**Book Problems**

3.31

**a)**

par(mfrow=c(2,2))

hw4.book.1 <- read.csv("~/hw4-book-1.txt", sep="")

x <- hw4.book.1$Cycfail

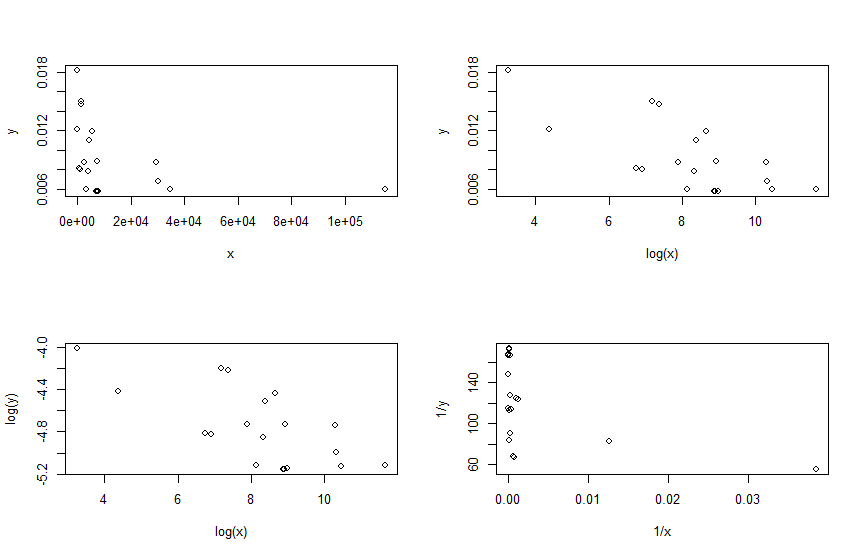
y <- hw4.book.1$Strampl

plot(x,y)

plot(log(x),y)

plot(log(x),log(y))

plot(1/x,1/y)



**b)** y vs log(x)

**c)** yhat = 0.0197092 + -0.0012805log(xi)

0.0197092 + -0.0012805log(5000) =.0088

3.33

**a)**  It could be a linear transformation; it looks like strength goes down as thickness is increased

**b)**

Residuals:

Min 1Q Median 3Q Max

-5.6278 -2.2024 0.2857 2.4414 4.8790

Coefficients:

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

(Intercept) 1.452e+01 4.754e+00 3.055 0.00717 \*\*

thickness 4.323e-02 1.981e-02 2.183 0.04337 \*

I(thickness^2) -6.001e-05 1.786e-05 -3.359 0.00372 \*\*

---

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

Residual standard error: 3.269 on 17 degrees of freedom

Multiple R-squared: 0.7797, Adjusted R-squared: 0.7538

F-statistic: 30.09 on 2 and 17 DF, p-value: 2.599e-06

1.452e+01 + 4.323e-02\*500 -6.001e-05 \* 500^2 = 21.1325

**Lect 13-4**

**a)**

cleaned <- read.csv("C:/Users/jaesu/OneDrive/Desktop/School\_code/STAT390/hw-1/cleaned.csv")

x <- cleaned$Drink.week

y <- cleaned$Money.week

plot(sqrt(x), y)

abline(lm.1)

lm.1 <- lm(y ~ x)

summary(lm.1)

Residuals:

Min 1Q Median 3Q Max

-12.287 -2.314 -2.301 2.699 22.699

Coefficients:

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

(Intercept) 2.3145 1.6823 1.376 0.178

x 4.9862 0.9939 5.017 1.53e-05 \*\*\*

---

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

Residual standard error: 6.944 on 35 degrees of freedom

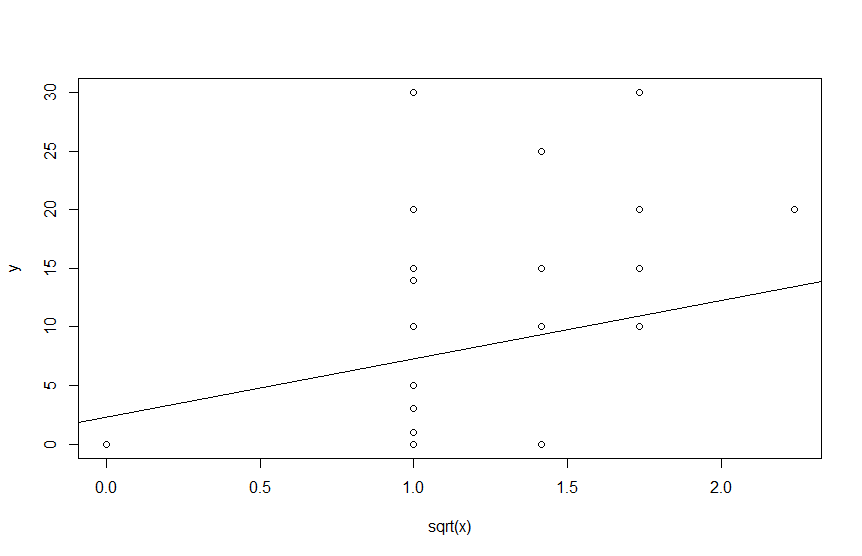
Multiple R-squared: 0.4183, Adjusted R-squared: 0.4017

F-statistic: 25.17 on 1 and 35 DF, p-value: 1.526e-05

Intercept) 2.3145

x 4.9862

The prediction of y from x is a linear equation with coefficients, the intercept is just where the line starts at 0 and the x coefficient is how much y increases or decreases due to change in x

**b)** 

**c)**

#R^2 = 0.4183 the variance of amount of money spend of week is due to the drinks per week

**d)**

#Se = 6.944 Typical error from the from the fit line is 6.944

**Lect 14-1**

**a)**

par(mfrow=c(2,2))

transform\_dat <- read.csv("C:/Users/jaesu/Downloads/transform\_dat.txt", sep="")

x <- transform\_dat$x

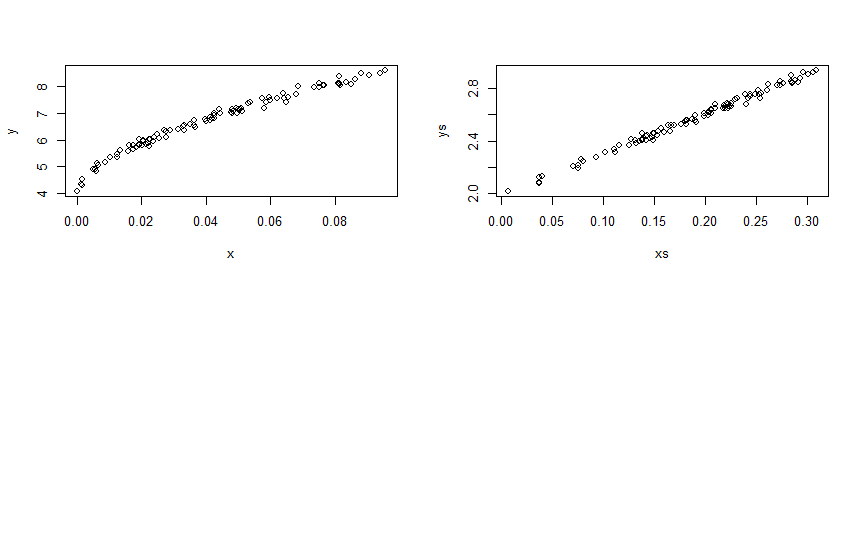
y <- transform\_dat$y

plot(x,y)

xs <- sqrt(x)

ys <- sqrt(y)

plot(xs,ys)



**b)**

lm.1 <- lm(ys ~xs)

summary(lm.1)

Residuals:

Min 1Q Median 3Q Max

-0.044793 -0.013386 0.000501 0.011582 0.042927

Coefficients:

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

(Intercept) 1.99403 0.00552 361.2 <2e-16 \*\*\*

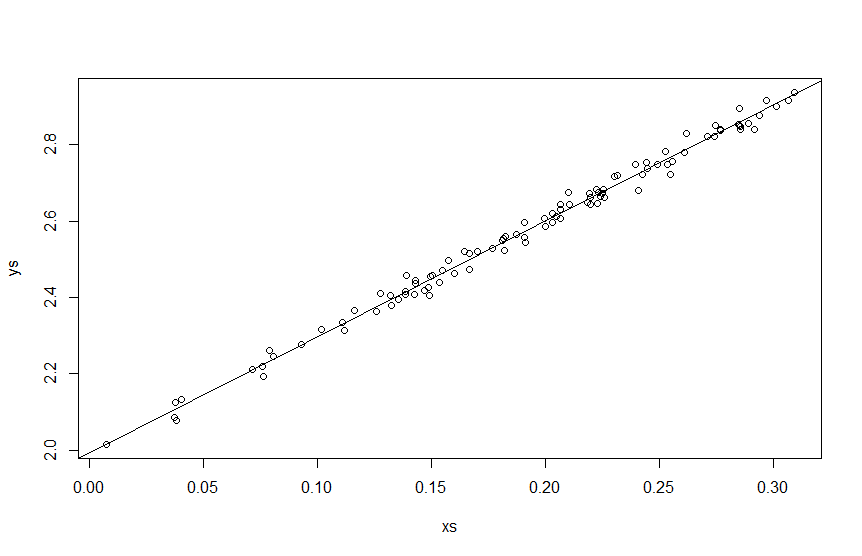
xs 3.03409 0.02710 112.0 <2e-16 \*\*\*

---

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

Residual standard error: 0.01911 on 98 degrees of freedom

Multiple R-squared: 0.9922, Adjusted R-squared: 0.9922



**c)**

# 99% of transformed y is due to transformed and the typical error is 0.019

**e)**

x <- transform\_dat$x

y <- transform\_dat$y

lm.2 <- lm(y ~ sqrt(x) + I(x))

summary(lm.2)

Residuals:

Min 1Q Median 3Q Max

-0.24317 -0.07320 0.00536 0.05339 0.23265

Coefficients:

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

(Intercept) 3.94770 0.05228 75.506 < 2e-16 \*\*\*

sqrt(x) 12.50864 0.61266 20.417 < 2e-16 \*\*\*

I(x) 8.01869 1.69229 4.738 7.37e-06 \*\*\*

---

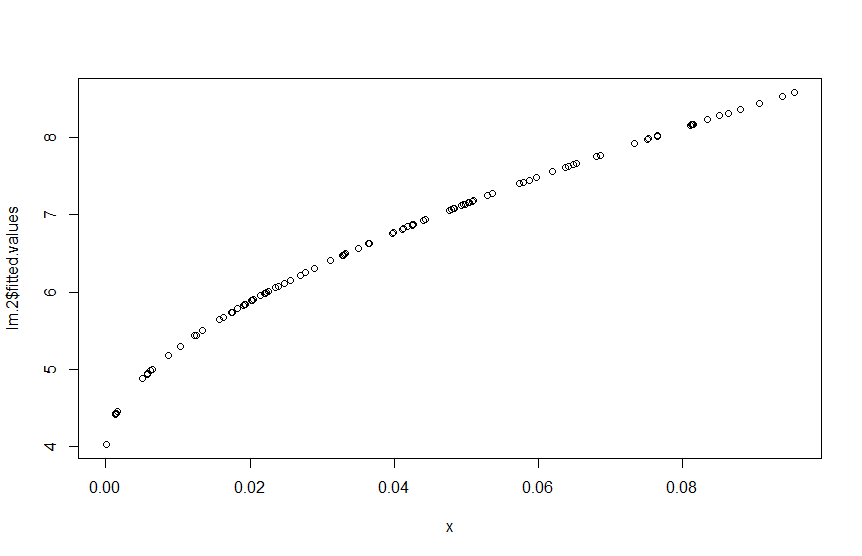
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0992 on 97 degrees of freedom

