Question 1:

```
> # (1) Perform a t-test for American Express with the null hypothesis
that the mean of its return is zero.
> read.table("/Users/Yisha1/Documents/d_logret_6stocks.txt",header=T)
       Date       Pfizer        Intel    Citigroup      AmerExp
Exxon     GenMotor
1  1-Aug-00 -0.001438612  0.049981263  0.044275101  0.017410003
0.0102248940  0.093294017
2  1-Sep-00  0.017489274 -0.255619266 -0.033536503  0.012656982
0.0379890200 -0.032209239
3  2-Oct-00 -0.017046116  0.034546736 -0.011645582 -0.004897625
0.0003305550 -0.019602167
4  1-Nov-00  0.012012934 -0.072550667 -0.022674793 -0.038275870
-0.0036500200 -0.094891600
5  1-Dec-00  0.016278701 -0.102497868  0.010708311  0.000000000
-0.0052520490  0.012461253
6  2-Jan-01 -0.008063083  0.090223122  0.039900620 -0.066129678
-0.0141692430  0.022971579
7  1-Feb-01 -0.000422980 -0.112194230 -0.055096146 -0.030733152
-0.0140468950  0.000824088
8  1-Mar-01 -0.040906294 -0.035702138 -0.038726816 -0.026380545
-0.0002400080 -0.012105099
9  2-Apr-01  0.024190228  0.069994483  0.038511978  0.011868735
0.0388974880  0.024082196
10 1-May-01 -0.002978787 -0.058260610  0.019333184 -0.002446047
0.0028442560  0.020148775
11 1-Jun-01 -0.029781389  0.034634870  0.013258067 -0.035641970
-0.0068134640  0.053440295
12 2-Jul-01  0.012504432  0.008168789 -0.022187219  0.017739418
-0.0194814020 -0.005100405
13 1-Aug-01 -0.030663200 -0.027529477 -0.038475736 -0.044368019
-0.0146074300 -0.061635162
14 4-Sep-01  0.019815480 -0.135934121 -0.053479798 -0.098043942
-0.0082241460 -0.105946472
15 1-Oct-01  0.019063731  0.077211653  0.050835509  0.006689711
0.0006100500 -0.016274333
16 1-Nov-01  0.015543895  0.126580684  0.023566060  0.048543672
-0.0207262340  0.085210960
17 3-Dec-01 -0.036145791 -0.016421934  0.022871285  0.035242521
0.0215788660 -0.009657415
18 2-Jan-02  0.019356687  0.046876533 -0.025940517  0.002871379
-0.0028078170  0.022139216
19 1-Feb-02 -0.006050198 -0.088680731 -0.020151007  0.007237226
0.0269480740  0.019672220
20 1-Mar-02 -0.013187975  0.027384065  0.039197815  0.050683167
0.0258072640  0.057331233
21 1-Apr-02 -0.038640426 -0.026448085 -0.058277811  0.001375340
-0.0378280050  0.025768635
22 1-May-02 -0.020012226 -0.014900615  0.000481346  0.015691714
-0.0001183520 -0.010495544
```

```
23 3-Jun-02  0.004989620 -0.179572434 -0.046948457 -0.068454444
0.0106401330 -0.065487824
24 1-Jul-02 -0.034159152  0.012261550 -0.062746165 -0.011860070
-0.0465282000 -0.060041503
25 1-Aug-02  0.011452067 -0.051537916  0.022330581  0.009740522
-0.0130506960  0.016998701
26 3-Sep-02 -0.056822917 -0.079127863 -0.043102044 -0.063162423
-0.0457869330 -0.090010126
27 1-Oct-02  0.039382501  0.095369960  0.097624046  0.067951966
0.0233571050 -0.068058029
28 1-Nov-02 -0.001620779  0.082000518  0.022127194  0.029514688
0.0172318270  0.083238291
29 2-Dec-02 -0.013493147 -0.127500953 -0.043258124 -0.040869439
0.0017395890 -0.032155007
30 2-Jan-03 -0.000914625  0.002562217 -0.008110182  0.002151752
-0.0098600090 -0.006417575
31 3-Feb-03 -0.007697729  0.042681011 -0.012956568 -0.024428147
0.0012277850 -0.025617995
32 3-Mar-03  0.018994390 -0.025156666  0.014203546 -0.004565156
0.0116929920 -0.001942487
33 1-Apr-03 -0.005686915  0.053056729  0.056727624  0.057647618
0.0031710110  0.030362391
34 1-May-03  0.005686915  0.054144721  0.021322255  0.041490099
0.0176708400 -0.002801910
35 2-Jun-03  0.041784483 -0.000213046  0.018444872  0.001579917
-0.0059815860  0.008214181
36 1-Jul-03 -0.010109859  0.077829522  0.023189447  0.024870758
-0.0039908770  0.016906014
37 1-Aug-03 -0.045266311  0.060434430 -0.014198430  0.008620388
0.0281661160  0.046380496
38 2-Sep-03  0.006546894 -0.016587184  0.021075597  0.000112293
-0.0129172300 -0.001791893
39 1-Oct-03  0.017184425  0.078321576  0.020888904  0.018572284
-0.0002498100  0.018169063
40 3-Nov-03  0.028255616  0.007861351 -0.003462108 -0.011445240
-0.0015018840  0.006155458
41 1-Dec-03  0.022153888 -0.019719492  0.013782077  0.024270976
0.0541511150  0.096343714
42 2-Jan-04  0.015748075 -0.021237664  0.011862818  0.031325870
-0.0022191900 -0.031390331
43 2-Feb-04  0.002115176 -0.018679024  0.006780909  0.013019280
0.0171231800 -0.009458693
