Name: Parth Pareek
UNI: PP2547
Date: 3/2/2016
Assignment: HW5

1.

a. SE = 0.087474
b. CI = 5.614 to 6.017

2.

a. Confirmed from code output

```
Call:
lm(formula = dat$Denmark ~ dat$Year)

Residuals:
      Min        1Q    Median        3Q       Max
-0.003225 -0.001339  0.000089  0.001119  0.003790

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.987e-01  4.080e-02  14.673   <2e-16 ***
dat$Year    -4.289e-05  2.069e-05  -2.073   0.0442 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.001803 on 43 degrees of freedom
Multiple R-squared:  0.09083,    Adjusted R-squared:  0.06968
F-statistic: 4.296 on 1 and 43 DF,  p-value: 0.04424
```

```
Call:
lm(formula = dat$Netherlands ~ dat$Year)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0031437 -0.0008246  0.0002819  0.0009287  0.0021478

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.724e-01  2.792e-02   24.08  < 2e-16 ***
dat$Year    -8.084e-05  1.416e-05   -5.71 9.64e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.001233 on 43 degrees of freedom
Multiple R-squared:  0.4313,    Adjusted R-squared:  0.418
F-statistic: 32.61 on 1 and 43 DF,  p-value: 9.637e-07
```

```
Call:
lm(formula = dat$Canada ~ dat$Year)

Residuals:
       Min         1Q     Median         3Q        Max
-1.494e-03 -6.161e-04 -8.312e-05  4.951e-04  1.284e-03

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.338e-01  5.480e-02  13.390 3.98e-11 ***
dat$Year    -1.112e-04  2.768e-05  -4.017 0.000738 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.000768 on 19 degrees of freedom
  (24 observations deleted due to missingness)
Multiple R-squared:  0.4592,    Adjusted R-squared:  0.4307
F-statistic: 16.13 on 1 and 19 DF,  p-value: 0.0007376
```

```
Call:
lm(formula = dat$USA ~ dat$Year)

Residuals:
       Min         1Q     Median         3Q        Max
-5.343e-04 -1.800e-04 -1.714e-05  2.571e-04  3.743e-04

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.201e-01  1.860e-02  33.340  < 2e-16 ***
dat$Year    -5.429e-05  9.393e-06  -5.779 1.44e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0002607 on 19 degrees of freedom
  (24 observations deleted due to missingness)
Multiple R-squared:  0.6374,    Adjusted R-squared:  0.6183
F-statistic:  33.4 on 1 and 19 DF,  p-value: 1.439e-05
```

b. From code output:

Country	t-stat	p-value
Denmark	-2.07	0.044
Netherlands	-5.71	<0.001
Canada	-4.02	0.0007
USA	-5.78	<0.001

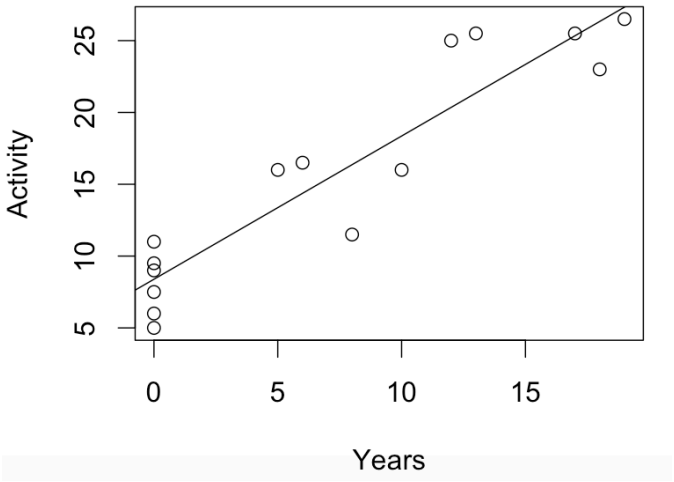
p-value is less than 0.005 for each country, indicating that proportion of male birth is truly declining

c. T-stat is calculated as Intercept/SE → SE for US is the least compared to other countries (order e^-0.05) and hence, can have the highest t-stat

d. Standard error is a function of SD and n-size. SD for USA is smaller than for Canada

