.1

##### b)

Compute the standard deviation of each asset

SDr1<-sqrt((0.05-Er1)^2\*0.25+(0.08-Er1)^2\*0.25+(0.12-Er1)^2\*0.25+(0.15-Er1)^2\*0.25)  
SDr2<-sqrt((0.15-Er2)^2\*0.25+(0.12-Er2)^2\*0.25+(0.08-Er2)^2\*0.25+(0.05-Er2)^2\*0.25)  
cat("Standard deviation of asset 1=", SDr1, "\n")

## Standard deviation of asset 1= 0.03807887

cat("Standard deviation of asset 2=", SDr2, "\n")

## Standard deviation of asset 2= 0.03807887

##### c)

Based on expected return and the variance of return, which asset do you prefer and why? On what other basis might you prefer one asset to another?

Answer: Since R1 and R2 have the same expected return and variance, I have no prefernece between the two. I might prefer one asset over another if I have other information about the two assets, if I believe one is going to perform better.

##### d) Assume you hold a portfolio with 50% held in asset 1 and 50% held in asset 2. What is the expected return and variance of your portfolio? Note that the portfolio return is a random variable written as R = 0.5R1+0.5R2.

cat("The expected value of the portfolio is: \n","E(R)=E(0.5R1+0.5R2)=0.5E(R1)+0.5E(R2)=",0.5\*Er1+0.5\*Er2,'\n')

## The expected value of the portfolio is:   
## E(R)=E(0.5R1+0.5R2)=0.5E(R1)+0.5E(R2)= 0.1

cat("The variance of the portfolio is: \n","Var(R)=Var(0.5R1+0.5R2)=0.25Var(R1)+0.25Var(R2)+2\*0.5\*0.5Cov(R1,R2)=0.25Var(R1)+0.25Var(R2)=",0.25\*SDr1^2+0.25\*SDr2^2,'\n')

## The variance of the portfolio is:   
## Var(R)=Var(0.5R1+0.5R2)=0.25Var(R1)+0.25Var(R2)+2\*0.5\*0.5Cov(R1,R2)=0.25Var(R1)+0.25Var(R2)= 0.000725

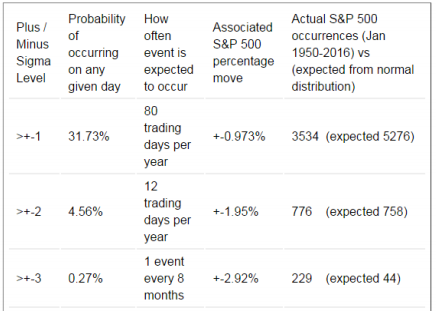
1. In this problem we are going to reproduce the first table from the article [article link](https://sixfigureinvesting.com/2016/03/modeling-stock-market-returns-with-laplace-distribution-instead-of-normal/)

Our numbers will be slightly different but close.To grab historical data for the S&P500 do the following

library(quantmod)  
spydata=getSymbols("^GSPC",from="1950-01-01",auto.assign=FALSE)  
spyrets=dailyReturn(Ad(spydata))  
spyrets=spyrets[-1]  
sd(spyrets)

## [1] 0.00967957

Note we find the historical standard deviation of daily returns to be 0.967% and the quoted paper uses 0.973%. We assume 250 trading days in a year. Simply reproduce the following table using our data from R [so it will be slightly different]. The expected number is 250days*66 years/day*prob of occurrence per year.



mu.spyrets<-mean(spyrets)  
sd.spyrets<-sd(spyrets)  
   
# Define X=daily return, X is Normally distributed with mean and standard deviation equal to historical values  
  
# Define empty columns of the final table  
 c1<-character(3)  
 c2<-character(3)  
 c3<-character(3)  
 c4<-character(3)  
 c5<-character(3)  
   
for(i in 1:3){  
 # prob of occuring on any given day  
 col2.num<-pnorm(mu.spyrets-i\*sd.spyrets, mean=mu.spyrets, sd=sd.spyrets)+1-pnorm(mu.spyrets+i\*sd.spyrets, mean=mu.spyrets, sd=sd.spyrets)  
 # how often event is expected to occur  
 if (i<3){col3.num<-ceiling(col2.num\*250) } else{col3.num<-ceiling(col2.num\*250\*8/12) }  
 # associated S&P 500 percentage move  
 col4.num<-i\*sd.spyrets\*100  
 # actual S&P 500 occurences  
 col5.num.actual<-sum(spyrets>mu.spyrets+i\*sd.spyrets)+sum(spyrets<mu.spyrets-i\*sd.spyrets)  
 # expected S&P 500 occurences  
 col5.num.expected<-ceiling(col2.num\*250\*66)  
   
