

# Hw 3

*Ben Chu*

*April 1, 2018*

Loading some data, packages, and function.

```

library(tidyverse)
library(psych)
library(purrr)
library(stats)
library(DescTools)
library(car)
library(jtools)
library(lmSupport)
library(lm.beta)
library(lsr)
load("C:/Users/Branly Mclanbry/Downloads/GSS1991HW.RData")
load("C:/Users/Branly Mclanbry/Downloads/TOP2003.RData")
hw1 <- GSS1991HW %>%
  janitor::clean_names()
options(contrasts = c('contr.helmert', 'contr.poly'))
pphehe <- function(x,var) {
  (qqnorm(x, main = var))
  (qqline(x))
}
denss <- function(x,var) {
  plot(density(x), main = var)
}

skurt.1 <-function(x,var) {
  print(var)
  print(round(DescTools::Kurt(x, method = 2, conf.level = .99, R = 2000),2))
  print(round(DescTools::Skew(x, method = 2, conf.level = .99, R = 2000),2))
}

transformer <- function(x,var){
  squareroot <- (x+1)^.5
  inverse <- 1/(x+1)
  log <- log10(x+1)
  print(var)
  print("squareroot")
  print(round(DescTools::Skew(squareroot,na.rm=TRUE, method=2,conf.level=.99, R=2000),2))
  print(round(DescTools::Kurt(squareroot,na.rm=TRUE, method=2,conf.level=.99, R=2000),2))
  print("log")
  print(round(DescTools::Skew(log,na.rm=TRUE, method=2,conf.level=.99, R = 2000),2))
  print(round(DescTools::Kurt(log,na.rm=TRUE, method=2,conf.level=.99, R = 2000),2))
  print("inverse")
  print(round(DescTools::Skew(inverse,na.rm=TRUE, method=2,conf.level=.99, R = 2000),2))
  print(round(DescTools::Kurt(inverse,na.rm=TRUE, method=2,conf.level=.99, R = 2000),2))
}

p_list <- list(hw1$educ,hw1$maeduc,hw1$prestg80)
p_names <- names(hw1[2:4])

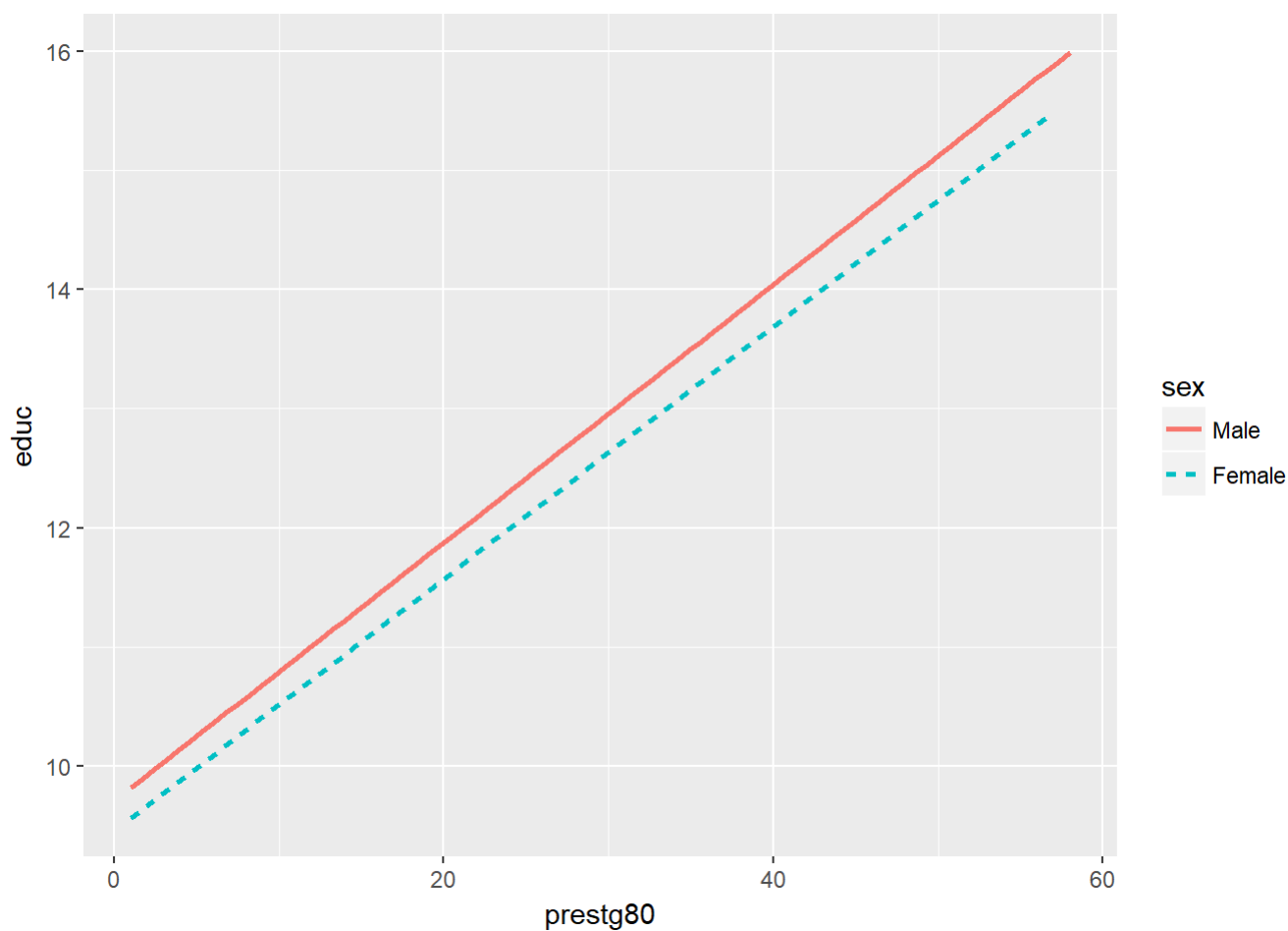
```

# HW1

```
gss.dat <- lm(prestg80~sex*educ, hw1)
summ(gss.dat,center = TRUE, digits = 5, confint = TRUE)
```

```
## MODEL INFO:
## Observations: 1162
## Dependent Variable: prestg80
##
## MODEL FIT:
## F(3,1158) = 134.3184, p = 0
## R-squared = 0.25815
## Adj. R-squared = 0.25622
##
## Standard errors: OLS
##           Est.      2.5%    97.5%   t val.      p
## (Intercept) 26.80581 25.80841 27.80322 52.67521 0 ***
## sex         -0.23471 -1.55352  1.0841  -0.34882 0.72729
## educ         2.14216  1.8066   2.47771 12.51221 0 ***
## sex:educ      0.45576 -0.01272  0.92423  1.90676 0.0568 .
##
## All continuous predictors are mean-centered.
```

