4.5.a

> plastic<-read.table(file.choose(),header=F)

> summary(plastic)

> y<-plastic$V1

> x<-plastic$V2

> plastic.lm<-lm(y~x)

> summary(plastic.lm)

> confint(plastic.lm,level=1-.1/2)

> anova(plastic.lm)

V1 V2

Min.: 196.0 Min.: 16

1st Qu.: 212.5 1st Qu.: 22

Median : 226.5 Median :28

Mean: 225.6 Mean: 28

3rd Qu.: 239.2 3rd Qu.: 34

Max.: 253.0 Max.: 40

Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-5.1500 -2.2188 0.1625 2.6875 5.5750

Coefficients:

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

(Intercept) 168.60000 2.65702 63.45 < 2e-16 \*\*\*

x 2.03438 0.09039 22.51 2.16e-12 \*\*\*

---

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

Residual standard error: 3.234 on 14 degrees of freedom

Multiple R-squared: 0.9731, Adjusted R-squared: 0.9712

F-statistic: 506.5 on 1 and 14 DF, p-value: 2.159e-12

4.5.b

>cor(x,y)

They are positively correlated and yes, it does match the procedure in part a.

4.5.c

A family confidence coefficient indicates the proportion of families of estimates that are entirely correct when repeated samples are selected and the specified confidence intervals for the entire family are calculated for each sample

4.12.a

> typo<-read.table(file.choose(),header=F)

> y<-typo$V1

> x<-typo$V2

> typo.lm<-lm(y~x)

> summary(typo.lm)

The estimated regression function is Y=1.0878 + 17.9704X + error

Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-4.349 -3.027 -1.233 1.347 10.208

Coefficients:

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

(Intercept) 1.0878 2.7519 0.395 0.701

x 17.9704 0.1681 106.886 <2e-16 \*\*\*

---

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

Residual standard error: 4.69 on 10 degrees of freedom

Multiple R-squared: 0.9991, Adjusted R-squared: 0.999

F-statistic: 1.142e+04 on 1 and 10 DF, p-value: < 2.2e-16

4.12.b

> plot(x,y)

> abline(typo.lm)

> origin.lm<-lm(y~0+x)

> summary(origin.lm)

> abline(origin.lm)

Yes, it appears that the regression is a good fit because the p-value <.02

4.12.c

> g<-predict(typo.lm,data.frame(x),se.fit=T,interval ="confidence",level=.98)

> anova(typo.lm)

It appears that management should consider revising their cost of typographical errors

Ho: There is no lack of fit, E{Y}= B0+ B1\*X

Ha: There is lack of fit, E{Y}≠B0+ B1\*X

$fit

fit lwr upr

1 126.88093 121.84586 131.91600

2 216.73316 212.84780 220.61852

3 180.79227 176.56109 185.02345

4 180.79227 176.56109 185.02345

5 252.67406 248.93020 256.41791

6 450.34897 444.10761 456.59032

7 540.20120 531.98152 548.42088

8 450.34897 444.10761 456.59032

9 324.55584 320.42798 328.68370

10 180.79227 176.56109 185.02345

11 72.96959 66.91257 79.02661

12 108.91048 103.55337 114.26759

$se.fit

1 2 3 4 5 6 7 8

1.821813 1.405818 1.530946 1.530946 1.354618 2.258277 2.974084 2.258277

9 10 11 12

1.493562 1.530946 2.191578 1.938334

$df

[1] 10

$residual.scale

[1] 4.690277

4.12.d

> h<-predict(typo.lm,data.frame(x=10),se.fit=T,interval ="prediction",level=.98)

$fit

fit lwr upr

1 180.7923 167.1564 194.4282

$se.fit

[1] 1.530946

$df

[1] 10

$residual.scale

[1] 4.690277

4.13.a

> e<-typo.lm$resid

> sum(e)

[1] 1.831868e-15

> plot(typo.lm$fitted,e)

It appears that the residuals do not sum to zero, most likely due to the stated typo graphical errors. The plot seems to vary widely

