Name: Raj Shah

PSID: 1499521

**Homework 4**

Prob 1:

Code:

*library(readxl)*

*LungCapData <- read\_excel("C:/Users/Raj Shah/Desktop/Studies UH/MATH 6359 Stastical Computing/Homework/Answers/HW4/LungCapData.xls")*

*attach(LungCapData)*

*lm\_1<-lm(LungCap~Gender,data=LungCapData)*

*summary(lm\_1)*

Output

Call:

lm(formula = LungCap ~ Gender, data = LungCapData)

Residuals:

Min 1Q Median 3Q Max

-7.1343 -1.7557 0.1657 1.9657 6.3657

Coefficients:

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

(Intercept) 7.4057 0.1387 53.377 < 2e-16 \*\*\*

Gendermale 0.9036 0.1950 4.634 4.26e-06 \*\*\*

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.625 on 723 degrees of freedom

Multiple R-squared: 0.02884, Adjusted R-squared: 0.0275

F-statistic: 21.47 on 1 and 723 DF, p-value: 4.262e-06

Code:

*lm\_2<-lm(LungCap~Age+I(Age^2),data=LungCapData)*

*summary(lm\_2)*

Output

Call:

lm(formula = LungCap ~ Age + I(Age^2), data = LungCapData)

Residuals:

Min 1Q Median 3Q Max

-4.8446 -1.0150 -0.0212 1.0442 4.1942

Coefficients:

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

(Intercept) 0.610275 0.433197 1.409 0.159

Age 0.647707 0.076549 8.461 <2e-16 \*\*\*

I(Age^2) -0.004354 0.003184 -1.367 0.172

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.525 on 722 degrees of freedom

Multiple R-squared: 0.6727, Adjusted R-squared: 0.6718

F-statistic: 742 on 2 and 722 DF, p-value: < 2.2e-16

Code:

*lm\_3<-lm(LungCap~Age\*Gender,data=LungCapData)*

*summary(lm\_3)*

Output

Call:

lm(formula = LungCap ~ Age \* Gender, data = LungCapData)

Residuals:

Min 1Q Median 3Q Max

-4.5454 -0.9223 0.0288 0.9777 4.3432

Coefficients:

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

(Intercept) 0.593760 0.250953 2.366 0.01824 \*

Age 0.547160 0.019212 28.480 < 2e-16 \*\*\*

Gendermale 0.998278 0.346172 2.884 0.00405 \*\*

Age:Gendermale 0.003119 0.026700 0.117 0.90705

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.437 on 721 degrees of freedom

Multiple R-squared: 0.7098, Adjusted R-squared: 0.7086

F-statistic: 587.8 on 3 and 721 DF, p-value: < 2.2e-16

Assume the following model:

yi=Beta0+gamma1\*zi + ei; ei~N(0,sig^2)

There is a significant impact of Factor Variable (“Gender”) on the response variable (“Lung Capacity”) as seen from the very low p-value(4.26e-06). The “Male” is the reference group that has significance on the Lung capacity.

Assume the following model:

yi = Beta0 + Beta1\*xi + Beta2\*(xi)^2+ ei ; ei ~ N(0; sig^2)

The hypotheses to test for quadratic term:

H0 : Beta2 = 0; vs Ha : Beta2 != 0

There is no quadratic relationship between the response variable (“Lung Capacity”) and continuous variable (“Age”) indicated by a higher p-value. we fail to reject H0 due to p-value of 0.172 for quadratic term, hence

claiming no quadratic relationship between Lung Capacity and Age.

Assume the following model:

yi=Beta0+Beta1\*(xi)+gamma1\*z+phi1\*xi\*zi

Hypotheses to test for interaction:

H0 : phi1 = 0; vs; Ha : phi1 != 0

We fail to reject H0 due to p-value of 0.907 for interaction term. There is no significance of the interaction term.

Prob2:

Code:

*my\_f\_test<-function(x,z){*

*ssdb=0*

*ssdw=0*

*xbar=mean(x)*

*for (c in (1:length(z.levels))){*

*xbar\_i=mean(x[z==z.levels[c]])*

*ssdb=round(ssdb+sum((length(x[z==z.levels[c]]))\*((xbar\_i-xbar)^2)),2)*

*for (i in (1:length(x[z==z.levels[c]]))){*

*x\_ij=x[z==z.levels[c]][i]*

*ssdw=round(ssdw+sum((x\_ij-xbar\_i)^2),2)*

*}*

*}*

*k=length(z.levels)*

*n=length(x)*

*dofb=k-1*

*dofw=n-k*

*msb=round((ssdb/dofb),2)*

*msw=round((ssdw/dofw),2)*

*f=round(msb/msw,4)*

*p\_val=round(pf(q = f,df1=dofb,df2 = dofw,lower.tail = FALSE),4)*

*result<-matrix(c(dofb,dofw,ssdb,ssdw,f,p\_val),ncol=6)*

*colnames(result)<-c("DofB","DofW","SSD\_B","SSD\_W","F Val","Pr(>F)")*

*result<-as.table(result)*

*print(result)*

*}*

Output

Call 1:

> x<-runif(n=50,min=100,max=200)

> z.levels<-c("a","b","c")

> z<-as.factor(sample(z.levels,size=50,replace=T))

> my\_f\_test(x,z)

DofB DofW SSD\_B SSD\_W F Val

A 2.0000 47.0000 2982.7100 38535.6100 1.8189

Pr(>F)

A 0.1734

> anova(lm(x~z))

Analysis of Variance Table

Response: x

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

z 2 2983 1491.36 1.8189 0.1734

Residuals 47 38536 819.91

Call 2:

> x<-runif(n=50,min=100,max=200)

> z.levels <- letters[1:8]

> z<-as.factor(sample(z.levels,size=50,replace=T))

> my\_f\_test(x,z)

DofB DofW SSD\_B SSD\_W F Val

A 7.0000 42.0000 4291.5500 22053.3400 1.1676

Pr(>F)

A 0.3416

> anova(lm(x~z))

Analysis of Variance Table

Response: x

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

z 7 4291.6 613.08 1.1676 0.3416

Residuals 42 22053.3 525.08

Prob 3:

Code:

*library(ISwR)*

*data("lung")*

*attach(lung)*

*anova(lm(volume~method))*

Output

Analysis of Variance Table

Response: volume

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

method 2 1.0811 0.54056 2.6893 0.1004

Residuals 15 3.0150 0.20100

As we see from the P-value(0.1004) which is higher than 0.05, the method variable has no strong significance on the lung volume.

I don’t think it makes sense to compare methods without controlling the patient variable to measure the Lung Volume. As it can depend of a lot of factors like lifestyle, respiratory problems etc., which could have significant impact on the lung volume.

*Code:*

*anova(lm(volume~method+subject))*

Output

Analysis of Variance Table

Response: volume

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

method 2 1.08111 0.54056 6.4953 0.01557 \*

subject 5 2.18278 0.43656 5.2457 0.01271 \*

Residuals 10 0.83222 0.08322

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Here the p-value has reduced in comparison with one – way anova. We see that the variables method and subject together have a better significance on lung volume than method variable alone. Also the significance is not strong as indicated by the no. of ‘\*’.

Prob 4: Description of Data

**Active**

Test group receiving active training; these children had their walking and placing reflexes trained for four three-minute sessions that took place every day from their second to their eighth week of life.

**Passive**

Passive training group; these children received the same types of social and gross motor stimulation, but did not have their specific walking and placing reflexes trained.

**None**

Group receiving no training,These children had no special training, but were tested along with the children who underwent active or passive training.

**ctr.8w**

Eighth-week controls; these children had no training and were only tested at the age of 8 weeks.

Code:

*library(ISwR)*

*data("zelazo")*

*attach(zelazo)*

*zel <- unlist(zelazo,use.names=FALSE)*

*baby<-c(rep("Active",length(1:6)),rep("Passive",length(7:12)),rep("None",length(13:17)),rep("Ctr.8w",length(18:23)))*

*f\_test<-factor(baby)*

*anova(lm(zel~f\_test))*

*summary(lm(zel~f\_test))*

Analysis of Variance Table

Response: zel

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

f\_test 3 16.428 5.4759 2.4749 0.0927 .

Residuals 19 42.040 2.2126

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:

lm(formula = zel ~ f\_test)

Residuals:

Min 1Q Median 3Q Max

-2.4500 -0.9167 -0.3750 0.5458 3.6250

Coefficients:

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

(Intercept) 10.1250 0.6073 16.673 8.45e-13 \*\*\*

f\_testCtr.8w 2.3333 0.8588 2.717 0.0137 \*

f\_testNone 1.3250 0.9007 1.471 0.1576

f\_testPassive 1.2500 0.8588 1.456 0.1619

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.487 on 19 degrees of freedom

Multiple R-squared: 0.281, Adjusted R-squared: 0.1674

F-statistic: 2.475 on 3 and 19 DF, p-value: 0.0927

From the f-test results we see due to the high p-value we fail to reject the null hypothesis of equal means for all training methods. There is no significant difference across the training groups.

*pairwise.t.test(zel,f\_test,p.adjust.method = "bonferroni")*

Pairwise comparisons using t tests with pooled SD

data: zel and f\_test

Active Ctr.8w None

Ctr.8w 0.082 - -

None 0.946 1.000 -

Passive 0.971 1.000 1.000

P value adjustment method: bonferroni

The difference of means between the group is not significantly different as seen from the high p-value.

The group Active and 8 week Control have the smallest adjusted p-value of 0.082.