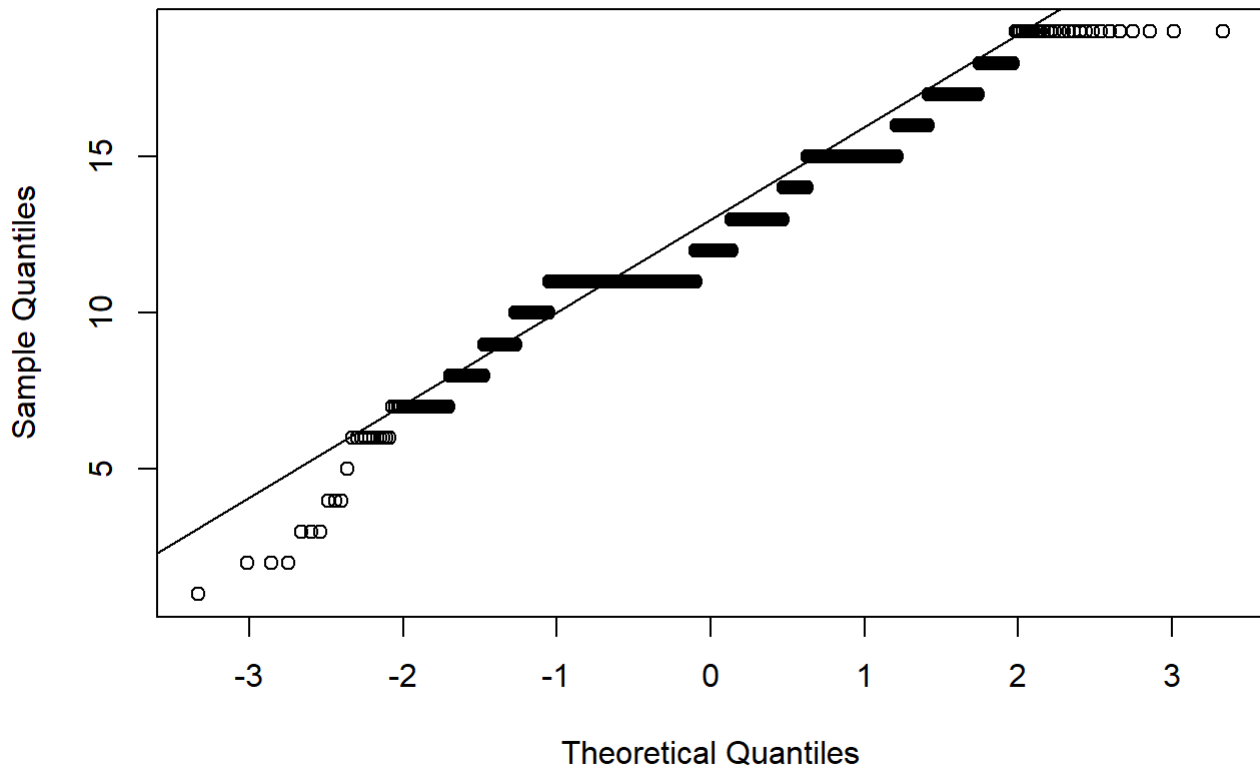
```
ggplot(gss.dat, aes(x = prestg80, educ, color = sex)) +
  geom_smooth(aes(linetype = sex), method = 'lm', se = F)
```



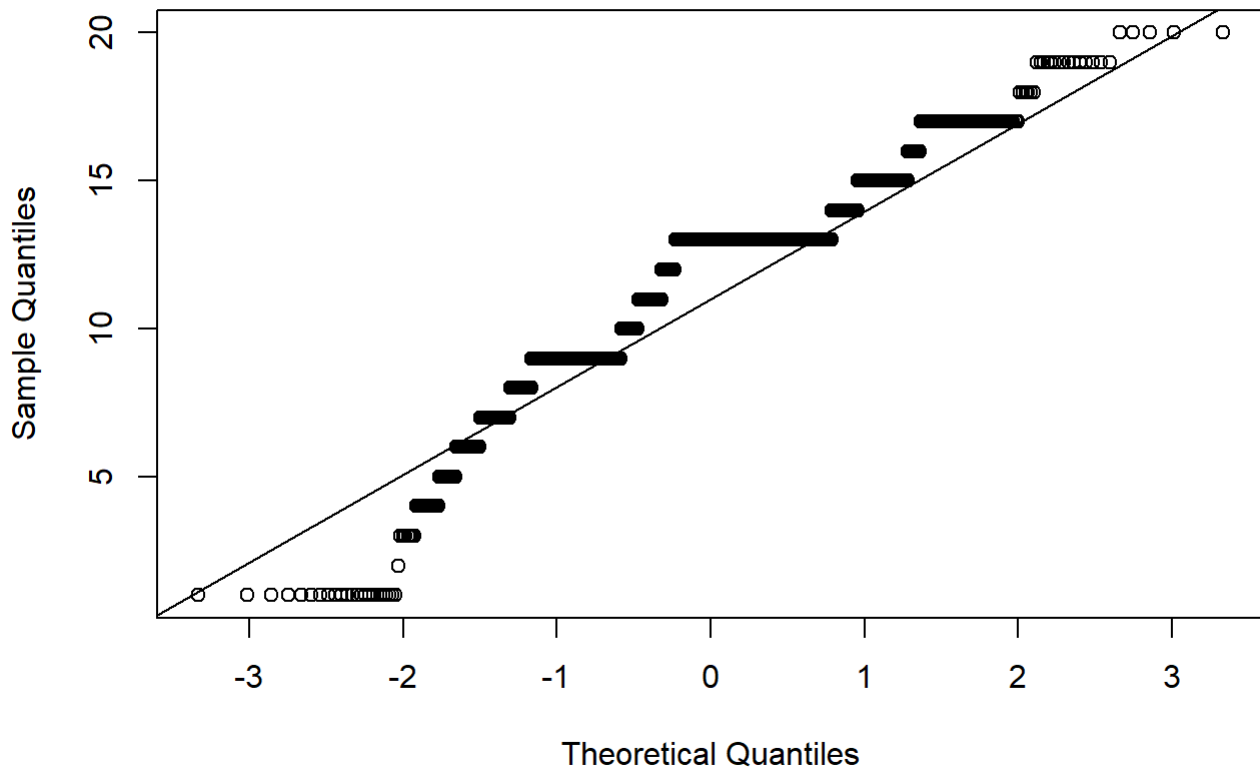
# HW2

```
walk2(p_list,p_names,pphehe)
```