44 1-Mar-04 -0.019288230 -0.030753805  0.012267738 -0.012145545
-0.0060304690 -0.007941261
45 1-Apr-04  0.008607804 -0.024068646 -0.027843588 -0.024949111
0.0098634440  0.001620126
46 3-May-04 -0.003063819  0.045791862 -0.015263851  0.015239967
0.0099553100 -0.014176433
47 1-Jun-04 -0.013135825 -0.014787260  0.000692103  0.006594513
0.0114509890  0.011337234
```

```
48 1-Jul-04 -0.030491723 -0.053760665 -0.019188415 -0.009580051
0.0180838070 -0.033399340
49 2-Aug-04  0.011876253 -0.058250748  0.023904782 -0.002001822
0.0007736270 -0.013614662
50 1-Sep-04 -0.028332050 -0.025811490 -0.023595125  0.012265109
0.0204755860  0.012073829
51 1-Oct-04 -0.024200939  0.045251691  0.006452318  0.014388280
0.0079454680 -0.042109935
52 1-Nov-04 -0.015356644  0.003157084  0.003644451  0.021085951
0.0198988810  0.006031965
53 1-Dec-04 -0.014084690  0.019040089  0.032148678  0.005093112
0.0000864354  0.016341604
54 3-Jan-05 -0.046516472 -0.017862074  0.007701610 -0.022982941
0.0028427590 -0.036824626
55 1-Feb-05  0.039975516  0.030472706 -0.008076244  0.006507102
0.0909272820 -0.007985210
56 1-Mar-05 -0.000338104 -0.013929818 -0.026065490 -0.021854120
-0.0261940260 -0.083992068
57 1-Apr-05  0.014633051  0.005252870  0.023245386  0.011111802
-0.0191303460 -0.042013994
58 2-May-05  0.014630589  0.060803225  0.001318328  0.009356124
-0.0041946140  0.079608491
59 1-Jun-05 -0.005088825 -0.015344193 -0.008162243 -0.004091884
0.0097251450  0.032753690
60 1-Jul-05 -0.017295755  0.018252426 -0.022110024  0.014246467
0.0095867970  0.034619924
61 1-Aug-05 -0.014040733 -0.022132340  0.002713407  0.001894712
0.0105471960 -0.025993870
62 1-Sep-05 -0.008682706 -0.018343450  0.016994806  0.016950229
0.0256082320 -0.047977476
63 3-Oct-05 -0.060303366 -0.020818266  0.002497608 -0.003389887
-0.0538313140 -0.048092196
64 1-Nov-05  0.002411637  0.058709923  0.038299120  0.024183203
0.0319235510 -0.070676054

> ae<-read.table("d_logret_6stocks.txt", header=T)
> t.test(ae[,5],mu=0)

        One Sample t-test

data:  ae[, 5]
t = 0.18782, df = 63, p-value = 0.8516
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.006792534  0.008201838
sample estimates:
   mean of x
0.0007046519
#[,5] stands for column five which is american express
# Therefore, p-value = 0.8516 which is fail reject null. It means
```

```
return of American express is 0.

#(2) Perform a Wilcoxon signed-rank test for American Express with the
null hypothesis that the mean of its return is zero.
> wilcox.test(ae[,5],mu=0)

        Wilcoxon signed rank test with continuity correction

data:  ae[, 5]
V = 1153, p-value = 0.3225
alternative hypothesis: true location is not equal to 0
> # Therefore, return of American Express is 0

> # (3) Perform a two-sample t-test to conclude if the mean returns of
Pfizer and American Express are same or not.
> t.test(ae[,2],ae[,5])

        Welch Two Sample t-test

data:  ae[, 2] and ae[, 5]
t = -1.0028, df = 118.21, p-value = 0.318
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.014118044  0.004626111
sample estimates:
    mean of x     mean of y
-0.0040413145  0.0007046519
> #Therefore, p-value = 0.318 which is also fail to reject null. There
is no difference between the mean returns of Pfizer and American
Express

> # (4) Perform a two-sample Wilcoxon test to conclude if the mean
returns of Pfizer and American Express are same or not.
> wilcox.test(ae[,2],ae[,5])

        Wilcoxon rank sum test with continuity correction

data:  ae[, 2] and ae[, 5]
W = 1757, p-value = 0.1662
alternative hypothesis: true location shift is not equal to 0
> # Therefore, p-value = 0.1662 which is fail reject null. There is no
difference between the mean returns of Pfizer and American Express.

> # (5) Compare the variance of returns for Pfizer and American
Express.
> var.test(ae[,2],ae[,5])

        F test to compare two variances

data:  ae[, 2] and ae[, 5]
```

```
F = 0.5914, num df = 63, denom df = 63, p-value = 0.03896
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3592924 0.9734621
sample estimates:
ratio of variances
          0.591403
> # Therefore, p-value = 0.03896 which is reject null. The variance
between Pfizer and American Express is not equal.
```

