Lab Number One

Ben Chu

January 23, 2018

```
library(tidyverse)
```

cleaning the data

```
uncleanlab1 <- read.csv("C:/Users/Branly Mclanbry/Downloads/lab1.csv")
lab1 <- uncleanlab1 %>%
  mutate(
    race = case_when(
      qa == 1 ~ "white",
      qa == 2 ~ "african-american",
      qa == 3 ~ "hispanic/latino",
      qa == 4 ~ "asian",
      qa == 5 ~ "native american",
      qa == 6 ~ "other"),
    support = (q1a + q1b +q1c))
```

1a

These two are not statistically correlated with each other. $R^2 = .0007$, $F(1,76) = 0.06$, $p = .81$

```
regdat<-lm(support~qb,lab1)
summary(regdat)
```

```
##
## Call:
## lm(formula = support ~ qb, data = lab1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8024 -4.5804 -0.5012  4.1141 11.5147
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.40497    3.30304   3.150  0.00233 **
## qb          -0.03171    0.13499  -0.235  0.81490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.113 on 76 degrees of freedom
## Multiple R-squared:  0.0007256,  Adjusted R-squared:  -0.01242
## F-statistic: 0.05519 on 1 and 76 DF,  p-value: 0.8149
```

```
cor.test(lab1$support,lab1$qb)
```

```
##
##  Pearson's product-moment correlation
##
## data:  lab1$support and lab1$qb
## t = -0.23492, df = 76, p-value = 0.8149
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2479816  0.1967729
## sample estimates:
##         cor
## -0.02693739
```

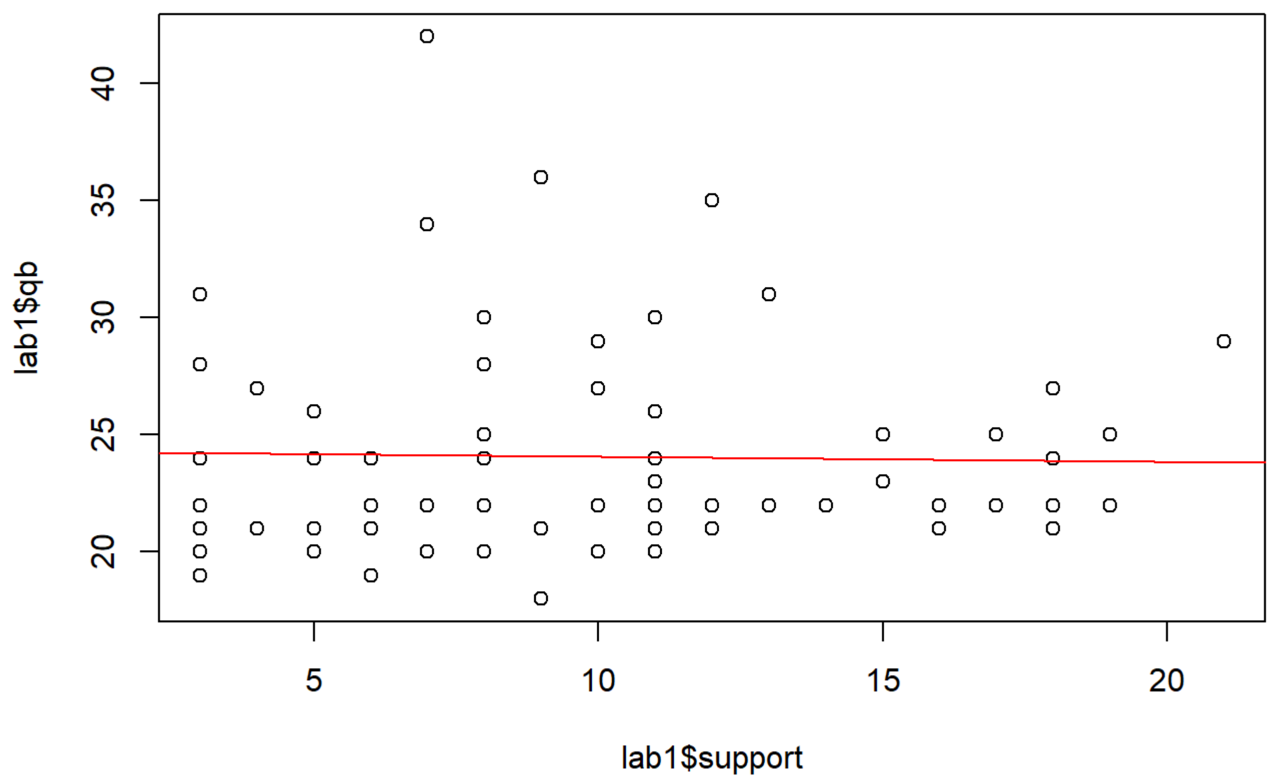
1b

```
newage = data.frame(qb=21)
predict(regdat,newage)
```

```
##        1
## 9.739008
```

1c

```
plot(lab1$support,lab1$qb)
abline(lm(lab1$qb~lab1$support),col = "red")
```



