Homework 7

**1. a)**

> x=scan()

1: 23.1 32.8 31.8 32.0 30.4 24.0 39.5 24.2 52.5 37.9 30.5 25.1 12.4 35.1 31.5 21.1

17:

Read 16 items

> x

[1] 23.1 32.8 31.8 32.0 30.4 24.0 39.5 24.2 52.5 37.9 30.5 25.1 12.4 35.1 31.5

[16] 21.1

> y=scan()

1: 10.5 16.7 18.2 17.0 16.3 10.5 23.1 12.4 24.9 22.8 14.1 12.9 8.8 17.4 14.9 10.5

17:

Read 16 items

> y

[1] 10.5 16.7 18.2 17.0 16.3 10.5 23.1 12.4 24.9 22.8 14.1 12.9 8.8 17.4 14.9

[16] 10.5

> model=lm(y~x)

> model

Call:

lm(formula = y ~ x)

Coefficients:

(Intercept) x

0.5180 0.5016

**Thus, the fitted simple linear model is y= 0.5016x+0.5180.**

To test

**

> anova(model)

Analysis of Variance Table

Response: y

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

x 1 309.927 309.927 101.54 8.496e-08 \*\*\*

Residuals 14 42.731 3.052

---

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

**Since p-value<α, then reject Ho. Thus, there is a strong relationship between chemical yield(y) and another chemical(x).**

**b)** > confint(model,level=0.9)

5 % 95 %

(Intercept) -2.2427613 3.2788045

x 0.4139049 0.5892431

**Thus, 90% confidence interval for the slope is (0.4139049, 0.5892431).**

**c)** > model2=lm(y~x-1)

> model2

Call:

lm(formula = y ~ x - 1)

Coefficients:

x

0.5174

**Thus, the fitted simple linear model2 is y= 0.5174x.**

**d)** > summary(model)$r.squared

[1] 0.8788327

> summary(model2)$r.squared

[1] 0.9899623

**Since , then the coefficient of determination for the model with intercept is smaller than that with no intercept. Thus, this implies that model2 is appropriate for the representation of the given data.**

**2. a)** > library(MASS)

> Cars93

> dim(Cars93)

[1] 93 27

**Thus, there are 27 variables in included in the data set.**

**b)**

> y=Cars93$MPG.city

> x1=Cars93$EngineSize

> x2=Cars93$Weight

> x3=Cars93$Passengers

> x4=Cars93$Price

> model=lm(y~x1+x2+x3+x4)

> model

Call:

lm(formula = y ~ x1 + x2 + x3 + x4)

Coefficients:

(Intercept) x1 x2 x3 x4

46.389413 0.196119 -0.008207 0.269622 -0.035804

**c**) > summary(model)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4)

Residuals:

Min 1Q Median 3Q Max

-6.1207 -1.9098 0.0522 1.1294 13.9580

Coefficients:

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

(Intercept) 46.389413 2.097516 22.116 < 2e-16 \*\*\*

x1 0.196119 0.588880 0.333 0.740

x2 -0.008207 0.001343 -6.111 2.63e-08 \*\*\*

x3 0.269622 0.424951 0.634 0.527

x4 -0.035804 0.049179 -0.728 0.469

---

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

Residual standard error: 3.06 on 88 degrees of freedom

Multiple R-squared: 0.7165, Adjusted R-squared: 0.7036

F-statistic: 55.59 on 4 and 88 DF, p-value: < 2.2e-16

**Since p-value< 2.2e-16<2.63e-08, then reject Ho. Thus, Engine Size(x1), Passengers(x3), Price(x4) are marked as statistically significant.**

**d)** > library(MASS)

> step =stepAIC(model, direction="both")