 # pu into Table cells  
 c1[i]<-paste(">+-",i)  
 c2[i]<-paste(round(100\*col2.num,digits = 2),"%")  
 if (i<3){c3[i]<-paste(col3.num,"trading days per year", sep=" ")}  
 else{c3[i]<-paste(col3.num,"event every 8 months", sep=" ")}  
   
 c4[i]<-paste("+-",round(100\*i\*sd.spyrets,digits = 2),"%")  
 c5[i]<-paste(col5.num.actual,"(expected",col5.num.expected,")")  
}  
  
# Make Table  
myTable<-data.frame(c1,c2,c3,c4,c5)  
names(myTable)<-c("Plus/Minus Sigma Level",  
 "Prob of occuring on any given day",  
 "How often event is expected to occur",  
 "Associated percentage move",  
 "Actual occurences vs Expected from normal distribution")  
  
library(knitr)  
kable(myTable)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plus/Minus Sigma Level | Prob of occuring on any given day | How often event is expected to occur | Associated percentage move | Actual occurences vs Expected from normal distribution |
| >+- 1 | 31.73 % | 80 trading days per year | +- 0.97 % | 3589 (expected 5236 ) |
| >+- 2 | 4.55 % | 12 trading days per year | +- 1.94 % | 783 (expected 751 ) |
| >+- 3 | 0.27 % | 1 event every 8 months | +- 2.9 % | 235 (expected 45 ) |

# (9)

Test using the S&P 500 data used previously that the daily standard deviation of returns from 1990 to 1995 equals the standard deviation of returns from 2000 to 20005. Use the R routine var.test(x,y).

library(quantmod)  
spy1=getSymbols("^GSPC",from="1990-01-01",to="1995-12-31",auto.assign=FALSE)  
spy2=getSymbols("^GSPC",from="2000-01-01",to="2005-12-31",auto.assign=FALSE)  
  
var.test(dailyReturn(Ad(spy1)), dailyReturn(Ad(spy2)))

##   
## F test to compare two variances  
##   
## data: dailyReturn(Ad(spy1)) and dailyReturn(Ad(spy2))  
## F = 0.3638, num df = 1516, denom df = 1507, p-value < 2.2e-16  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.3288920 0.4024051  
## sample estimates:  
## daily.returns  
## daily.returns 0.3637992  
## attr(,"names")  
## [1] "ratio of variances"

Answer: Since the p-value is less than 0.05 we reject the null hypothesis. Our conclusion at the 5% level of confidence is, there is sufficient sample evidence to support the claim that the standard deviation of daily adjusted returns from 1990 to 1995 is difference from that from 2000 to 2005.

# (10)

Again using the historical data, note that the growth of $1 over this period can be obtained using the command prod((1+spyrets)).

##### a.

What would the growth be if someone had missed the 10 best performing days over this period [this is the argument to not market time but instead buy and hold].

spydata=getSymbols("^GSPC",from="1950-01-01",auto.assign=FALSE)  
spyrets=as.numeric(dailyReturn(Ad(spydata))) # no time stamp, pure numeric vecotr  
spyrets=spyrets[-1]  
  
# if buy and hold  
prod(1+spyrets)

## [1] 128.401

# To sort by returns using order()  
spyrets.sorted<-spyrets[order(spyrets,decreasing =TRUE)]  
   
# missed the 10 best performing days 1950 until now  
spyrets.nobest<-spyrets.sorted[11:length(spyrets.sorted)]  
cat("If missed the 10 best performing days, growth=",prod(1+spyrets.nobest),'\n')

## If missed the 10 best performing days, growth= 61.9599

##### b.

What would the growth be if someone had missed the 10 worst performing days over this period [this is the argument to market time and instead buy and hold].

# missed the 10 worst performing days 1950 until now  
spyrets.noworst<-spyrets.sorted[1:(length(spyrets.sorted)-10)]  
cat("If missed the 10 worst performing days, growth=",prod(1+spyrets.noworst),'\n')

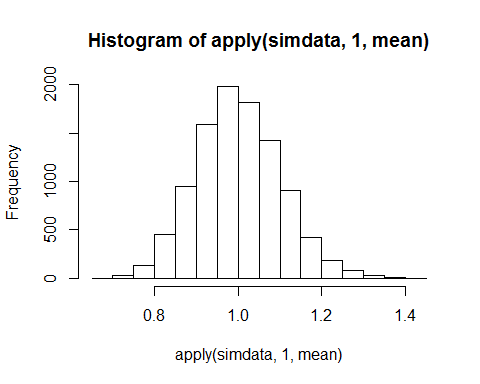
## If missed the 10 worst performing days, growth= 334.0691

# (11)

Recall that the R command apply allows us to easily and quickly calculate the mean (or sd or sum or any function) of each row or column of a matrix. Using this command we can quickly write code to demonstrate the Central Limit Theorem.