Multiple R-squared: 0.9919, Adjusted R-squared: 0.9918

F-statistic: 5954 on 2 and 97 DF, p-value: < 2.2e-16

**f)**



**Lect 15-2**

**a)**

hw4 <- read.csv("~/hw4.txt", sep="")

x1 <- hw4$Depth

x2 <- hw4$Content

ys <- hw4$Strength

lm.1 <- lm(ys ~ x1 + x2 + I(x1^2) + I(x2^2) + x1 \* x2)

summary(lm.1)

Residuals:

Min 1Q Median 3Q Max

-8.349 -2.188 0.279 1.649 13.613

Coefficients:

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

(Intercept) -140.22976 136.13743 -1.030 0.3331

x1 -16.47521 9.07116 -1.816 0.1069

x2 12.82710 8.25854 1.553 0.1590

I(x1^2) 0.09555 0.07206 1.326 0.2214

I(x2^2) -0.24339 0.12744 -1.910 0.0925 .

x1:x2 0.49864 0.23543 2.118 0.0670 .

---

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

Residual standard error: 7.023 on 8 degrees of freedom

Multiple R-squared: 0.7561, Adjusted R-squared: 0.6037

F-statistic: 4.961 on 5 and 8 DF, p-value: 0.02307

# b1 = -16.47521, b2 = 12.82710, b3 = 0.09555, b4 = -0.2433, b5 = 0.49864

**b)**

#You cannot interpret the regression coefficients because the term x1\*x2 means the change in x1 impacts x2 which means that the coefficients are uninterpretable

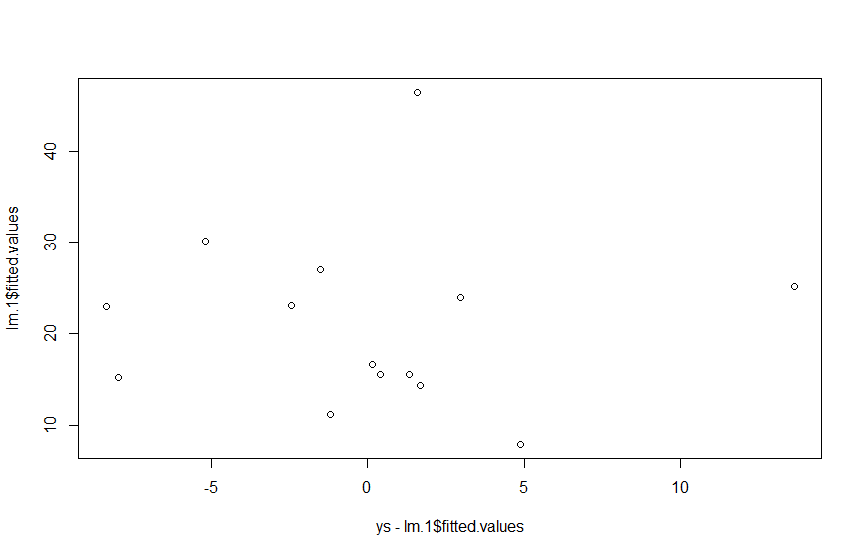
**c)**

# R^2 = .76 which means that these predictors are pretty good at determining the strength

d)

Standard error is 7 which means on average the fit value is 7 units away from the actual

e) plot(ys - lm.1$fitted.values, lm.1$fitted.values)

Plot looks random which means the model fits fine 

f)

lm.2 <- lm(ys ~ x1 + x2)

summary(lm.2)

Residuals:

Min 1Q Median 3Q Max

-12.867 -4.475 -1.526 3.878 19.321

Coefficients:

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

(Intercept) 14.8893 23.2447 0.641 0.5349

x1 0.6607 0.2737 2.414 0.0344 \*

x2 -0.0284 0.6423 -0.044 0.9655

---

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

Residual standard error: 9.019 on 11 degrees of freedom

Multiple R-squared: 0.447, Adjusted R-squared: 0.3465

F-statistic: 4.446 on 2 and 11 DF, p-value: 0.03845

g)