4.13.b

Ho: There is no lack of fit, E{Y}= B0+ B1\*X

Ha: There is lack of fit, E{Y}≠B0+ B1\*X

P-value > .01 so we cannot reject null hypothesis. There is no lack of fit

> anova(typo.lm)

The p-value is 2.2e-16. there is no lack of fit

Analysis of Variance Table

Response: y

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

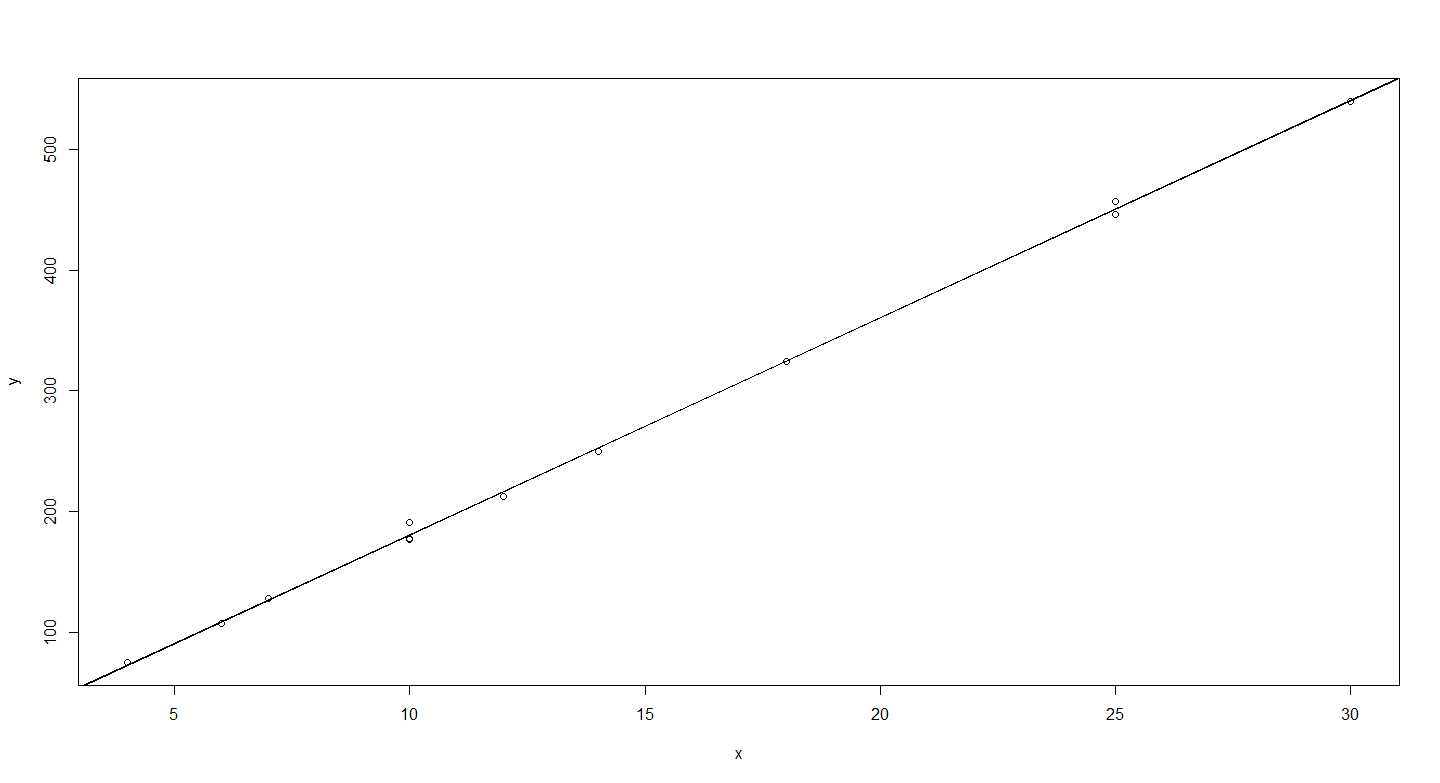
x 1 251326 251326 11425 < 2.2e-16 \*\*\*