2a

These two are statistically correlated with each other. R^2 = .31, F(1,76) = 34.3, p < .001

```
beldat <-lm(support~q3,lab1)
summary(beldat)
```

```
##
## Call:
## lm(formula = support ~ q3, data = lab1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4246  -3.2194  -0.0963   3.5754   7.9037
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.7082     0.8259   6.912 1.29e-09 ***
## q3            1.3881     0.2370   5.856 1.14e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.245 on 76 degrees of freedom
## Multiple R-squared:  0.3109, Adjusted R-squared:  0.3019
## F-statistic:  34.3 on 1 and 76 DF,  p-value: 1.137e-07
```

```
cor.test(lab1$support,lab1$q3)
```

```
##
##  Pearson's product-moment correlation
##
## data:  lab1$support and lab1$q3
## t = 5.8563, df = 76, p-value = 1.137e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3825674 0.6940346
## sample estimates:
##       cor
## 0.5576258
```

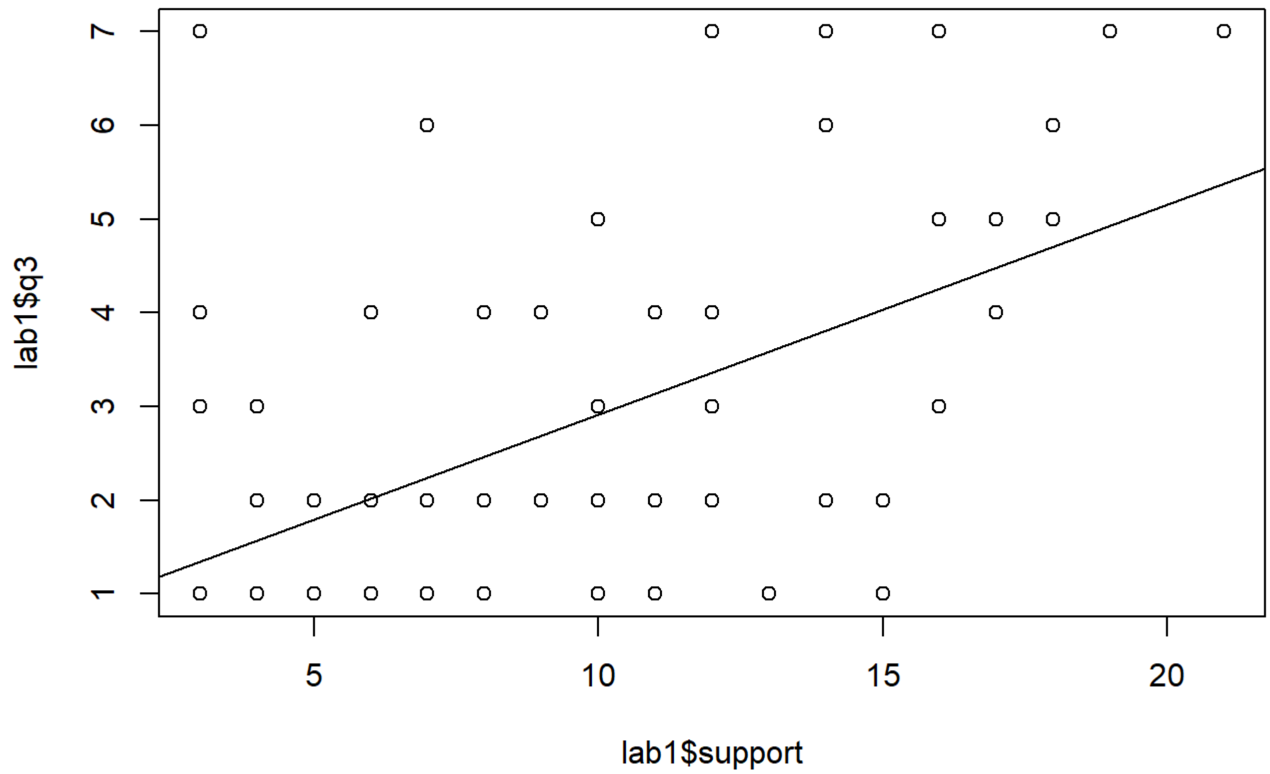
2b

```
newbel <- data.frame(q3=7)
predict(beldat,newbel)
```

```
##        1
## 15.42458
```

2c

```
plot(lab1$support,lab1$q3)
abline(lm(lab1$q3~lab1$support))
```



3a

These two are statistically correlated with each other. $R^2 = .06$, $F(1,76) = 5.13$, $p < .05$

```
knowdat <-lm(support~q2,lab1)
summary(knowdat)
```

```
##
## Call:
## lm(formula = support ~ q2, data = lab1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8345 -4.0297 -0.4972  3.9378 10.1004
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.2900     2.0022   2.642   0.0100 **
## q2            0.9349     0.4130   2.264   0.0264 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.95 on 76 degrees of freedom
## Multiple R-squared:  0.06317,    Adjusted R-squared:  0.05084
## F-statistic: 5.124 on 1 and 76 DF,  p-value: 0.02645
```

```
cor.test(lab1$support,lab1$q2)
```

```
##
##  Pearson's product-moment correlation
##
## data:  lab1$support and lab1$q2
## t = 2.2637, df = 76, p-value = 0.02645
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03050249 0.44875980
## sample estimates:
##       cor
## 0.2513273
```