Question 2:

```
> # Do the data provide sufficient evidence to indicate that rats
exposed to a 5°C environment have a higher mean blood pressure than
rats exposed to a 26°C environment? Test by using α = .05.

> BP5C <- c(384, 369, 354, 375, 366, 423)
> BP26C <- c(152, 157, 179, 182, 176, 149)
> shapiro.test(BP5C)

        Shapiro-Wilk normality test

data:  BP5C
W = 0.8694, p-value = 0.2238

> shapiro.test(BP26C)

        Shapiro-Wilk normality test

data:  BP26C
W = 0.85134, p-value = 0.1614

> var.test(BP5C, BP26C)

        F test to compare two variances

data:  BP5C and BP26C
F = 2.6306, num df = 5, denom df = 5, p-value = 0.3121
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.3680964 18.7989696
sample estimates:
ratio of variances
          2.630558

> t.test(BP5C, BP26C, alternative="greater", var.equal=T)

        Two Sample t-test

data:  BP5C and BP26C
```

```
t = 18.509, df = 10, p-value = 2.285e-09
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 191.8422      Inf
sample estimates:
mean of x mean of y
 378.5000  165.8333

> # The p-value for the group at 5°C is 0.1614, and the p-value for
the group at 26°C is 0.2238. They are both normal distribution since
both p-values are greater than a = 0.1.
># For variance test, F-test is 0.3123 which is greater than a = 0.05.
Fail to reject the null hypothesis. Therefore, they are the same
> # P-value for Pooled-Variance T-test is less than a = 0.05.
Therefore, we reject the null hypothesis at the significance level a =
0.05. So the rats have higher blood pressure at 5 degree celsius than
at 26 degree celsius
```