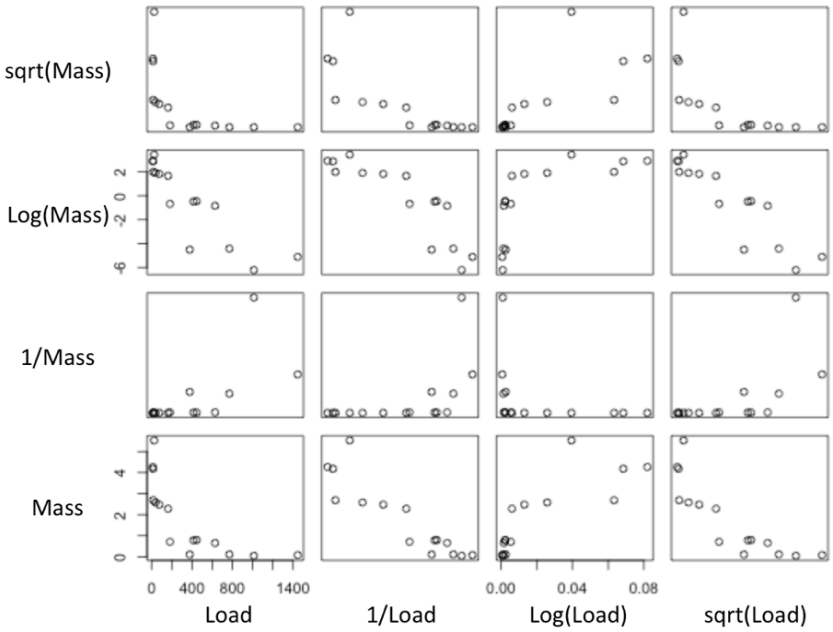
e. The proportion of males is calculated from the total population which varies for every country. This may lead to different SD for each of the countries.

3. One sided p-value = 8.57e^-0.05
   95% CI = 7.37 to 17.85
   Amount of activity associated with years of playing: Intercept = 8.38; coefficient = 0.9971 → 1 point increase is associated with 0.9971 years of playing music



4.

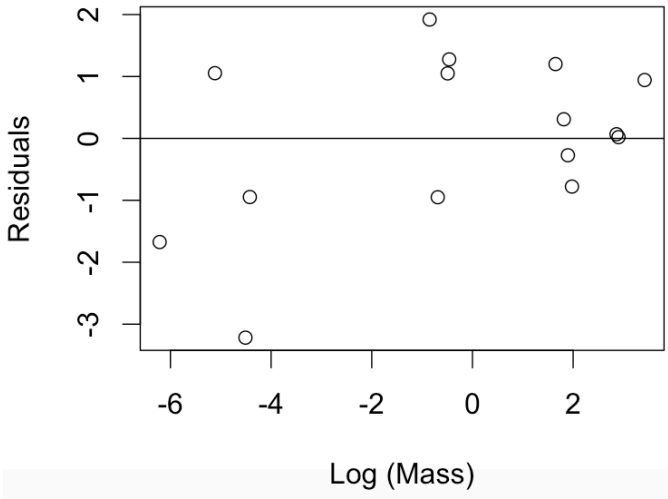
a. Scatter plot



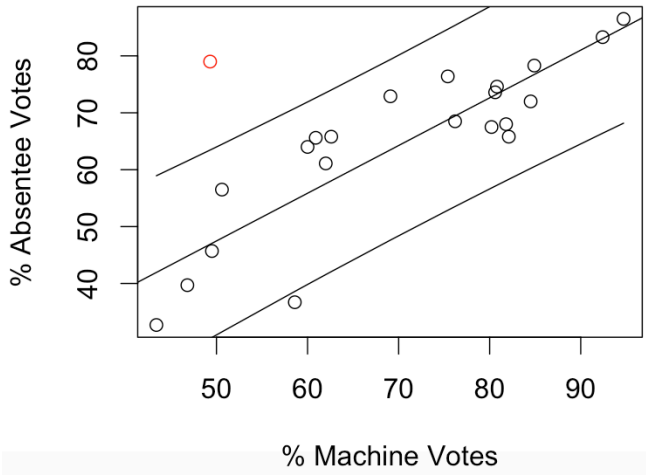
b. sqrt(load) and log(mass) seem the best transformation

   Y = 3.7965 – 0.2621 X Y → log(Mass) and X → sqrt(load)

c. The residuals vs. fitted values have been plotted below.