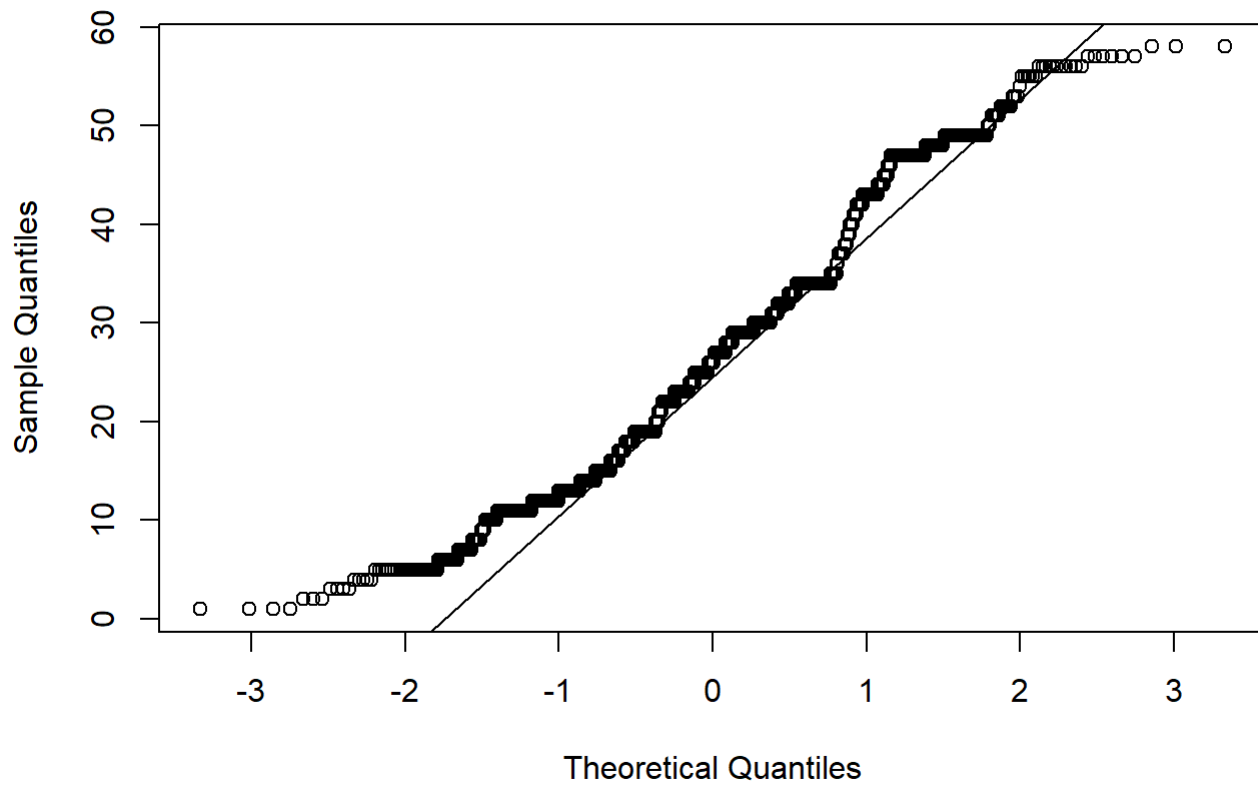
**educ**



**maeduc**

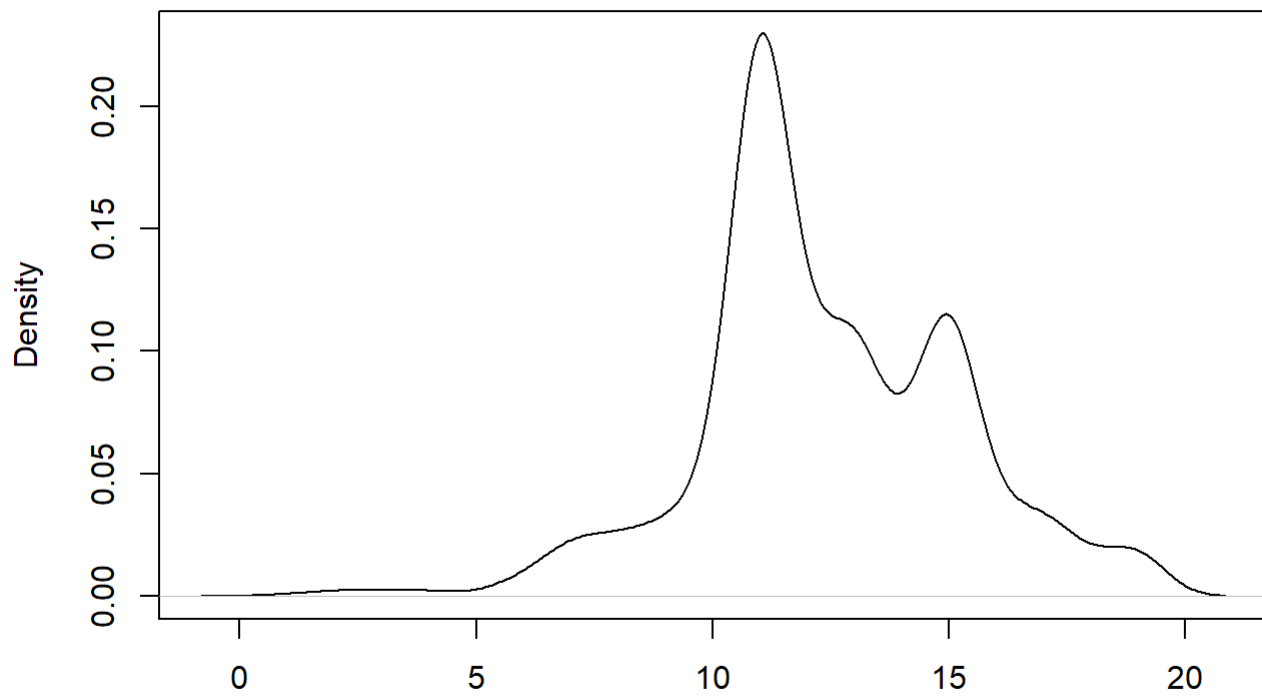


# prestg80



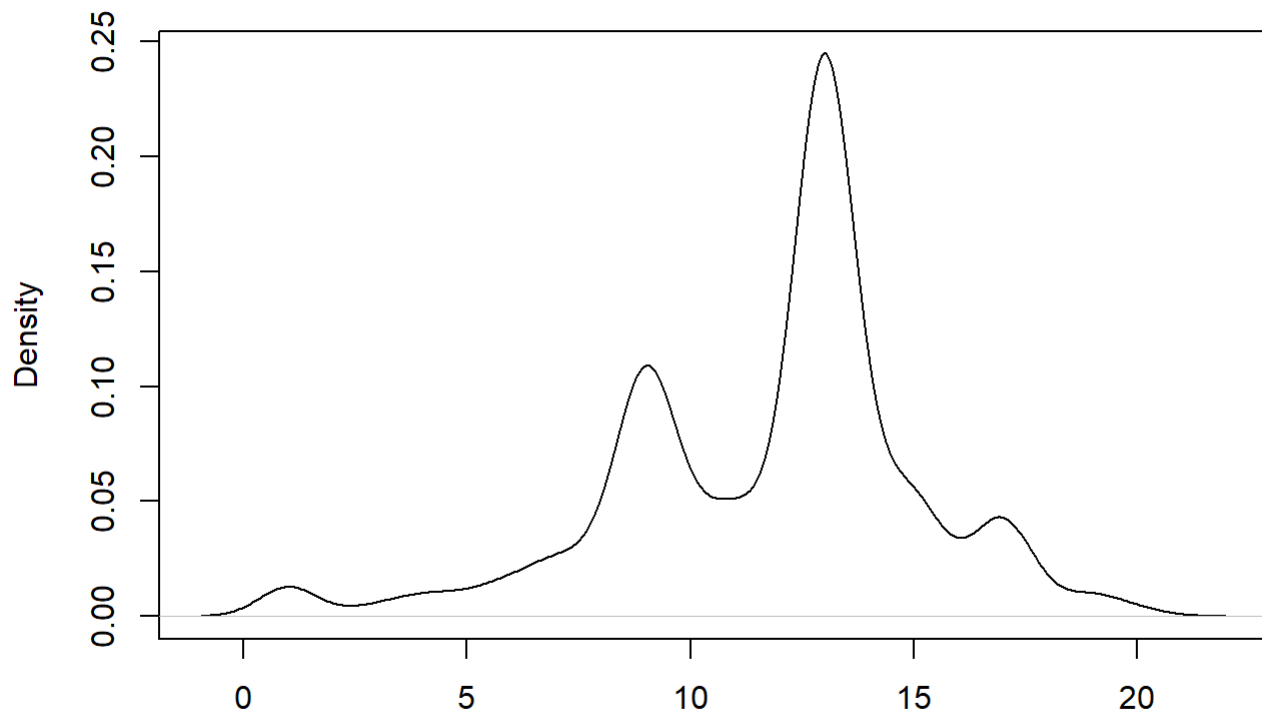
```
walk2(p_list,p_names,denss)
```

**educ**



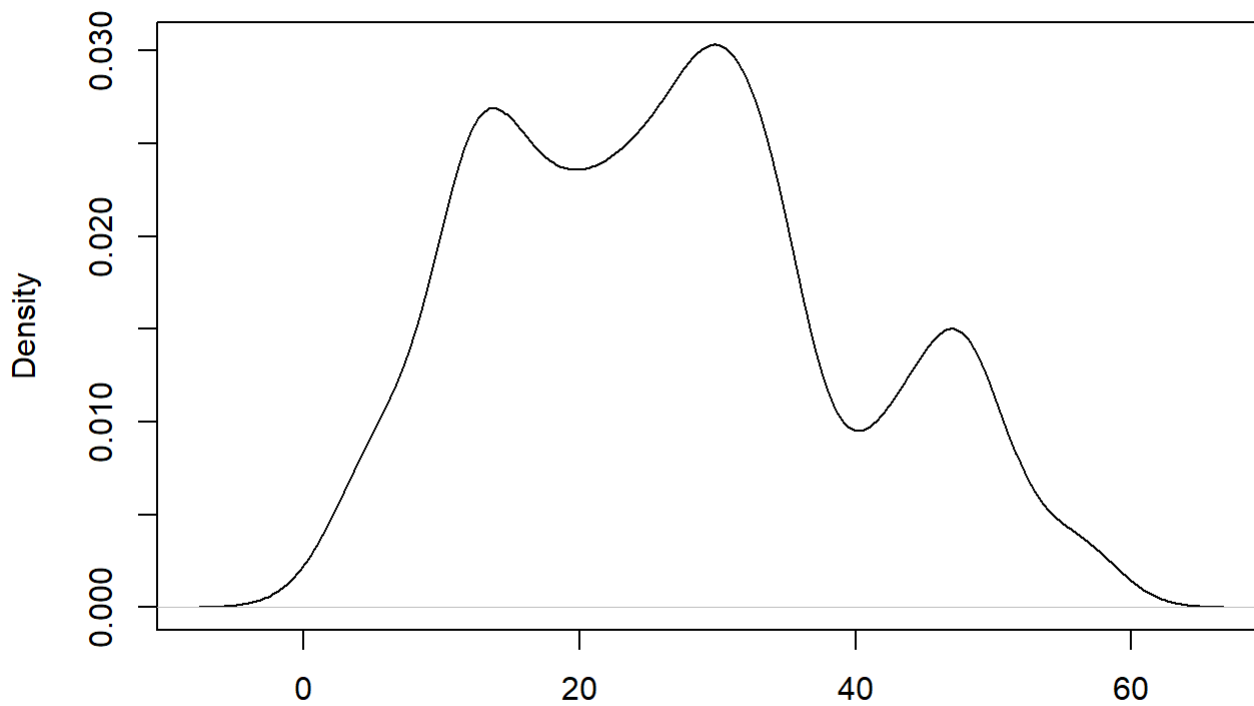
N = 1162 Bandwidth = 0.6117

**maeduc**



N = 1162 Bandwidth = 0.6549

## prestg80



N = 1162 Bandwidth = 2.878

```
walk2(p_list,p_names,skurt.1)
```

```
## [1] "educ"
##   kurt lwr.ci upr.ci
##   0.77  0.23  1.54
##   skew lwr.ci upr.ci
##  -0.09 -0.37  0.15
## [1] "maeduc"
##   kurt lwr.ci upr.ci
##   1.05  0.65  1.59
##   skew lwr.ci upr.ci
##  -0.72 -0.92 -0.52
## [1] "prestg80"
##   kurt lwr.ci upr.ci
##  -0.66 -0.83 -0.49
##   skew lwr.ci upr.ci
##   0.33  0.23  0.43
```

```
walk2(p_list,p_names,transformer)
```



```

## [1] "educ"
## [1] "squareroot"
##   skew lwr.ci upr.ci
## -0.72 -1.26 -0.30
##   kurt lwr.ci upr.ci
##   2.68  1.23  4.77
## [1] "log"
##   skew lwr.ci upr.ci
## -1.75 -2.70 -1.04
##   kurt lwr.ci upr.ci
##   8.84  4.06 15.18
## [1] "inverse"
##   skew lwr.ci upr.ci
##   6.81  3.82  9.24
##   kurt lwr.ci upr.ci
##  78.87 30.42 151.92
## [1] "maeduc"
## [1] "squareroot"
##   skew lwr.ci upr.ci
## -1.51 -1.73 -1.27
##   kurt lwr.ci upr.ci
##   3.61  2.79  4.64
## [1] "log"
##   skew lwr.ci upr.ci
## -2.62 -2.88 -2.34
##   kurt lwr.ci upr.ci
##   9.33  7.61 11.45
## [1] "inverse"
##   