Question 3:

```
> # (a) According to the data, can you conclude, at the significance
level of 0.10, that the corneal thickness is not equal for affected
versus unaffected eyes? Please write the entire R code to check the
assumptions necessary and to perform the test.

> affect <- c(488, 478, 480, 426, 440, 410, 458, 460)
> notaffect <- c(484, 478, 492, 444, 436, 398, 464, 476)
> diff = affect - notaffect
> shapiro.test(diff)

        Shapiro-Wilk normality test

data:  diff
W = 0.94404, p-value = 0.6512
# p-value = 0.6515. Fail to reject null. Data is normal.

> # (b) Calculate a 90% confidence interval for the mean difference in
thickness.
> t.test(diff,mu=0,conf.level = 0.9)

        One Sample t-test

data:  diff
t = -1.053, df = 7, p-value = 0.3273
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
 -11.196549   3.196549
sample estimates:
mean of x
```

```
       -4

> # p -value = 0.3273. Fail to reject null. Therefore, there is no
difference between eye affected and not effected
```

Question 4:

```
> # (a). Please check the normality of the data.
> time <- c(28, 25, 27, 31, 10, 26, 30, 15, 55, 12, 24, 32, 28, 42,
38)
> shapiro.test(time)

        Shapiro-Wilk normality test

data:  time
W = 0.94167, p-value = 0.4038
# p-vlaue = 0.4038 which is larger than 0.1. Fail to reject null.
Therefore, data is normal.

> # (b). Please test the research hypothesis at the significance level
α = 0.05.
> t.test(time, alternative = "greater", mu = 25, conf.level = 0.95)

        One Sample t-test

data:  time
t = 1.0833, df = 14, p-value = 0.1485
alternative hypothesis: true mean is greater than 25
95 percent confidence interval:
 22.99721      Inf
sample estimates:
mean of x
     28.2

> # p-value - 0.1485. Fail to reject null. Therefore we can’t say the
length of time has increased.
```