The following code simulates drawing 10000 samples of size 100 from an exponential distribution with shape parameter=1, and then drawing the resulting histogram.

set.seed(02138)  
simdata=matrix(nrow=10000,ncol=100,rexp(100\*10000,1))  
hist(apply(simdata,1,mean))



Explain in easy to understand language what the above code is doing. Using the above code, show how the Central Limit Theorem works for the Gamma distribution with shape=5 and rate=5 parameters [R code is rgamma(n,5,5)]. Use sample sizes of 10, 50 and 100. For each simulation run, report the mean of the 10000 sample means produced and the standard deviation of the 10000 sample means produced. Comment on what you find.

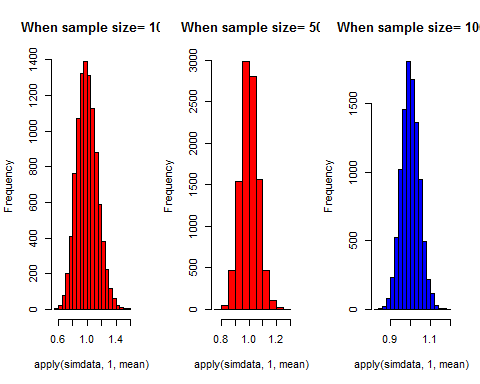
Answer: The first line of code draws 100\*10000 independent random samples from an exponential distribution with shape parameter=1, put them into a matrix, each row of the resulting matrix is a random sample of size 100, therefore, the code essentially generates 10000 samples of size 100. The second line of codes plots a histogram of the means of each sample, by applying mean function to each row.

set.seed(02138)  
n<-c(10,50,100)  
par(mfrow=c(1,3))  
for(i in 1:3){  
 sample.size<-n[i]  
 cat("When sample size=",sample.size,'\n')  
   
 simdata<-matrix(nrow=10000, ncol=sample.size, rgamma(sample.size\*10000,5,5))  
 cat("The mean of the 10000 sample means=",mean(apply(simdata,1,mean)),'\n')  
 cat("The standard deviation of the 10000 sample means=",sd(apply(simdata,1,mean)),'\n')  
 hist(apply(simdata,1,mean),main = paste("When sample size=",sample.size), col = sample.size)  
}

## When sample size= 10   
## The mean of the 10000 sample means= 0.9976744   
## The standard deviation of the 10000 sample means= 0.1419841

## When sample size= 50   
## The mean of the 10000 sample means= 1.000864   
## The standard deviation of the 10000 sample means= 0.06296471

## When sample size= 100   
## The mean of the 10000 sample means= 0.9994539   
## The standard deviation of the 10000 sample means= 0.04455103



Answer: The sample means are all close to the true mean which is 1. However, the standard deviation decreases substantially while increasing the sample size. This tendency is in agreement with the Central Limit Theorem. Compare the histograms, we can easily see that the centrality is all around 1, while the dispersion is smaller and smaller while increasing the sample size.

Statistical Methods Review Problems. This part of this homework is to simply reignite the brain cells involving confidence intervals and hypothesis testing. The R cookbook is the best place to go to see how to compute confidence intervals and hypothesis tests in R. For each problem below, use R to answer the question and clearly state the conclusion of your test, or clearly give the confidence interval.

# (12)

In a study of the length of time that students require to earn bachelor's degrees, 80 students are randomly selected and they are found to have a mean of 4.8 years and standard deviation of 2.2 years. Construct a 95% confidence interval estimate of the population mean. Does the resulting confidence interval contradict the fact that 39% of students earn their bachelor's degrees in four years?

mu<-4.8  
sigma<-2.2  
c(mu-1.96\*sigma/sqrt(80),mu+1.96\*sigma/sqrt(80))

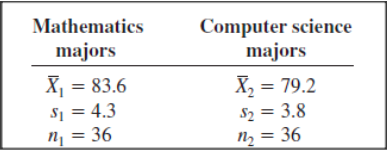
## [1] 4.317904 5.282096

Answer: The 95% confidence interval is (4.32,5.28), it doesn't contradict the fact that 39% of students earn their bachelor's degrees in 4 years. The 95% CI does Not say that, 95% of students earn their degrees between (4.32,5.28). Rather, a confidence interval is an observed interval, in principle different from sample to sample, that frequently includes the value of an unobserved parameter of interest (in this case, the length of time getting a bachelor's degree) if the experiment is repeated. How frequently the observed interval contains the parameter is determined by the confidence level (usually 95%). More specifically, the meaning the CI is that, if CI are constructed across many separate data analyses of replicated experiments, the proportion of such intervals that contain the true value of the parameter will match the given confidence level.