Residuals 10 220 22

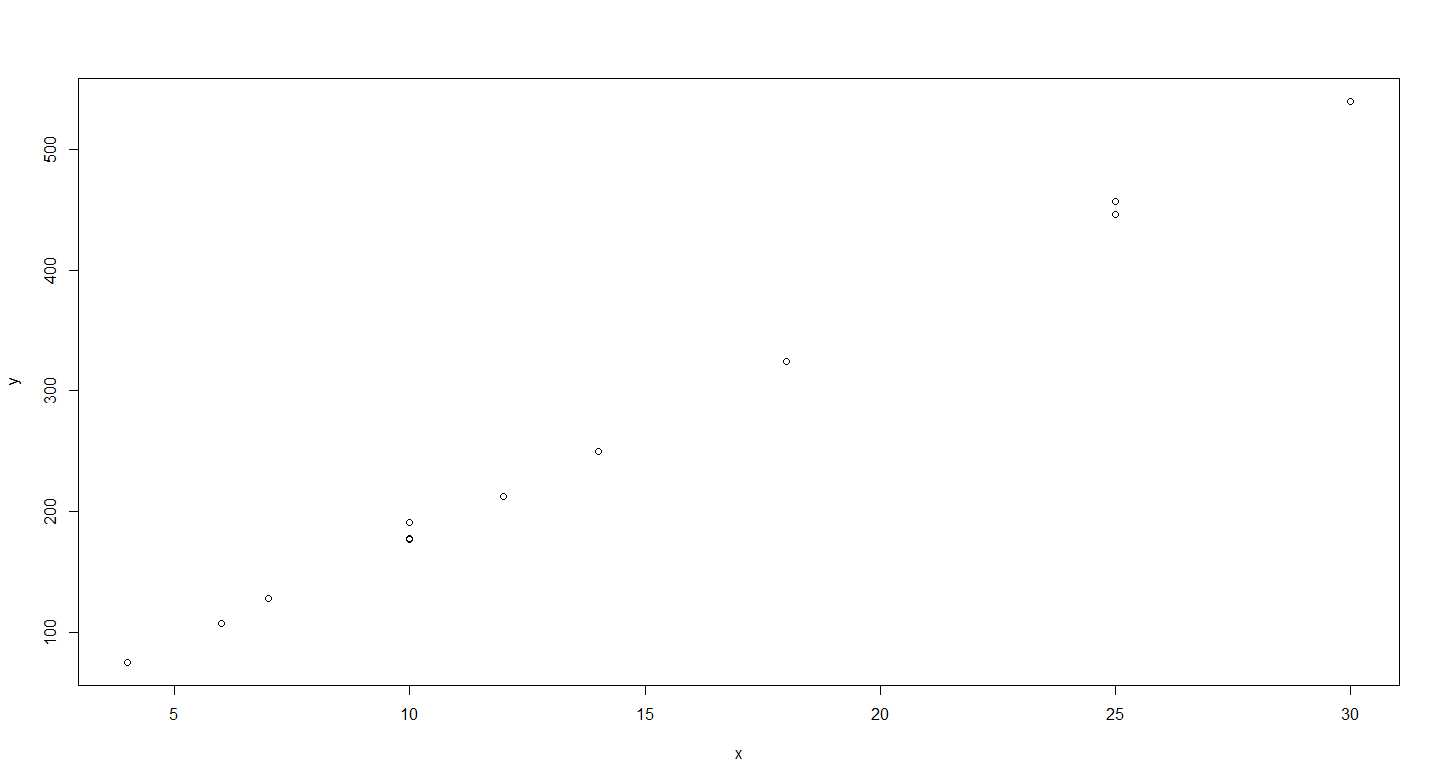
---

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

4.12.b



4.12.b



4.13.a