# R^2 = .447 which means the model is okay at determining the strength

h)

# Yes, because with only the lower order terms gives a lower R^2 which is the amount strength is determined by depth and content

i)

# Se for model 2 is 9.01 and model 1 is 7.023 which means that on average model 1 is closer to the actual value

# than in model 2

**Lect 15-3**

dat1 <- read.table("C:/Users/jaesu/Downloads/hw\_3\_dat1.txt", quote="\"", comment.char="", header = TRUE)

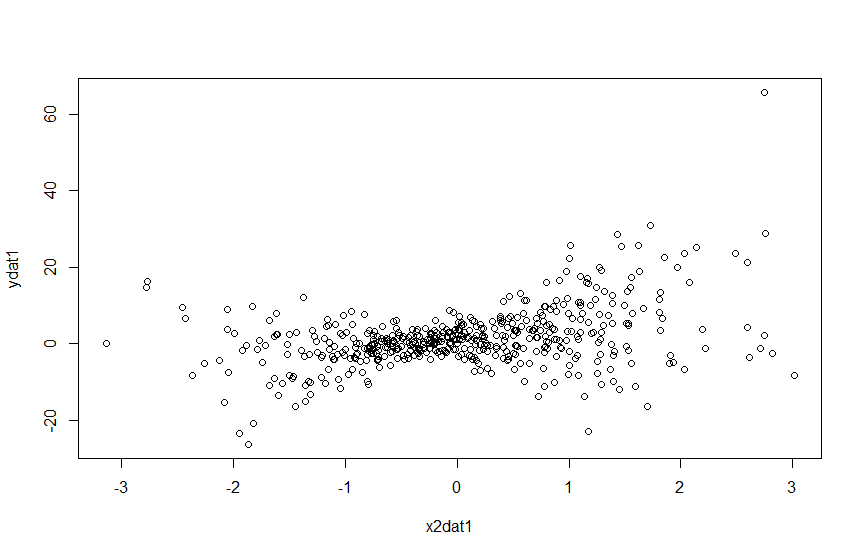
ydat1 <- dat1$y

x1dat1 <- dat1$x1

x2dat1 <- dat1$x2

plot(x1dat1, ydat1)

plot(x2dat1, ydat1)



cor(dat1)

x1 x2 y

x1 1.00000000 0.05645841 0.3145897

x2 0.05645841 1.00000000 0.3861130

y 0.31458969 0.38611298 1.0000000

lm.1 <- lm(ydat1 ~ x1dat1 + x2dat1)

summary(lm.1)

# R^2 = .235 Pretty weak

lm.2 <- lm(ydat1 ~ x1dat1 + x2dat1 + x1dat1:x2dat1)

summary(lm.2)

# R^2 = .9356 much stronger

lm.3 <- lm(ydat1 ~ x1dat1 + x2dat1 + I(x1dat1^2) + I(x2dat1^2) + x1dat1:x2dat1)

summary(lm.3)

# R^2 = .9358 not much stronger at all

# ydat1 ~ x1dat1 + x2dat1 + x1dat1:x2dat1 this is probably the most accurate model, there is almost no colinearity

# between x1 & x1 given that their correlation coefficent is about 5%, and the interaction is not over fitting

# as the model jumped to be three times more accurate with the interaction and adding quadratic makes it

# barely more accurate which makes it a lot more prone to overfitting. The model is definately nonlinea

dat2 <- read.csv("C:/Users/jaesu/Downloads/hw\_3\_dat2.txt", sep="")

ydat2 <- dat2$y

x1dat2 <- dat2$x1

x2dat2 <- dat2$x2

plot(x1dat2, x2dat2)

plot(x1dat2, ydat2)

plot(x2dat2, ydat2)

cor(dat2)

x1 x2 y

x1 1.0000000 0.8914847 0.4970352

x2 0.8914847 1.0000000 0.5066167

y 0.4970352 0.5066167 1.0000000

plot(x1dat2, x2dat2)