So, what the 95% CI really tells us is that, if replicate the experiment many many times, 95% of times we will see the observed CI include the true average length of time that students requir to earn a bachelor's degree. We can say that, there is a 95% chance that (4.32,5.28) includes the true average length of time that students require to get a bachelor's degree.

# (13)

Two groups of students are given a problem-solving test, and the results are compared. Find and interpret the 95% confidence interval of the true difference in means.



library(BSDA)  
tsum.test(83.6,4.3,36,79.2,3.8,36)

##   
## Welch Modified Two-Sample t-Test  
##   
## data: Summarized x and y  
## t = 4.6005, df = 68.957, p-value = 1.859e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 2.491991 6.308009  
## sample estimates:  
## mean of x mean of y   
## 83.6 79.2

Answer: The 95% CI is (2.491991, 6.308009). If we replicate the experiment many many times, 95% of times we will see the observed CI includes the true difference in means; or, there is a 95% chance that (2.491991, 6.308009) covers the true difference in means.

# (14)

A recent Gallup poll consisted of 1012 randomly selected adults who were asked whether "cloning of humans should or should not be allowed." Results showed that 892 of those surveyed indicated that cloning should not be allowed. Construct a 95% confidence interval estimate of the proportion of adults believing that cloning of humans should not be allowed.

prop.test(892,1012,correct = FALSE)

##   
## 1-sample proportions test without continuity correction  
##   
## data: 892 out of 1012, null probability 0.5  
## X-squared = 588.92, df = 1, p-value < 2.2e-16  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.8600478 0.8999133  
## sample estimates:  
## p   
## 0.8814229

Answer: 95% confidence interval is (0.8600478, 0.8999133). There is a 95% chance that (86.00%, 89.99%) includes the true proportion of adults believing that cloning of human should not be allowed.

# (15)

In a sample of 80 Americans, 55% wished that they were rich. In a sample of 90 Europeans, 45% wished that they were rich. Find and interpret the 95% confidence interval of the true difference in proportions.

prop.test(c(0.55\*80, 0.45\*90), c(80, 90), correct = FALSE)

##   
## 2-sample test for equality of proportions without continuity  
## correction  
##   
## data: c(0.55 \* 80, 0.45 \* 90) out of c(80, 90)  
## X-squared = 1.6942, df = 1, p-value = 0.1931  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## -0.04982832 0.24982832  
## sample estimates:  
## prop 1 prop 2   
## 0.55 0.45

Answer: 95% confidence interval is (-0.04982832, 0.24982832). There is a 95% chance that (-4.98%, 24.98%) includes the true difference (Americans vs Europeans) in proportions of those who wished they were rich.

# (16)

For this problem, download data for the last year (9/01/2014 to 9/01/2015) for AAPL and CAT. We are going to work through some items in the R Cookbook so look there if you get stuck. [use adjusted closing prices for everything].

##### (a)

Run a hypothesis test to see if AAPL or CAT daily returns are normally distributed. Interpret the output, what is the conclusion of the hypothesis test?

AAPL<-getSymbols("AAPL",from="2014-09-01",to="2015-09-01",auto.assign=FALSE)  
CAT<-getSymbols("CAT",from="2014-09-01",to="2015-09-01",auto.assign=FALSE)  
  
shapiro.test(as.numeric(dailyReturn(Ad(AAPL))))

##   
## Shapiro-Wilk normality test  
##   
## data: as.numeric(dailyReturn(Ad(AAPL)))  
## W = 0.97627, p-value = 0.0003108

shapiro.test(as.numeric(dailyReturn(Ad(CAT))))

##   
## Shapiro-Wilk normality test  
##   
## data: as.numeric(dailyReturn(Ad(CAT)))  
## W = 0.96256, p-value = 3.671e-06

Answer: Both p-values are less than 0.05 we reject the null hypothesis. Our conclusion at the 5% level of confidence is that, there is sufficient sample evidence to support the claim that daily returns of AAPL and CAT are not normally distributed.

