1.

(a)

> x=c(23.1,32.8,31.8,32,30.4,24.0,39.5,24.2,52.5,37.9,30.5,25.1,12.4,35.1,31.5,21.1)

> y=c(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)

> model=lm(y~x)

> model

Call:

lm(formula = y ~ x)

Coefficients:

(Intercept) x

0.5180 0.5016

The assumption is **

> 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

(b)

> confint(model,level=0.9)

5 % 95 %

(Intercept) -2.2427613 3.2788045

x 0.4139049 0.5892431

(c)

> model2=lm(y~x-1)

> model2

Call:

lm(formula = y ~ x - 1)

Coefficients:

x

0.5174

(d)

> summary(model)$r.squared

[1] 0.8788327

> summary(model2)$r.squared

[1] 0.9899623

Model2 is better.

2.

(a)

> library(MASS)

> data(Cars93)

> dim(Cars93)

[1] 93 27

(b)

> model=lm(Cars93$MPG.city~Cars93$EngineSize+Cars93$Weight+Cars93$Passengers+Cars93$Price)

> model

Call:

lm(formula = Cars93$MPG.city ~ Cars93$EngineSize + Cars93$Weight +

Cars93$Passengers + Cars93$Price)

Coefficients:

(Intercept) Cars93$EngineSize Cars93$Weight Cars93$Passengers Cars93$Price

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

(d)

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

Start: AIC=212.87

Cars93$MPG.city ~ Cars93$EngineSize + Cars93$Weight + Cars93$Passengers +

Cars93$Price

Df Sum of Sq RSS AIC

- Cars93$EngineSize 1 1.04 824.89 210.99

- Cars93$Passengers 1 3.77 827.62 211.29

- Cars93$Price 1 4.96 828.82 211.43

<none> 823.85 212.87

- Cars93$Weight 1 349.67 1173.52 243.77

Step: AIC=210.99

Cars93$MPG.city ~ Cars93$Weight + Cars93$Passengers + Cars93$Price

Df Sum of Sq RSS AIC

- Cars93$Passengers 1 3.20 828.10 209.35

- Cars93$Price 1 4.84 829.74 209.53

<none> 824.89 210.99

+ Cars93$EngineSize 1 1.04 823.85 212.87

- Cars93$Weight 1 627.12 1452.01 261.57

Step: AIC=209.35

Cars93$MPG.city ~ Cars93$Weight + Cars93$Price

Df Sum of Sq RSS AIC

- Cars93$Price 1 11.96 840.05 208.68

<none> 828.10 209.35

+ Cars93$Passengers 1 3.20 824.89 210.99

+ Cars93$EngineSize 1 0.47 827.62 211.29

- Cars93$Weight 1 1050.34 1878.44 283.52

Step: AIC=208.68

Cars93$MPG.city ~ Cars93$Weight

Df Sum of Sq RSS AIC

<none> 840.05 208.68

+ Cars93$Price 1 11.96 828.10 209.35

+ Cars93$Passengers 1 10.31 829.74 209.53

+ Cars93$EngineSize 1 0.06 839.99 210.67

- Cars93$Weight 1 2065.52 2905.57 322.09

3.

(a)

> library(UsingR)

> data(npdb)

> attach(npdb)

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

age, amount, field, ID, state, year

> 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

4.

> library(UsingR)

> 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

5.

(a)

> library(UsingR)

>data(babies)

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

(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]="0"

> babies.gestation[babies.gestation<259]="1"

> 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