# colinearity between x1 & x2

lm.11 <- lm(ydat2 ~ x1dat2 + I(x1dat2^2))

summary(lm.11)

Residuals:

Min 1Q Median 3Q Max

-12.4868 -1.9684 -0.0062 2.0696 13.9185

Coefficients:

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

(Intercept) 0.8977 0.2089 4.296 2.09e-05 \*\*\*

x1dat2 4.6551 0.1714 27.165 < 2e-16 \*\*\*

I(x1dat2^2) 5.6122 0.1183 47.458 < 2e-16 \*\*\*

---

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

Residual standard error: 3.839 on 497 degrees of freedom

Multiple R-squared: 0.8639, Adjusted R-squared: 0.8633

F-statistic: 1577 on 2 and 497 DF, p-value: < 2.2e-16

# R^2 value is 0.8639

lm.12 <- lm(ydat2 ~ x2dat2 + I(x2dat2^2))

summary(lm.12)

Residuals:

Min 1Q Median 3Q Max

-13.5117 -2.0642 -0.1169 2.0078 14.5754

Coefficients:

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

(Intercept) 0.8764 0.1881 4.658 4.1e-06 \*\*\*

x2dat2 4.6263 0.1527 30.289 < 2e-16 \*\*\*

I(x2dat2^2) 5.3314 0.1008 52.876 < 2e-16 \*\*\*

---

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

Residual standard error: 3.486 on 497 degrees of freedom

Multiple R-squared: 0.8878, Adjusted R-squared: 0.8874

F-statistic: 1966 on 2 and 497 DF, p-value: < 2.2e-16

# R^2 value is 0.8878

# The model used should be ydat2 ~ x2dat2 + I(x2dat2^2) there is colinearity between x1 and x2 which means

# that one is unnecessary and should be removed, I chose x1 to be removed because R^2 is better for that one

**Lect 15-3**

par(mfrow=c(2,2))

mult <- read.table("C:/Users/jaesu/Downloads/hw\_3\_mult\_simple\_dat.txt", quote="\"", comment.char="")

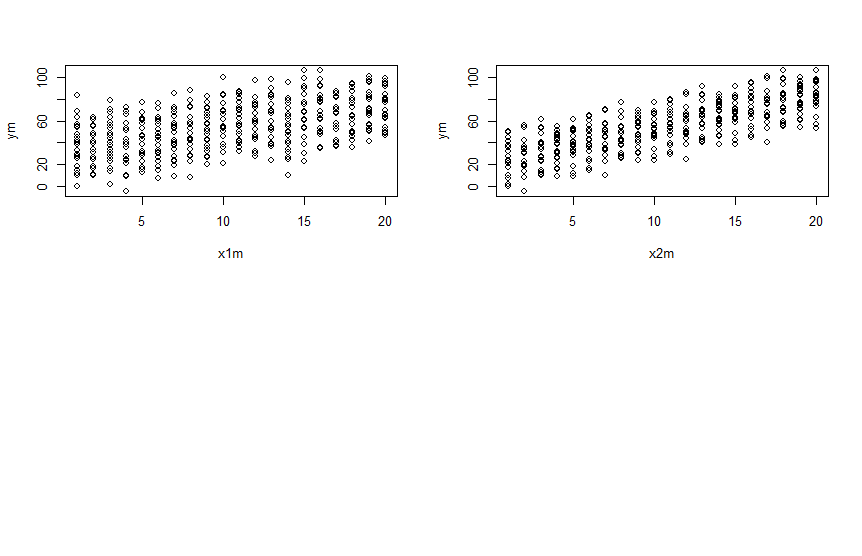
x1m <- mult$V1

x2m <- mult$V2

ym <- mult$V3

plot(x1m, ym)

plot(x2m, ym)



a)

lm.1 <- lm(ym ~ x1m + x2m)

summary(lm.1)

lm(formula = ym ~ x1m + x2m)

