Stephen Hanna 109097796

HW5 AMS 394

Problem (1):

(1)

```
> stocks <-
read.table("http://www.ams.sunysb.edu/~xing/statfinbook/_BookData/Chap03/d_logret_6stocks.txt",
header=T)

> fit <- lm(stocks[,2] ~ stocks[,6])

> summary(fit)


Call:

lm(formula = stocks[, 2] ~ stocks[, 6])


Residuals:

      Min        1Q    Median        3Q       Max

-0.049930 -0.013003 -0.000505  0.017353  0.049231


Coefficients:

             Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.005325   0.002756  -1.932  0.05794 .

stocks[, 6]  0.354649   0.119729   2.962  0.00433 **

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.02178 on 62 degrees of freedom

Multiple R-squared:  0.124,     Adjusted R-squared:  0.1098

F-statistic: 8.774 on 1 and 62 DF,  p-value: 0.004328
```

Response: The coefficient of correlation is .354649 and the coefficient of the intercept is -.005325.

(2)

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

Response: stocks[, 2]
            Df    Sum Sq   Mean Sq F value   Pr(>F)
stocks[, 6]  1 0.0041609 0.0041609   8.774 0.004328 **
Residuals   62 0.0294022 0.0004742
```

Response: The probability is lower than .01, thus we can reject the null hypothesis and conclude that the regression effects are significant with over 99% confidence.

(3)

```
> Pfizer = stocks[,2]
> Exxon = stocks[,6]
> Citigroup = stocks[,4]
> data = data.frame(cbind(Pfizer,Exxon,Citigroup))
> stacked = stack(data)
> anova(lm(values ~ ind, stacked))
Analysis of Variance Table

Response: values
           Df   Sum Sq    Mean Sq F value Pr(>F)
ind         2 0.001934 0.00096712  1.4351 0.2407
Residuals 189 0.127366 0.00067390
```

(4)

```
	1-sample proportions test with continuity correction

data:  sum(Citigroup > 0) out of 64, null probability 0.5
X-squared = 1.2656, df = 1, p-value = 0.2606
```

alternative hypothesis: true p is not equal to 0.5

95 percent confidence interval:

 0.4484671 0.6983808

sample estimates:

       p

0.578125

Response: The p value is greater than .05, so we cannot reject the null hypothesis that the proportion of Citigroup's returns that are greater than 0 is not significantly greater than .5.

Problem (2):

(1)

```
> data(juul)

> juul <- juul[,4:5]

> juul <- juul[complete.cases(juul),]

> one <- juul[juul[,"tanner"] == 1,]

> two <- juul[juul[,"tanner"] == 2,]

> three <- juul[juul[,"tanner"] == 3,]

> four <- juul[juul[,"tanner"] == 4,]

> five <- juul[juul[,"tanner"] == 5,]

> stacked = do.call("rbind", list(one,two,three,four,five))

> stacked$tanner <- as.factor(stacked$tanner)

> anova(lm(stacked[,1] ~ stacked[,2]))

Analysis of Variance Table


Response: stacked[, 1]

             Df   Sum Sq Mean Sq F value    Pr(>F)

stacked[, 2]   4 12696217 3174054  228.35 < 2.2e-16 ***

Residuals    787 10939116   13900

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Response: Since the p value is less than .01, we can say with more than 99% confidence that the five levels of tanner give significantly different results.

(2)

```
> mean(one[,1])

[1] 207.4727

> mean(two[,1])

[1] 352.6714

> mean(three[,1])

[1] 483.2222

> mean(four[,1])

[1] 513.0172

> mean(five[,1])

[1] 465.3344
```

(3)

```
> one <- one[,1]

> two <- two[,1]

> three <- three[,1]

> four <- four[,1]

> five <- five[,1]

> shapiro.test(one)


        Shapiro-Wilk normality test


data:  one

W = 0.96947, p-value = 3.764e-06


> shapiro.test(two)


        Shapiro-Wilk normality test
```

data: two

W = 0.9606, p-value = 0.02704

> shapiro.test(three)

Shapiro-Wilk normality test

data: three

W = 0.96348, p-value = 0.1657

> shapiro.test(four)

Shapiro-Wilk normality test

data: four

W = 0.94686, p-value = 0.01309

> shapiro.test(five)

Shapiro-Wilk normality test

data: five

W = 0.97828, p-value = 0.0001284

Response: Only one tanner level has normally distributed data, so non parametric tests will be used to compare all pairs of values.

> wilcox.test(one,two)

Wilcoxon rank sum test with continuity correction

```
data:  one and two
W = 3550.5, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> wilcox.test(one,three)

	Wilcoxon rank sum test with continuity correction

data:  one and three
W = 712, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> wilcox.test(one,four)

	Wilcoxon rank sum test with continuity correction

data:  one and four
W = 300.5, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> wilcox.test(one,five)

	Wilcoxon rank sum test with continuity correction

data:  one and five
W = 5006.5, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Response: Tanner level one is significantly different from all other tanner levels, demonstrated by every pair wilcox test with one showing a p value less than .01.

```
> wilcox.test(two,three)

	Wilcoxon rank sum test with continuity correction

data:  two and three
W = 783, p-value = 5.733e-06
alternative hypothesis: true location shift is not equal to 0

> wilcox.test(two,four)

	Wilcoxon rank sum test with continuity correction

data:  two and four
W = 693, p-value = 1.579e-10
alternative hypothesis: true location shift is not equal to 0

> wilcox.test(two,five)

	Wilcoxon rank sum test with continuity correction

data:  two and five
W = 5702, p-value = 7.593e-10
alternative hypothesis: true location shift is not equal to 0
```

Response: Tanner level two is significantly different from all other tanner levels, demonstrated by every pair wilcox test with one showing a p value less than .01.

> wilcox.test(three,four)

Wilcoxon rank sum test with continuity correction

data: three and four

W = 1084, p-value = 0.1426

alternative hypothesis: true location shift is not equal to 0

> wilcox.test(three,five)

Wilcoxon rank sum test with continuity correction

data: three and five

W = 7332.5, p-value = 0.5295

alternative hypothesis: true location shift is not equal to 0

Response: Tanner level three is NOT significantly different from tanner levels four and five, demonstrated by those pairs wilcox tests showing a p value greater than .05.

> wilcox.test(four,five)

Wilcoxon rank sum test with continuity correction

data: four and five

W = 10996, p-value = 0.005231

alternative hypothesis: true location shift is not equal to 0

Response: Tanner level four is significantly different than tanner level five, as demonstrated by the p value being lower than .01.

Problem (3):

```
> data(survey)

> tbl = table(survey$Smoke, survey$Exer)

> chisq.test(tbl)


        Pearson's Chi-squared test


data:  tbl

X-squared = 5.4885, df = 6, p-value = 0.4828


Warning message:

In chisq.test(tbl) : Chi-squared approximation may be incorrect
```

Response: Since the p value is greater than .05, we cannot reject the null hypothesis that the smoking habit is independent of the exercise level.