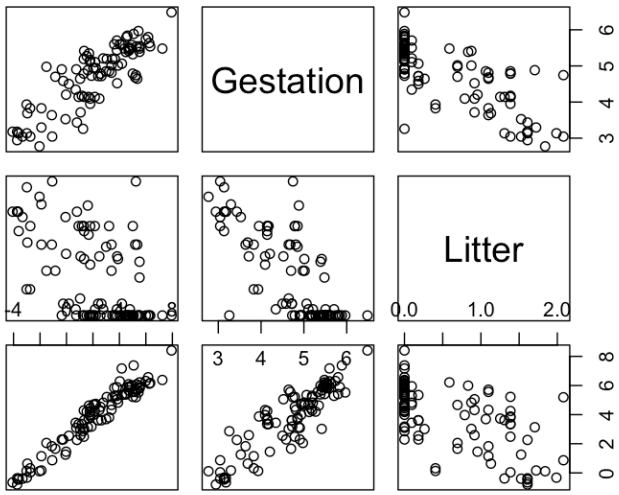
5. a and b.



c. Predicted value at 49.3 = 46.88664%
Standard Error = 7.915031
Standard Deviations away = 4.057
p-value = 0.00067

d. Bonferroni adjusted p-value = 22*0.00067 = 0.01479

6. a.



b. Estimates in Display 9.15 can be confirmed from the fitted model (summary below)

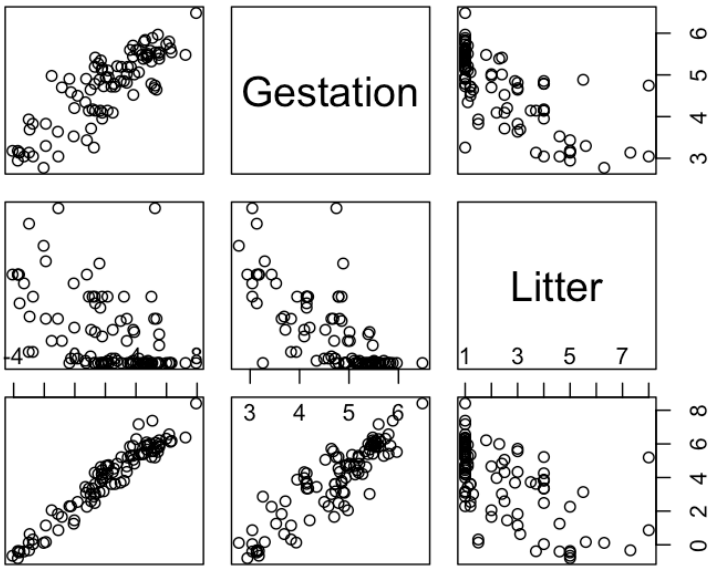
```
Call:
lm(formula = Brain ~ ., data = dat_trans)

Residuals:
     Min       1Q   Median       3Q      Max
-0.95415 -0.29639 -0.03105  0.28111  1.57491

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.85482    0.66167   1.292  0.19962
Body         0.57507    0.03259  17.647  < 2e-16 ***
Gestation    0.41794    0.14078   2.969  0.00381 **
Litter      -0.31007    0.11593  -2.675  0.00885 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4748 on 92 degrees of freedom
Multiple R-squared:  0.9537,    Adjusted R-squared:  0.9522
F-statistic: 631.6 on 3 and 92 DF,  p-value: < 2.2e-16
```

c. Litter on non-transformed scale



The distribution looks similar but we should still prefer the log transformed scales since effects can be gauged easily and back-transformation is simpler.