Residuals:

Min 1Q Median 3Q Max

-22.999 -6.061 0.200 5.392 22.249

Coefficients:

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

(Intercept) 2.8781 1.2377 2.325 0.0206 \*

x1m 1.9268 0.0777 24.799 <2e-16 \*\*\*

x2m 2.9980 0.0777 38.585 <2e-16 \*\*\*

---

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

Residual standard error: 8.961 on 397 degrees of freedom

Multiple R-squared: 0.8412, Adjusted R-squared: 0.8404

F-statistic: 1052 on 2 and 397 DF, p-value: < 2.2e-16

b)

lm.2 <- lm(ym ~ x1m)

summary(lm.2)

Residuals:

Min 1Q Median 3Q Max

-51.480 -14.898 0.256 15.373 47.732

Coefficients:

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

(Intercept) 34.3572 2.0260 16.96 <2e-16 \*\*\*

x1m 1.9268 0.1691 11.39 <2e-16 \*\*\*

---

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

Residual standard error: 19.51 on 398 degrees of freedom

Multiple R-squared: 0.2459, Adjusted R-squared: 0.244

F-statistic: 129.8 on 1 and 398 DF, p-value: < 2.2e-16

lm.3 <- lm(ym ~ x2m)

summary(lm.3)

Residuals:

Min 1Q Median 3Q Max

-33.999 -9.122 0.734 10.459 29.730

Coefficients:

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

(Intercept) 23.1098 1.4842 15.57 <2e-16 \*\*\*

x2m 2.9980 0.1239 24.20 <2e-16 \*\*\*

---

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

Residual standard error: 14.29 on 398 degrees of freedom

Multiple R-squared: 0.5953, Adjusted R-squared: 0.5943

F-statistic: 585.5 on 1 and 398 DF, p-value: < 2.2e-16

c)

# The beta for each regression model for x1 and x2 are the respective betas for the combined model. This makes

d)

# sense because if one x is held constant then the beta is the linear relationship between the x and the

# predicted y

min(x1m)

x <- seq(min(x1m), max(x1m), length =100)

y <- seq(min(x2m), max(x2m), length =100)

f <- function(x,y) {

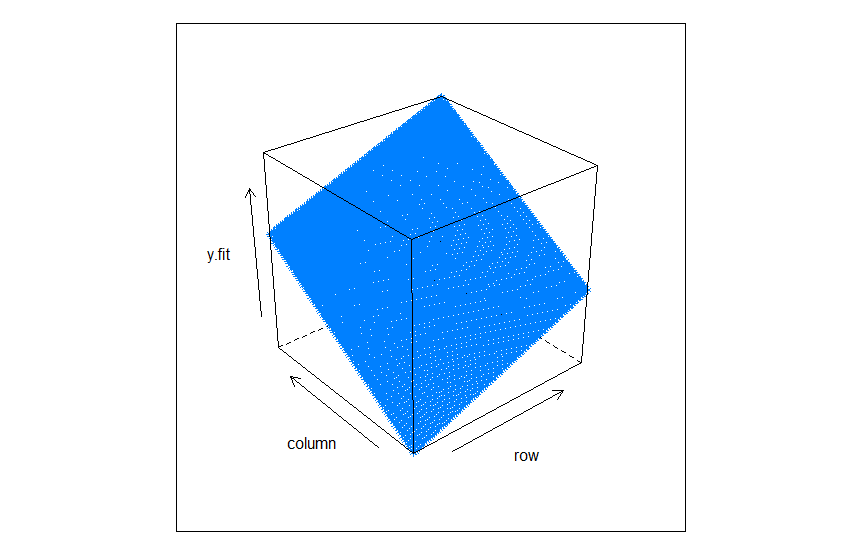
r <- lm.1$coefficients[1] + lm.1$coefficients[2]\*x + lm.1$coefficients[3]\*y

}

y.fit <- outer(x,y,f)

y.fit

library(lattice)