##### (b)

Run a hypothesis test to see if AAPL or CAT monthly returns are normally distributed. Interpret the output, what is the conclusion of the hypothesis test?

shapiro.test(as.numeric(monthlyReturn(Ad(AAPL))))

##   
## Shapiro-Wilk normality test  
##   
## data: as.numeric(monthlyReturn(Ad(AAPL)))  
## W = 0.89676, p-value = 0.1207

shapiro.test(as.numeric(monthlyReturn(Ad(CAT))))

##   
## Shapiro-Wilk normality test  
##   
## data: as.numeric(monthlyReturn(Ad(CAT)))  
## W = 0.96987, p-value = 0.8928

Answer: Both p-values are larger than 0.05 we accept the null hypothesis. Our conclusion at the 5% level of confidence is that, there is no sufficient sample evidence to reject the claim that monthly returns of AAPL and CAT are normally distributed.

##### (c)

Test if AAPL and CAT have the same average daily return (see comparing the means of two samples in the R Cookbook).

tsum.test(mean(as.numeric(dailyReturn(Ad(AAPL)))),   
 sd(as.numeric(dailyReturn(Ad(AAPL)))),   
 length(as.numeric(dailyReturn(Ad(AAPL)))),  
 mean(as.numeric(dailyReturn(Ad(CAT)))),   
 sd(as.numeric(dailyReturn(Ad(CAT)))),   
 length(as.numeric(dailyReturn(Ad(CAT)))))

##   
## Welch Modified Two-Sample t-Test  
##   
## data: Summarized x and y  
## t = 1.1644, df = 500.01, p-value = 0.2448  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.001100890 0.004304077  
## sample estimates:  
## mean of x mean of y   
## 0.0003627482 -0.0012388453

Answer: The p-value is larger than 0.05 we accept the null hypothesis. OUr conclusion at the 5% level of confidence is that, there is no sufficient sample evidence to reject the claim that AAPl and CAT have the same average daily returns.

##### (d)

Test if AAPL and CAT have the same average monthly return (see comparing the means of two samples in the R Cookbook).

tsum.test(mean(as.numeric(monthlyReturn(Ad(AAPL)))),   
 sd(as.numeric(monthlyReturn(Ad(AAPL)))),   
 length(as.numeric(monthlyReturn(Ad(AAPL)))),  
 mean(as.numeric(monthlyReturn(Ad(CAT)))),   
 sd(as.numeric(monthlyReturn(Ad(CAT)))),   
 length(as.numeric(monthlyReturn(Ad(CAT)))))

##   
## Welch Modified Two-Sample t-Test  
##   
## data: Summarized x and y  
## t = 1.2846, df = 23.877, p-value = 0.2112  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.01859108 0.07984341  
## sample estimates:  
## mean of x mean of y   
## 0.006328462 -0.024297703

Answer: The p-value is larger than 0.05 we accept the null hypothesis. OUr conclusion at the 5% level of confidence is that, there is no sufficient sample evidence to reject the claim that AAPl and CAT have the same average monthly returns.

##### (e)

Test if AAPL and CAT have the same standard deviation of daily returns. Use the R routine var.test(dailyReturn(Ad(AAPL)),dailyReturn(Ad(CAT)))

var.test(dailyReturn(Ad(AAPL)),dailyReturn(Ad(CAT)))

##   
## F test to compare two variances  
##   
## data: dailyReturn(Ad(AAPL)) and dailyReturn(Ad(CAT))  
## F = 1.1961, num df = 252, denom df = 252, p-value = 0.1558  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.9338966 1.5320305  
## sample estimates:  
## daily.returns  
## daily.returns 1.196143  
## attr(,"names")  
## [1] "ratio of variances"

Answer: The p-value is larger than 0.05 we accept the null hypothesis. OUr conclusion at the 5% level of confidence is that, there is no sufficient sample evidence to reject the claim that AAPl and CAT have the same standard deviation of daily returns.

##### (f)

Calculate the correlation between AAPL and CAT daily returns, and compute a confidence interval for the correlation. ???

cor(dailyReturn(Ad(AAPL)),dailyReturn(Ad(CAT)))

## daily.returns  
## daily.returns 0.468505

cor.test(dailyReturn(Ad(AAPL)),dailyReturn(Ad(CAT)))

##   
## Pearson's product-moment correlation  
##   
## data: dailyReturn(Ad(AAPL)) and dailyReturn(Ad(CAT))  
## t = 8.4016, df = 251, p-value = 3.297e-15  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3663443 0.5595049  
## sample estimates:  
## cor   
## 0.468505

Answer: The correlation is 0.468505, the p-value (3.297e-15) is less than 0.05, so the correlation between AAPL and CAT daily returns is significantly different from zero. The confidence interval is (0.3663443, 0.5595049), so there is a 95% chance that (0.3663443, 0.5595049) includes the true correlation.