Question 5:

```
> # (1)  Regress the return of Pfizer on the returns of Exxon (with
intercept). Report the
estimated coefficients.
> one<-read.table("/Users/Yisha1/Documents/
d_logret_6stocks.txt",header=T)
> attach(one)
> lm(one$Pfizer~one$Exxon)

Call:
```

```
lm(formula = one$Pfizer ~ one$Exxon)

Coefficients:
(Intercept)    one$Exxon
  -0.005325     0.354649
# There fore Pfizer = -0.0053525 + 0.354649 * Exxon

> # (2)  Regress the return of Pfizer on the returns of Exxon (without
intercept). Report the
estimated coefficients.
> lm(one$Pfizer~one$Exxon+0)

Call:
lm(formula = one$Pfizer ~ one$Exxon + 0)

Coefficients:
one$Exxon
   0.3183
> # Therefore, Pfizer = 0.3183 * Exxon

> # (3)  Compute the correlation of Pfizer and Exxon, and test if
their correlation is zero.
> cor(one$Pfizer,one$Exxon)
[1] 0.3520965
> cor.test(one$Exxon,one$Pfizer)

        Pearson's product-moment correlation

data:  one$Exxon and one$Pfizer
t = 2.9621, df = 62, p-value = 0.004328
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1163578 0.5502798
sample estimates:
      cor
0.3520965
> # Therefore, correlation is not zero. It is 0.350965


Question 6:
> # (a) what is the correlation coefficient between these two
variables
> x<- c(2.4,1.6,2.0,2.6,1.4,1.6,2.0,2.2)
> y<- c(225,184,220,240,180,184,186,215)
> cor(x,y)
[1] 0.9129053


> # (b) write down the least squares regression equation
> summary(lm(y~x))
```

```
Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-19.5182  -1.2267  -0.7368   4.2632  14.4818 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)   104.06      18.65   5.581  0.00141 **
x              50.73       9.26   5.478  0.00155 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.29 on 6 degrees of freedom
Multiple R-squared:  0.8334,	Adjusted R-squared:  0.8056 
F-statistic: 30.01 on 1 and 6 DF,  p-value: 0.001546
> #LS regression equation:104.06+ 50.73 x
> cat("The ls regression equation is y=",lm(y~x)
$coefficients[1],"+x*",lm(y~x)$coefficients[2])
The ls regression equation is y= 104.0607 +x* 50.72874


> # (c) what is the coefficient of determination of your regression
> names(summary(lm(y~x)))
 [1] "call"          "terms"         "residuals"     "coefficients" 
"aliased"       "sigma"        
 [7] "df"            "r.squared"     "adj.r.squared" "fstatistic"   
"cov.unscaled" 
> summary(lm(y~x))$r.squared
[1] 0.8333961


> # (d) At a= 0.01n is there a significant linear relationship between
those two variables?
> anova(lm(y~x))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)   
x          1 3178.2  3178.2  30.014 0.001546 **
Residuals  6  635.3   105.9                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #There is a significant liner relationship since p-value is 0.001546
> #anova(lm(y~x))$Pr[1] is also same as summary(lm(y~x))$coeff[2,4]
[1] 0.001545633

> # (e) what is expected value when a company plans to spend 1800 usd
```

```
on investment
 > predict(lm(y~x),data.frame(x=1.8))

       1
195.3725
# company sales, y, is 195.3725 in thousand
```

Question 7:

```
> #With the rmr data set (ISwR package), plot metabolic rate versus
body weight. Fit a linear regression model to the relation. According
to the fitted model, what is the predicted metabolic rate for a body
weight of 80kg?
> library(ISwR)
> data(rmr)
> body <- rmr$body.weight
> rate <- rmr$metabolic.rate
> plot (body,rate)
> rmr.lm <- lm(rate~body)
> predict (rmr.lm,data.frame(body = 80))
       1
1375.989
> attach(rmr)
> lm(metabolic.rate~body.weight, data=rmr)

Call:
lm(formula = metabolic.rate ~ body.weight, data = rmr)

Coefficients:
(Intercept)  body.weight
     811.23         7.06
> fit <- lm(metabolic.rate~body.weight, data=rmr)
> summary(fit)

Call:
lm(formula = metabolic.rate ~ body.weight, data = rmr)

Residuals:
    Min      1Q  Median      3Q     Max
-245.74 -113.99  -32.05  104.96  484.81

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 811.2267    76.9755  10.539 2.29e-13 ***
body.weight   7.0595     0.9776   7.221 7.03e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 157.9 on 42 degrees of freedom
```

```
Multiple R-squared:  0.5539,	Adjusted R-squared:  0.5433
F-statistic: 52.15 on 1 and 42 DF,  p-value: 7.025e-09

> 811.2267 + 7.0595 * 80 # , or:
[1] 1375.987
> predict(fit, newdata=data.frame(body.weight=80))
       1
1375.989
> # answer is 1375.989
```

Probability the Mid !

