Advanced Data Analysis -HW2

Name:Rui Liu UNI:rl2987

1

```r
library(openxlsx)

data.1 <- read.xlsx("~/Desktop/2019 Spring/Advanced Data Analysis/hw/HW2/HW2(Diets).xlsx")
```

a)

```r
model.1.1 <- lm(Response~as.factor(Diet),data=data.1)
anova(model.1.1)
```

```
## Analysis of Variance Table
##
## Response: Response
##                 Df  Sum Sq  Mean Sq F value  Pr(>F)
## as.factor(Diet)  3 0.52652 0.175506  4.4891 0.01185 *
## Residuals       25 0.97740 0.039096
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs H_1:Not H_0

From the anova table above, we have `p-value = 0.012 < 0.05`, so we reject H_0 under significance level $\gamma = 0.05$

In conclusion, there are at least two means that are not equal.

b)

```r
t.test(Response~Diet, data=data.1[1:15,],conf.level=0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  Response by Diet
## t = 1.1625, df = 9.1239, p-value = 0.2745
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1325412  0.4139698
## sample estimates:
## mean in group 1 mean in group 2
##        3.745714        3.605000
```

Our test : $H_0 : \mu_1 - \mu_2 = 0$

According to the test, we have 95% confidence interval for $\mu 1 - \mu 2$ is (-0.13,0.41)

c) & d)

```r
mu_1 <- mean(data.1$Response[data.1$Diet==1])
mu_2 <- mean(data.1$Response[data.1$Diet==2])
mu_3 <- mean(data.1$Response[data.1$Diet==3])
mu_4 <- mean(data.1$Response[data.1$Diet==4])

n1 <- sum(data.1$Diet==1)
n2 <- sum(data.1$Diet==2)
n3 <- sum(data.1$Diet==3)
n4 <- sum(data.1$Diet==4)

L <- (mu_1+mu_2-mu_3-mu_4)/2
L
```

```
## [1] -0.08505952
```

```r
# From the anova table in part a), we know that MSE is 0.0391
SE_L <- sqrt(0.0391*sum(0.25/n1+0.25/n2+0.25/n3+0.25/n4))

t <- qt(1-0.05/2,df=29-4)

# hyphothesis test
abs(L) > t*SE_L
```

```
## [1] FALSE
```

```r
# confidence interval
L-SE_L*t
```

```
## [1] -0.2373727
```

```r
L+SE_L*t
```

```
## [1] 0.06725369
```

part c):

The confidence interval for L is : (-0.237,0.067)

part d):

Our test: $H_0 : L = \frac{\mu_1+\mu_2-\mu_3-\mu_4}{2} = 0$ vs. $H_1 : L \neq 0$

$\hat{L} = \frac{\hat{\mu_1}+\hat{\mu_2}-\hat{\mu_3}-\hat{\mu_4}}{2} = -0.08$

Since $|\hat{L}| < t_{\gamma/2}(df = n - k = 29 - 4)$, we fail to reject H_0

2

a)

```r
# Define dummy variables
D <- function(x,i) {
  temp <- rep(NA,length(x))

  for(j in 1:length(x)) {
    if (x[j] == i) {temp[j] <- 1}
    else if(x[j] == 4) {temp[j] <- -1}
```

```r
    else {temp[j] <- 0}
  }

  return(temp)
}

D1 <- D(data.1$Diet,1)
D2 <- D(data.1$Diet,2)
D3 <- D(data.1$Diet,3)

model.2.1 <- lm(data.1$Response~D1+D2+D3)
summary(model.2.1)
```

```
##
## Call:
## lm(formula = data.1$Response ~ D1 + D2 + D3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3857 -0.0950 -0.0525  0.1250  0.4443
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.71789    0.03698 100.549   <2e-16 ***
## D1           0.02783    0.06450   0.431   0.6698
## D2          -0.11289    0.06173  -1.829   0.0794 .
## D3          -0.11955    0.06801  -1.758   0.0910 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1977 on 25 degrees of freedom
## Multiple R-squared:  0.3501, Adjusted R-squared:  0.2721
## F-statistic: 4.489 on 3 and 25 DF,  p-value: 0.01185
```

From the defination and the table above, we have:

$$\hat{\mu} = 3.72$$

$$\hat{\alpha_1} = 0.028$$

$$\hat{\alpha_2} = -0.113$$

$$\hat{\alpha_3} = -0.120$$

$$\hat{\alpha_4} = -(\hat{\alpha_1} + \hat{\alpha_2} + \hat{\alpha_3}) = -(0.028 - 0.113 - 0.120) = 0.205$$

b)

```r
model.2.2 <- lm(data.1$Response~1)
anova(model.2.2,model.2.1)
```

```
## Analysis of Variance Table
##
## Model 1: data.1$Response ~ 1
## Model 2: data.1$Response ~ D1 + D2 + D3
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
```

```
## 1     28 1.5039
## 2     25 0.9774  3   0.52652 4.4891 0.01185 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our test: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ vs. H_α: Not H_0

From the anova table,we have `p-value=0.012 < 0.05`, so reject H_0

The `p-value` is exactly the same as the value in Problem 1 and we reached the same conclusion.

3

a)

Source	df	SS	MS	F
feed	2	23.43	11.715	17.564
error	7	4.67	0.667	
total	9	28.10		

b)

```
p.value <- 1-pf(17.69,df1=2,df2=7)
p.value
```

```
## [1] 0.001831386
```

Our test: $H_0 : \mu_1 = \mu_2 = \mu_3$ vs. H_α: Not H_0

From the calculation above, we know that `p-value = 0.0018 < 0.05`, so reject H_0

Therefore, we know that different chichen feeds have different effect.