skew lwr.ci upr.ci
##   5.13  4.27  5.98
##   kurt lwr.ci upr.ci
##  28.36 18.51 42.29
## [1] "prestg80"
## [1] "squareroot"
##   skew lwr.ci upr.ci
## -0.20 -0.32 -0.06
##   kurt lwr.ci upr.ci
## -0.54 -0.72 -0.30
## [1] "log"
##   skew lwr.ci upr.ci
## -0.96 -1.24 -0.74
##   kurt lwr.ci upr.ci
##   1.20  0.38  2.52
## [1] "inverse"
##   skew lwr.ci upr.ci
##   4.92  3.12  6.03
##   kurt lwr.ci upr.ci
##  37.33 17.68 53.95

```

## Creating new variables

```
hw1 <- hw1 %>%
  mutate(maeduc_sqrt = (maeduc+1)^.5,
         prestg80_sqrt = (prestg80+1)^.5,
         max_maeduc = max(maeduc),
         max_prest = max(prestg80),
         maeduc_sqrt_ref = ((max_maeduc + 1) - maeduc)^.5,
         prestg80_sqrt_ref = ((max_prest+1)-prestg80)^.5)
```

## Running analysis

```
gss.dat.2 <- lm(prestg80_sqrt~maeduc_sqrt_ref*educ,hw1)
summ(gss.dat.2, center = TRUE, digits = 5, confint = TRUE)
```

```
## MODEL INFO:
## Observations: 1162
## Dependent Variable: prestg80_sqrt
##
## MODEL FIT:
## F(3,1158) = 116.7009, p = 0
## R-squared = 0.23215
## Adj. R-squared = 0.23016
##
## Standard errors: OLS
##
```

	Est.	2.5%	97.5%	t val.	p
## (Intercept)	5.06491	4.99526	5.13455	142.53734	0 ***
## maeduc_sqrt_ref	0.16992	0.04222	0.29762	2.60803	0.00922 **
## educ	0.24125	0.21474	0.26776	17.83596	0 ***
## maeduc_sqrt_ref:educ	-0.03874	-0.07249	-0.00498	-2.24904	0.0247 *

```