1.(1)

```
> rand.vec<-rnorm(60,30,sqrt(16))
> rand.vec
 [1] 31.22665 33.82230 28.47936 29.00782 29.23694 28.68085 25.28197
 [8] 29.62076 26.35266 25.94503 20.41280 33.83555 30.18793 36.28133
[15] 26.07712 35.61231 30.29604 29.77874 29.06369 26.62158 33.33265
[22] 31.08605 34.79659 32.08481 30.72677 28.32942 20.47660 24.28966
[29] 33.03134 28.91218 27.01272 25.59961 24.28784 29.60684 34.95485
[36] 34.17065 33.53362 23.38987 27.22518 38.35401 34.46708 29.04980
[43] 34.93233 26.61658 26.99271 28.20720 27.61242 34.04999 35.90418
[50] 32.62101 25.19622 27.48257 35.46842 28.67139 32.16913 35.09071
[57] 31.69195 29.81952 34.47807 26.86646
```

1.(2)

```
> rand.mat <- matrix(rand.vec, 6, 10)
> rand.mat
         [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]
[1,] 31.22665 25.28197 30.18793 29.06369 30.72677 27.01272 33.53362
[2,] 33.82230 29.62076 36.28133 26.62158 28.32942 25.59961 23.38987
[3,] 28.47936 26.35266 26.07712 33.33265 20.47660 24.28784 27.22518
[4,] 29.00782 25.94503 35.61231 31.08605 24.28966 29.60684 38.35401
[5,] 29.23694 20.41280 30.29604 34.79659 33.03134 34.95485 34.46708
[6,] 28.68085 33.83555 29.77874 32.08481 28.91218 34.17065 29.04980
         [,8]     [,9]    [,10]
[1,] 34.93233 35.90418 32.16913
[2,] 26.61658 32.62101 35.09071
[3,] 26.99271 25.19622 31.69195
[4,] 28.20720 27.48257 29.81952
[5,] 27.61242 35.46842 34.47807
[6,] 34.04999 28.67139 26.86646
```

1.(3)

the values are (27.30204, 32.69796).

```
> LB <- qnorm(0.25, 30, sqrt(16))
> LB
[1] 27.30204
> UB <- qnorm(0.75, 30, sqrt(16))
> UB
[1] 32.69796
```

2.(1)

```
> x1<-sample(1:4,100,prob=c(0.1,0.15,0.3,0.45),replace=TRUE)
> head(x1)
[1] 4 4 3 3 2 4
```

2.(2)

```
> x2 <- numeric(100)
> for(i in 1:100) {
+tmp <- runif(1)
+if(tmp <= 0.1) x2[i] <- 1
+else if(0.1 < tmp & tmp <= 0.25) x2[i] <- 2
+else if(0.25 < tmp & tmp <= 0.55) x2[i] <- 3
+else x2[i] <- 4
+ }
> head(x2)
[1] 4 2 1 4 3 3
```

2.(3)

The probability is equal to 3<=X<=8 which is 0.7852687:

```
> ppois(8, 6) - ppois(2, 6)
[1] 0.7852687
```

3.(1)

```
> tb <- matrix(c(33, 124, 18, 94, 17, 109, 5, 60, 2, 30),
nrow=5,byrow=T)
> rownames(tb) <- c("18-24","25-29","30-34","35-39",">40")
> colnames(tb) <- c("Accidents","No Accidents")
> names(dimnames(tb)) <- c("Age","")
> tb


Age     Accidents No Accidents
  18-24        33          124
  25-29        18           94
  30-34        17          109
  35-39          5           60
  >40              2           30
```

3.(2)

```
> chisq.test(tb)

    Pearson's Chi-squared test

data:  tb
X-squared = 9.284, df = 4, p-value = 0.05438
```

The p value of the Chi-Square test is 0.05438 < 0.10, so we should reject H0(group (Accidents and No Accidents) and the Age range effects are independent), thus, based on the data, we conclude the group (Accidents and No Accidents) and the Age range effects are not independent at 0.10 level.