Start: AIC=212.87

y ~ x1 + x2 + x3 + x4

Df Sum of Sq RSS AIC

- x1 1 1.04 824.89 210.99

- x3 1 3.77 827.62 211.29

- x4 1 4.96 828.82 211.43

<none> 823.85 212.87

- x2 1 349.67 1173.52 243.77

Step: AIC=210.99

y ~ x2 + x3 + x4

Df Sum of Sq RSS AIC

- x3 1 3.20 828.10 209.35

- x4 1 4.84 829.74 209.53

<none> 824.89 210.99

+ x1 1 1.04 823.85 212.87

- x2 1 627.12 1452.01 261.57

Step: AIC=209.35

y ~ x2 + x4

Df Sum of Sq RSS AIC

- x4 1 11.96 840.05 208.68

<none> 828.10 209.35

+ x3 1 3.20 824.89 210.99

+ x1 1 0.47 827.62 211.29

- x2 1 1050.34 1878.44 283.52

Step: AIC=208.68

y ~ x2

Df Sum of Sq RSS AIC

<none> 840.05 208.68

+ x4 1 11.96 828.10 209.35

+ x3 1 10.31 829.74 209.53

+ x1 1 0.06 839.99 210.67

- x2 1 2065.52 2905.57 322.09

**Thus, the model select by AIC criteria is y~x2. That is**

> model2

Call:

lm(formula = y ~ x2)

Coefficients:

(Intercept) x2

47.048353 -0.008032

**3. a)**

> library(UsingR)

> data(npdb)

> attach(npdb)

The following objects are masked from npdb (position 3):

age, amount, field, ID, state, year

> boxplot(amount~year,data=npdb)



> boxplot(log(amount)~year,data=npdb)



**b**) > data1=subset(npdb, year==c("2000","2001","2002"),all=T)

> aov(amount~year,data=data1)

Call:

aov(formula = amount ~ year, data = data1)

Terms:

year Residuals

Sum of Squares 3.356470e+12 9.819813e+14

Deg. of Freedom 1 2288

Residual standard error: 655124.1

Estimated effects may be unbalanced

> summary(aov(amount~year,data=data1))

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

year 1 3.356e+12 3.356e+12 7.821 0.00521 \*\*

Residuals 2288 9.820e+14 4.292e+11

---

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

**Since p-value=0.00521<0.05, then reject Ho.**

**4.** > library(UsingR)

Loading required package: MASS

> mtcars

> attach(mtcars)

> model=lm(mpg~cyl+am+cyl\*am)

> anova(model)

Analysis of Variance Table

Response: mpg

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

cyl 1 817.71 817.71 94.6416 1.761e-10 \*\*\*

am 1 36.97 36.97 4.2791 0.04793 \*

cyl:am 1 29.44 29.44 3.4073 0.07551 .

Residuals 28 241.92 8.64

---

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

Thus, there is no evidence of a significant interaction effect for p-value= 0.07551. This test also shows that *cyl* and *am* are the main effects of *mpg* .

**5. a)**

> library(UsingR)

> babies

> data=data.frame(babies$gestation,babies$smoke,babies$ht,babies$wt1,babies$wt)

> data

**b)**

> attach(data)

> clean=subset(data,babies.gestation!=999&babies.smoke!=9&babies.ht!=99&babies.wt1!

=999&babies.wt!=999)

> clean

**c)** > BMI=(clean$babies.wt1/(clean$babies.ht\*clean$babies.ht))\*703

> BMI

**d)** > babies.gestation[babies.gestation<259]="1"

> babies.gestation[babies.gestation>=259]="0"

> premature=babies.gestation

> data1=data.frame(clean$babies.gestation,premature)

> data1

**e)** > data2=data.frame(babies.smoke,BMI,premature)

> data2

> model=glm(data2$premature~data2$babies.smoke+data2$BMI,family=binomial(logit))

> model

Call: glm(formula = data2$premature ~ data2$babies.smoke + data2$BMI,

family = binomial(logit))

Coefficients:

(Intercept) data2$babies.smoke data2$BMI

-3.34477 0.07464 0.03896

Degrees of Freedom: 1174 Total (i.e. Null); 1172 Residual

Null Deviance: 664.8

Residual Deviance: 662.9 AIC: 668.9