##
## All continuous predictors are mean-centered.
```

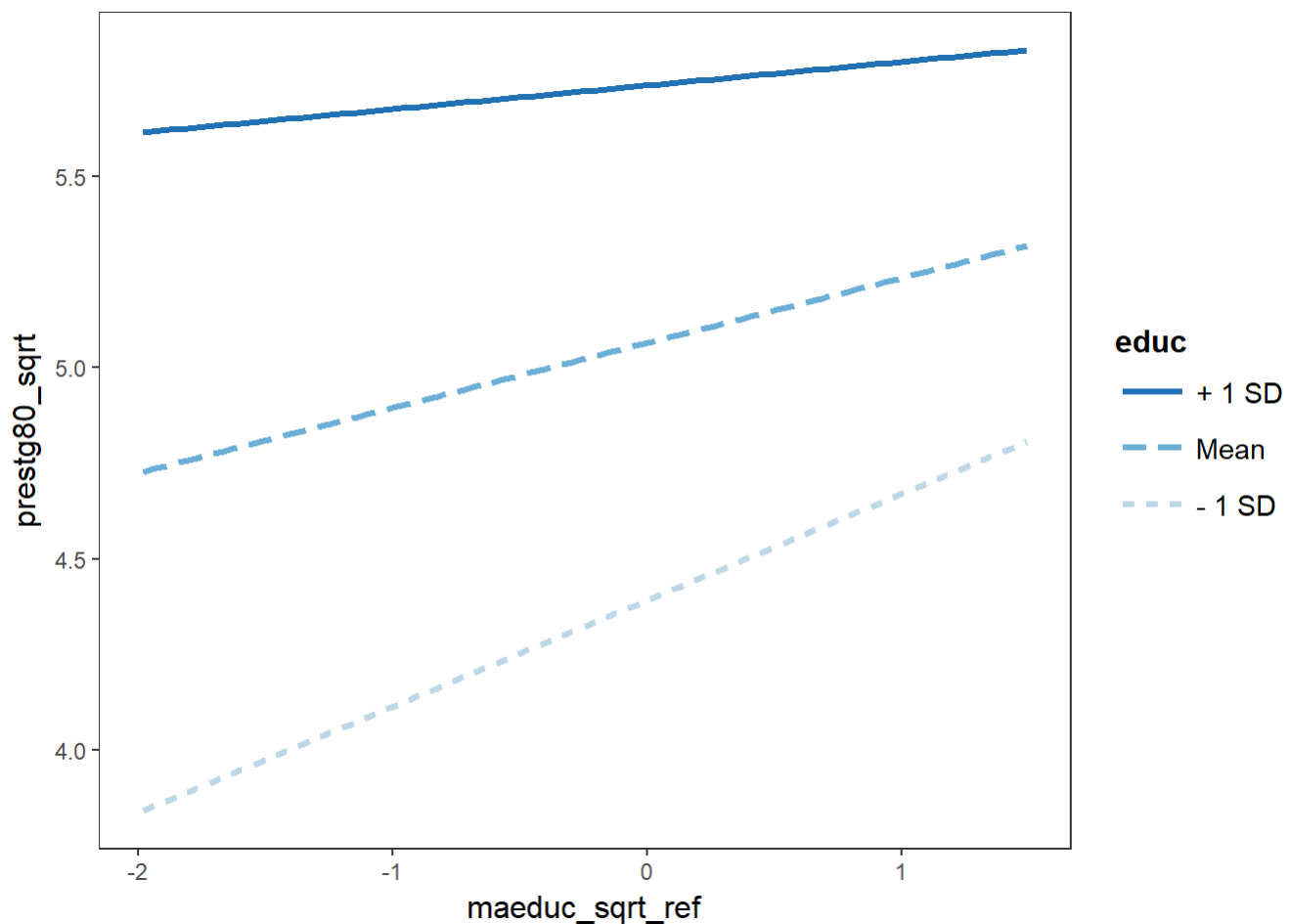
```
summary(gss.dat.2)
```

```
##
## Call:
## lm(formula = prestg80_sqrt ~ maeduc_sqrt_ref * educ, data = hw1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.1182 -0.7232  0.1374  0.8395  2.7321
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.13134    0.76112   0.173   0.8630
## maeduc_sqrt_ref    0.65066    0.22914   2.840   0.0046 **
## educ              0.35672    0.05546   6.432 1.84e-10 ***
## maeduc_sqrt_ref:educ -0.03874    0.01722  -2.249   0.0247 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.146 on 1158 degrees of freedom
## Multiple R-squared:  0.2321, Adjusted R-squared:  0.2302
## F-statistic: 116.7 on 3 and 1158 DF, p-value: < 2.2e-16
```

```
sim_slopes(gss.dat.2,educ,maeduc_sqrt_ref,johnson_neyman = FALSE, cont.int = TRUE, centered = c(
'educ','maeduc_sqrt_ref'), digits = 5)
```

```
## SIMPLE SLOPES ANALYSIS
##
## Slope of educ when maeduc_sqrt_ref = 0.57127 (+ 1 SD):
##   Est.   S.E.      p
## 0.21912 0.01518 0.00000
##
## Slope of educ when maeduc_sqrt_ref = 0 (Mean):
##   Est.   S.E.      p
## 0.24125 0.01353 0.00000
##
## Slope of educ when maeduc_sqrt_ref = -0.57127 (- 1 SD):
##   Est.   S.E.      p
## 0.26338 0.01814 0.00000
```

```
interact_plot(gss.dat.2,maeduc_sqrt_ref, educ,centered = c('educ','maeduc_sqrt_ref'))
```