4(1)

```
> Nonsmokers <- c(0.97, 0.72, 1, 0.81, 0.62, 1.32, 1.24, 0.99)
> Smokers <- c(0.48, 0.71, 0.98, 0.68, 1.18, 1.36, 0.78, 1.64)
> shapiro.test(Nonsmokers)

	Shapiro-Wilk normality test

data:  Nonsmokers
W = 0.95353, p-value = 0.7467
> shapiro.test(Smokers)

	Shapiro-Wilk normality test

data:  Smokers
W = 0.95193, p-value = 0.7307
> var.test(Nonsmokers, Smokers)

	F test to compare two variances

data:  Nonsmokers and Smokers
F = 0.37872, num df = 7, denom df = 7, p-value = 0.2235
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.07582146 1.89167862
sample estimates:
ratio of variances
         0.3787213
> t.test(Smokers, Nonsmokers, alternative = "less",  var.equal = TRUE)


	Two Sample t-test

data:  Smokers and Nonsmokers
t = 0.10768, df = 14, p-value = 0.5421
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf 0.3037443
sample estimates:
mean of x mean of y
  0.97625   0.95875
```

The p value of the normality test for Nonsmokers is 0.7467 > 0.05, we can not reject H0(data is normal), thus, data for Nonsmokers follows normal at 0.05 level, and the p value of the normality test for Smokers is 0.7307 > 0.05, we can not reject H0(data is normal), thus, data for Smokers also follows normal at 0.05 level, so both of the data are normal, we can use t test, the variance test shows that the p value is 0.2235 > 0.05, so that we can not reject H0(variances are equal), thus, the two data have equal variance at 0.05 level, at last, the t test with equal variance shows the p value is 0.5421 > 0.05, so we can not reject H0(smokers have higher or equal mean plasma ascorbic acid levels than the nonsmokers), thus, we can conclude smokers do not have a significant lower mean plasma ascorbic acid levels than the nonsmokers at 0.05 level.

4.(2)

```
>df<-data.frame(level=c(Nonsmokers, Smokers),
group=rep(c("Nonsmokers","Smokers"), each = 8))
> anova(lm(level ~ group, data = df))
Analysis of Variance Table

Response: level
                 Df  Sum Sq  Mean Sq   F value Pr(>F)
group      1 0.00122 0.001225  0.0116 0.9158
Residuals 14 1.47907 0.105648
```

The F value of the one-way ANOVA is 0.0116 with the p value to be 0.9158 > 0.05, so we can not reject the H0(two means are equal). Based on the data, we can conclude that there is no significant difference between plasma ascorbic acid levels of smokers and nonsmokers at 0.05 level.

5.(1)

```
>RadialTires<-c(4.2, 4.7, 6.6, 7, 6.7, 4.5, 5.7, 6, 7.4, 4.9, 6.2,
5.2)
>BeltedTires<-c(4.1, 4.9, 6.2, 6.9, 6.8, 4.3, 5.7, 5.8, 6.9, 4.7, 6,
4.9)
>shapiro.test(RadialTires)

    Shapiro-Wilk normality test

data:  RadialTires
W = 0.95549, p-value = 0.7181
>t.test(RadialTires, mu = 6, alternative = "greater")

    One Sample t-test

data:  RadialTires
t = -0.79268, df = 11, p-value = 0.7776
alternative hypothesis: true mean is greater than 6
95 percent confidence interval:
 5.210815      Inf
sample estimates:
mean of x
 5.758333
```

The p value of the normality test for radial tires is 0.7181 > 0.05, we can not reject H0(data is normal), thus, data for Nonsmokers follows normal at 0.05 level, and the p value of the one sample t test is 0.7776 > 0.05, so we can not reject H0(the mean gasoline consumption for the cars equipped with radial tires is less or equal than 6.0), thus we can conclude that the mean gasoline consumption for the cars equipped with radial tires is not significantly greater than 6.0 at 0.05 level.

5.(2)

```
>Diff<-RadialTires - BeltedTires
>shapiro.test(Diff)

    Shapiro-Wilk normality test

data:  Diff
W = 0.96933, p-value = 0.9036
>t.test(Diff, mu = 0, alternative = "less")

    One Sample t-test

data:  Diff
t = 2.7768, df = 11, p-value = 0.991
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
      -Inf 0.2607344
sample estimates:
mean of x
0.1583333
```

The p value of the normality test for difference is 0.9036 > 0.05, we can not reject H0(data is normal), thus, data of difference follows normal at 0.05 level, using t test we can find the p value is 0.991 > 0.05, so we can not reject H0(radial tires have more ore equal gasoline consumption than belted Tires), thus, we can conclude that cars equipped with radial tires do not give better fuel economy than those equipped with belted tires at 0.05 level.