**Lect 16-1**

par(mfrow=c(1,2))

ntrial = 5000

xmax = numeric(ntrial)

for(trial in 1:ntrial) {

x = rnorm(50, 0, 1)

hist(x, breaks=10)

xmax[trial] = max(x)

}

xmax

hist(xmax, main="")

xmin = numeric(ntrial)

for(trial in 1:ntrial) {

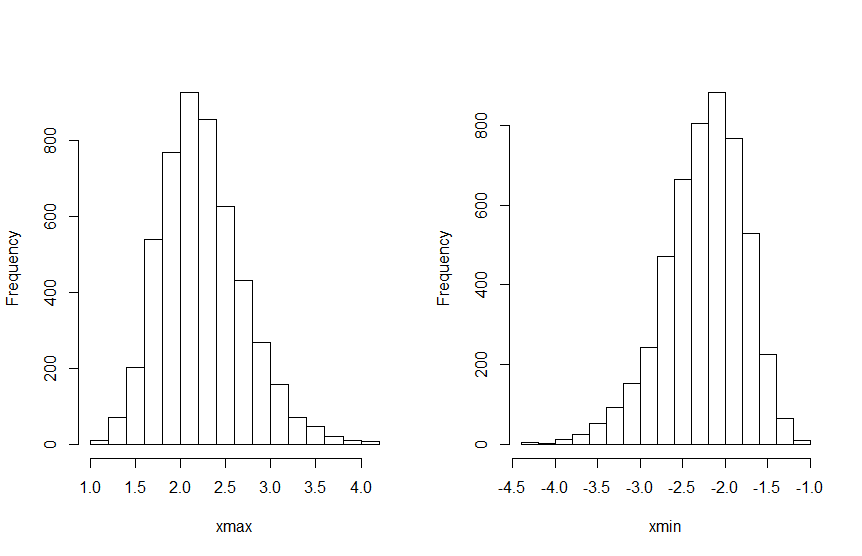
x = rnorm(50, 0, 1)

hist(x, breaks=10)

xmin[trial] = min(x)

}

hist(xmin, main="")



**Lect 16-2**

ntrial = 5000

xmean = numeric(ntrial)

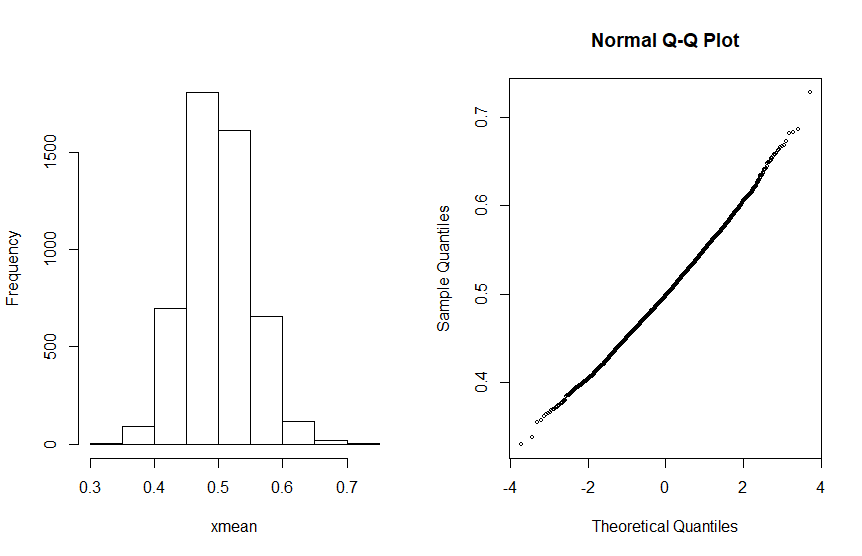
for(trial in 1:ntrial) {

x = rexp(100, 2)

xmean[trial] = mean(x)

}

hist(xmean, main="")



# The y intercept is about at .5 which is the predicted mean.

# The slope is about .05 which is far from the prediction .5/sqrt(5000) = .007