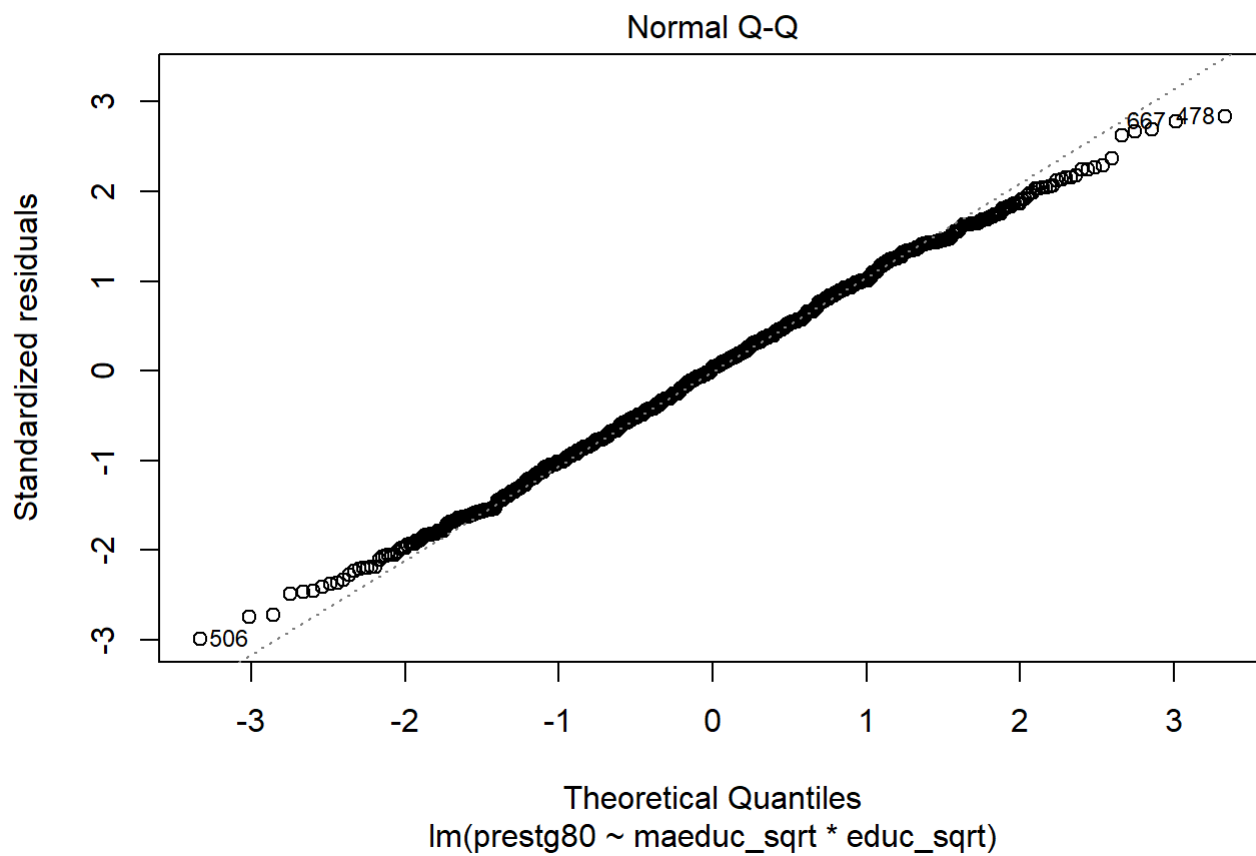
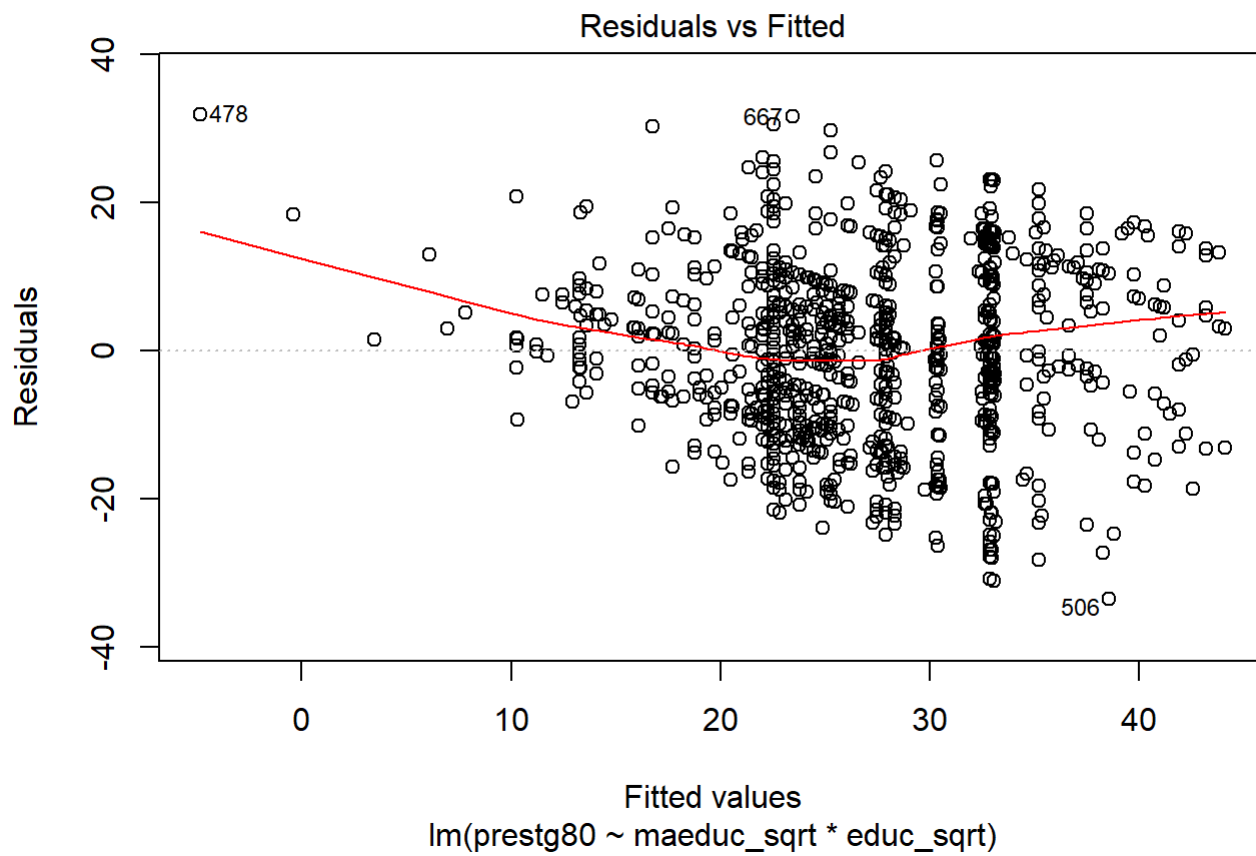
```
lm.beta(gss.dat.2)
```