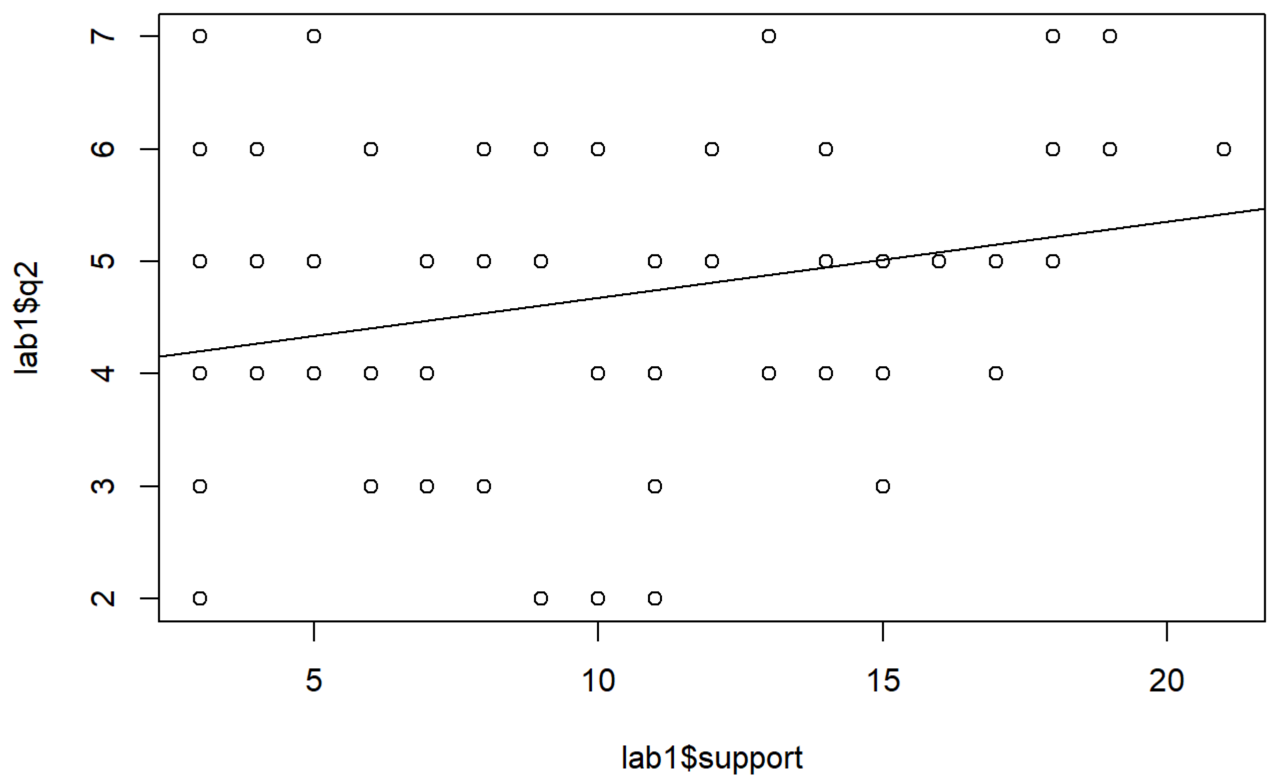
3b

```
newknow <- data.frame(q2=1)
predict(knowdat,newknow)
```

```
##      1
## 6.2249
```

3c

```
plot(lab1$support,lab1$q2)
abline(lm(lab1$q2~lab1$support))
```



4

```r
library(stats)
cor.test(lab1$support,lab1$qb)
```

```
##
##  Pearson's product-moment correlation
##
## data:  lab1$support and lab1$qb
## t = -0.23492, df = 76, p-value = 0.8149
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2479816  0.1967729
## sample estimates:
##         cor
## -0.02693739
```

```r
cor.test(lab1$support,lab1$q3)
```

```
##
##  Pearson's product-moment correlation
##
## data:  lab1$support and lab1$q3
## t = 5.8563, df = 76, p-value = 1.137e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3825674 0.6940346
## sample estimates:
##       cor
## 0.5576258
```

```
cor.test(lab1$support,lab1$q2)
```

```
##
##  Pearson's product-moment correlation
##
## data:  lab1$support and lab1$q2
## t = 2.2637, df = 76, p-value = 0.02645
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03050249 0.44875980
## sample estimates:
##       cor
## 0.2513273
```

Appears to me that personal benefit is the strongest correlation, because the correlation is the largest.

5

I do not think it is a good analysis because ethnicity is a factors instead of numeric. Continuous variables are necessry for correlations.

6

Restriction of range occurs with continuous variables that are not acurately/completely represented. I imagine that age may cause a problem regarding restriction of range. The dataset has a small range with all the 20 year olds.