CODES

```
#Question 1
library(Sleuth3)
attach(case0702)
dat <- case0702

pH <- dat$pH
time <- dat$Time
model <- lm(pH ~ log(time))
predict(model, newdata = list(time = 5), interval = "predict",
        level = 0.95, se.fit = TRUE)

n = nrow(dat)
timet <- log(time)
SSR <- sum(model$residuals^2)
sig_est <- sqrt(SSR/(n-2))
SE_est <- sig_est*sqrt(1 + 1/n + (((log(5)-mean(timet))^2) /
                                    ((n-1)*(var(timet)))))

#Question 2
library(Sleuth3)
attach(ex0724)
dat <- ex0724
summary(lm(dat$Denmark~dat$Year))
summary(lm(dat$Netherlands~dat$Year))
summary(lm(dat$Canada~dat$Year))
summary(lm(dat$USA~dat$Year))

#Question 3
library(Sleuth3)
attach(ex0728)
dat <- ex0728
control <- dat[which(Years == "0"),]
test <- dat[which(Years != "0"),]
t.test(test$Activity,control$Activity,var.equal = TRUE, alternative =
"greater") #for p-value
t.test(test$Activity,control$Activity,var.equal = TRUE) #for CI

lm(Activity~Years,data = dat)

#Question 4
library(lattice)
library(ggplot2)
library(Sleuth3)
attach(ex0817)
dat <- ex0817
dat_trans <- cbind(dat[1],log(dat[1]),(1/dat[1]),sqrt(dat[1]),
                   dat[2],log(dat[2]),(1/dat[2]),sqrt(dat[2]))
colnames(dat_trans) <- c("Load","Log(Load)","1/Load","sqrt(Load)",
                         "Mass","Log(Mass)","1/Mass","sqrt(Mass)")
pairs(dat_trans,horInd = 1:4,verInd = 5:8)

lm(dat_trans[,6]~dat_trans[,4])
```

```
plot(dat_trans[,6],lm(dat_trans[,6]~dat_trans[,4])$residuals, xlab = "Log
(Mass)", ylab = "Residuals")
abline(a = 0, b = 0)

#Question 5
library(Sleuth3)
attach(ex0820)
dat <- ex0820

#Part a
plot(DemPctOfMachineVotes[-22],DemPctOfAbsenteeVotes[-22], ylab = "%
Absentee Votes", xlab = "% Machine Votes")
points(DemPctOfMachineVotes[22],DemPctOfAbsenteeVotes[22], col= "red")

#Part b
model <- lm(DemPctOfAbsenteeVotes[-22]~DemPctOfMachineVotes[-22])
abline(a = model$coeff[1],b = model$coeff[2])
CIline <- predict(model, interval = "predict")

s1 <- smooth.spline(DemPctOfMachineVotes[-22],CIline[,2])
lines(s1)
s2 <- smooth.spline(DemPctOfMachineVotes[-22],CIline[,3])
lines(s2)

#Part c
preddata <- predict(model, newdata = list(DemPctOfMachineVotes = 49.3),
        interval = "predict")[1]

n = nrow(dat)-1
SSR <- sum(model$residuals^2)
sig_est <- sqrt(SSR/(n-2))
SE_est <- sig_est*sqrt(1 + 1/n + (((49.3-mean(DemPctOfMachineVotes[-
22]))^2) /
                                    ((n-1)*(var(DemPctOfMachineVotes[-
22])))))

dev <- (79-preddata)/SE_est
pval <- 2*(1-pt(dev, df = 19))

#Part d
bon_pval <- (n+1)*pval

#Question 6
library(lattice)
library(ggplot2)
library(Sleuth3)
attach(case0902)
dat <- case0902
dat_trans <- log(dat[,2:5])
pairs(dat_trans, horInd = 2:4, verInd = c(3,4,1),diag.panel = NULL)
model <- lm(Brain~.,data = dat_trans)
summary(model)
newdat <- data.frame(dat_trans[1:3],dat[5])
pairs(newdat, horInd = 2:4, verInd = c(3,4,1),diag.panel = NULL)
```