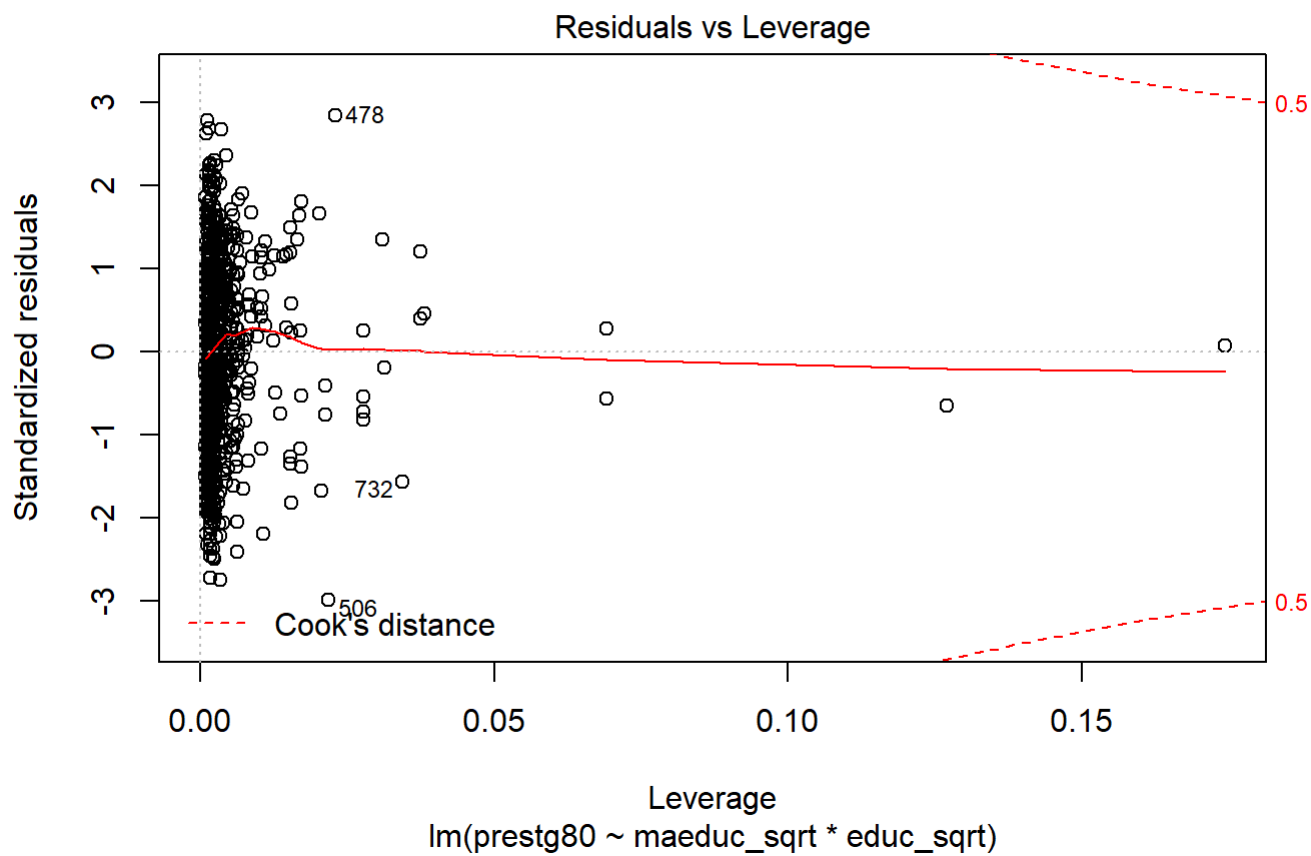
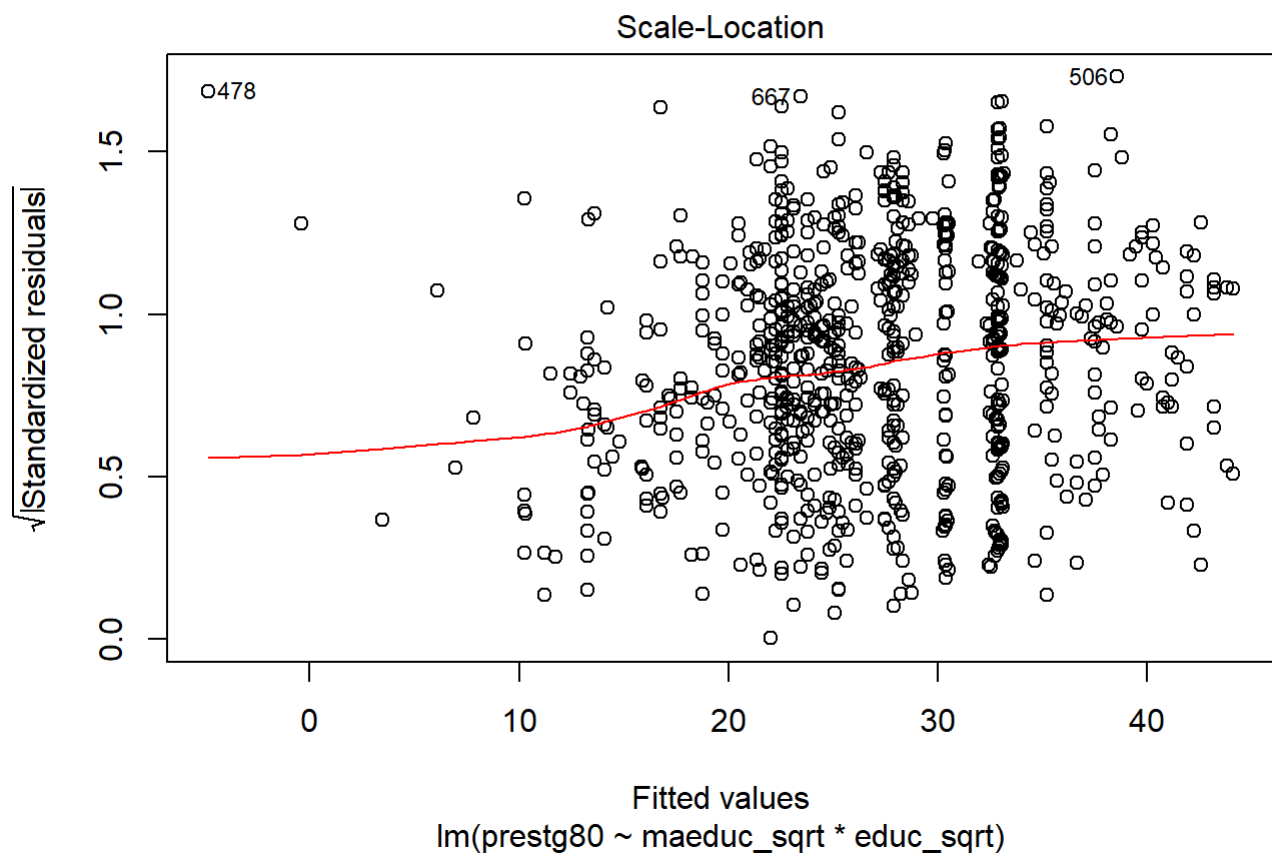
```
##  
## Call:  
## lm(formula = prestg80_sqrt ~ maeduc_sqrt_ref * educ, data = hw1)  
##  
## Standardized Coefficients::  
##           (Intercept)      maeduc_sqrt_ref           educ  
##           0.0000000      0.2846186      0.7615599  
## maeduc_sqrt_ref:educ  
##           -0.2644871
```

