Homework4

Yanyu Zheng yz2690

February 18, 2016

1. Chapter 5, problem 23

```
library("Sleuth3")
attach(ex0523)
anova(lm(Oxygen ~ Bone, ex0523))
```

```
## Analysis of Variance Table
##
## Response: Oxygen
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Bone      11 6.0675 0.55159  7.4268 9.73e-07 ***
## Residuals 40 2.9708 0.07427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The pValue is 9.73e-07, which means there's evidence that the means are dif- ferent for the different bones.

2. Chapter 5, problem 25

```
attach(ex0525)
anova(lm(log(Income2005) ~ Educ, ex0525))
```

```
## Analysis of Variance Table
##
## Response: log(Income2005)
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## Educ         4  217.65  54.413   62.87 < 2.2e-16 ***
## Residuals 2579 2232.12   0.865
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The pValue is less than 2.2e-16 for log scale, which means the evidence is very strong for the diffenrece in income in different groups.

```
## 12   13-15    16    >16
t.test(ex0525[Educ=="12","Income2005"],ex0525[Educ=="13-15","Income2005"],alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  ex0525[Educ == "12", "Income2005"] and ex0525[Educ == "13-15", "Income2005"]
## t = -4.9487, df = 1232.7, p-value = 4.252e-07
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -5346.347
## sample estimates:
## mean of x mean of y
##  36864.90  44875.96
```

```
t.test(ex0525[Educ=="13-15","Income2005"],ex0525[Educ=="16","Income2005"],alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  ex0525[Educ == "13-15", "Income2005"] and ex0525[Educ == "16", "Income2005"]
## t = -7.2686, df = 548.25, p-value = 6.29e-13
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -19426.6
## sample estimates:
## mean of x mean of y
##  44875.96  69996.97
```

```
t.test(ex0525[Educ=="16","Income2005"],ex0525[Educ==">16","Income2005"],alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  ex0525[Educ == "16", "Income2005"] and ex0525[Educ == ">16", "Income2005"]
## t = -1.4752, df = 770.26, p-value = 0.07029
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 798.1295
## sample estimates:
## mean of x mean of y
##  69996.97  76855.46
```

12 & 13-15 5346.347
13-15 & 16 19426.6
16 & >16 798.1295

3. Chapter 6, problem 12

```
attach(case0601)
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data3 = case0601 %>%
  filter(Handicap!="None") %>%
  group_by(Handicap) %>%
  summarise(mean = mean(Score),sd = sd(Score),n=n())
attach(data3)
```

```
## The following object is masked from case0601:
##
##     Handicap
```

```
coef = c(1/3,1/3,-1,1/3)
g = sum(coef*mean)
pooled = sqrt(sum(sd^2*(n-1))/sum(n-1))
SE = pooled*sqrt(sum(coef^2/n))
tStatistic = g/SE
## Planned comparison
pValue = 1-pt(tStatistic,52)
## Unplanned comparison
qTukey = qtukey(0.95,4,52)/sqrt(2)
```

Test statistic is 2.4063759. The pValue for the planned test is 0.0098502. The threshold for Tukey-Kramer test is 2.6540988, which is larger than our test Statistic. So pValue > 0.05.

4. Chapter 6, problem 15

- a

```
mean = c(30.20,28.80,26.20,31.10,30.20)
sd = c(3.82,5.26,4.66,4.91,3.53)
n = rep(9,5)
pooled = sqrt(sum((n-1)*sd^2)/sum(n-1))
```

The pooled estimate of the standard deviation is 4.484297

*b

-1/3, 1/2, 1/2,-1/3, -1/3

*c

```
coeff = c(1/3, -1/2, -1/2, 1/3, 1/3)
g = sum(coeff * mean)
sdG = pooled*sqrt(sum(coeff^2/n))
tStatistic = g/sdG
q = qt(0.975,40)
CI = c(g-sdG*q, g+sdG*q)
```

tStatistic is 2.1985622, qt(0.975,40) is 2.0210754, 95% confidence interval is 0.2421858, 5.7578142

5. Chapter 6, problem 16

```
LSD = qt(0.975,30)
# F value is large, no need for protected LSD
protectedLSD = qt(0.975,30)
tukey = qtukey(0.95,6,30)/sqrt(2)
bonferroni = qt(1-0.05/30,30)
sheffe = sqrt(5*qf(0.95,5,30))
```

LSD : 2.0422725
protectedLSD : 2.0422725
tukey : 3.0415943
bonferroni : 3.1888059
sheffe : 3.5591815

6. Chapter 6, problem 23

```
attach(ex1420)
```

```
## The following object is masked from ex0525:
##
##     Subject
```

```
TukeyHSD(aov(WtLoss24~Diet, ex1420))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = WtLoss24 ~ Diet, data = ex1420)
##
## $Diet
##                                       diff       lwr        upr     p adj
## Low-Fat-Low-Carbohydrate       -2.1828035 -4.224771 -0.1408361 0.0329364
## Mediterranean-Low-Carbohydrate -0.8849083 -2.932082  1.1622656 0.5656813
## Mediterranean-Low-Fat           1.2978952 -0.697418  3.2932084 0.2771180
```

From the result, we can see that the low-fat and the low-carbohydrate group is significantly different. 95% CI is (-4.224771,-0.1408361)