6.(1)

with intercept, the intercept estiamted is -0.004856159 and the slope estimated is 0.457523248.

```
>stock<-read.table("d_logret_6stocks.txt", header = TRUE)
>mod<-lm( GenMotor ~ Citigroup, data = stock)
>summary(mod)$coefficients
                Estimate  Std. Error   t value   Pr(>|t|)
(Intercept) -0.004856159 0.005368092 -0.904634 0.36916119
Citigroup    0.457523248 0.174196061  2.626484 0.01085527
>summary(mod)$coefficients[,1]
 (Intercept)    Citigroup
-0.004856159  0.457523248
```

without intercept, the slope estimated is 0.4527372.

```
>mod2 <- lm( GenMotor ~ Citigroup - 1, data = stock)
>summary(mod2)$coefficients
           Estimate Std. Error  t value   Pr(>|t|)
Citigroup 0.4527372  0.1738645 2.603965 0.01147972
>summary(mod2)$coefficients[,1]
[1] 0.4527372
```

6.(2)

```
>anova(mod)
Analysis of Variance Table
```

```
Response: GenMotor
          Df   Sum Sq   Mean Sq F value  Pr(>F)
Citigroup  1 0.012711 0.0127107  6.8984 0.01086 *
Residuals 62 0.114238 0.0018425
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>anova(mod2)
Analysis of Variance Table

Response: GenMotor
          Df   Sum Sq   Mean Sq F value  Pr(>F)
Citigroup  1 0.012458 0.0124576  6.7806 0.01148 *
Residuals 63 0.115746 0.0018372
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For both types of regressions in (1), the ANOVA tables are shown above, we can find that the p values are both lower than 0.05, so at 0.05 level, the regression effects are significant.

6.(3)

```
>cor(stock$GenMotor, stock$Citigroup)
[1] 0.3164246
>cor.test(stock$GenMotor, stock$Citigroup)

    Pearson's product-moment correlation

data:  stock$GenMotor and stock$Citigroup
t = 2.6265, df = 62, p-value = 0.01086
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.07657099 0.52165916
sample estimates:
      cor
0.3164246
```

The correlation of General Motor and Citigroup is 0.3164246, and the p value of the correlation test is 0.01086 < 0.05, so we should reject H0(correlation is 0), thus, based on the data, we can conclude that there is significant correlation between General Motor and Citigroup at 0.05 level.

6.(4)

```
>n <- nrow(stock)
>x <- sum(stock$Citigroup > stock$Pfizer)
>n
[1] 64
>x
[1] 37
>prop.test(x, n, 0.6)

    1-sample proportions test with continuity correction

data:  x out of n, null probability 0.6
```

```
X-squared = 0.052734, df = 1, p-value = 0.8184
alternative hypothesis: true p is not equal to 0.6
95 percent confidence interval:
 0.4484671 0.6983808
sample estimates:
       p
0.578125
>binom.test(x, n, 0.6)

    Exact binomial test

data:  x and n
number of successes = 37, number of trials = 64, p-value = 0.7988
alternative hypothesis: true probability of success is not equal to
0.6
95 percent confidence interval:
 0.4481586 0.7006192
sample estimates:
probability of success
              0.578125
```

As a rule of thumb, the approximation is satisfactory when the expected numbers of “successes” and “failures” are both larger than 5. Both of prop.test and binom.test show p values(0.8184 and 0.7988) are larger than 0.05, so we can not reject H0(the proportion is 0.6), thus, based on the data, we can conclude that the proportion of returns of Citigroup greater than Pfizer is 0.6 at 0.05 level.