### h3 Addressing residuals and multicollinearity

```
plot(gss.dat.2)
```







```
n.2 <- 1162
hat <- hatvalues(gss.dat.2)
mahun <- ((n.2-1)*(hat))-1
tail(sort(mahun),10)
```

```
##          936          200          732          621          1157          45          647
## 34.90754 35.43803 38.97695 42.54055 42.54055 43.39291 79.27967
##          183          634          641
## 79.33560 146.57972 201.44995
```

```
1-pchisq(201.45,df = 3)
```

```
## [1] 0
```

```
vif(gss.dat.2)
```

```
##          maeduc_sqrt          educ_sqrt maeduc_sqrt:educ_sqrt
##          41.42999          20.23956          83.41062
```

```
1/vif(gss.dat.2)
```

```
##          maeduc_sqrt          educ_sqrt maeduc_sqrt:educ_sqrt
##          0.02413710          0.04940818          0.01198888
```

## hw4

loading in data

```
load("C:/Users/Branly McInbry/Downloads/employee.RData")
hw2 <- employee
```

Ancova model

```
salary.dat <- aov(salbegin ~educ + minority,hw2)
Anova(salary.dat, type = "III")
```



```
## Anova Table (Type III tests)
##
## Response: salbegin
##           Sum Sq  Df    F value Pr(>F)
## (Intercept) 2.8383e+10  1 1195.9651 <2e-16 ***
## educ        1.7585e+10  9   82.3307 <2e-16 ***
## minority    4.1526e+07  1    1.7498 0.1866
## Residuals   1.0988e+10 463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
salary.dat.2 <- aov(salbegin~educ * minority,hw2)
Anova(salary.dat.2)
```

```
## Anova Table (Type II tests)
##
## Response: salbegin
##           Sum Sq  Df F value Pr(>F)
## educ        1.7585e+10  9 82.9828 <2e-16 ***
## minority    4.1526e+07  1  1.7636 0.1848
## educ:minority 2.5118e+08  7  1.5239 0.1569
## Residuals   1.0737e+10 456
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
etaSquared(salary.dat.2)
```

```
##           eta.sq eta.sq.part
## educ        0.600155086 0.620898929
## minority    0.001417217 0.003852674
## educ:minority 0.008572326 0.022859056
```

## hw4.c

```
salary.lm<- lm(salbegin ~educ + minority,hw2)
summ(salary.lm, center = TRUE, digits = 5, confint = TRUE)
```

```
## MODEL INFO:
## Observations: 474
## Dependent Variable: salbegin
##
## MODEL FIT:
## F(10,463) = 77.16408, p = 0
## R-squared = 0.62499
## Adj. R-squared = 0.61689
##
## Standard errors: OLS
##
```

	Est.	2.5%	97.5%	t val.		p
## (Intercept)	24859.05353	23514.63903	26203.46803	36.24094	0	***
## educ1	90.40843	-651.20705	832.02392	0.23893	0.81126	
## educ2	799.89444	-523.23021	2123.01909	1.1849	0.23667	
## educ3	406.77672	-5.43726	818.9907	1.93411	0.05371	.
## educ4	1574.32607	1245.93415	1902.71798	9.39616	0	***
## educ5	1829.20231	1326.9381	2331.46651	7.13802	0	***
## educ6	2052.02282	1579.91752	2524.12811	8.51905	0	***
## educ7	1847.8057	1592.02557	2103.58582	14.15916	0	***
## educ8	1598.18701	842.37979	2353.99423	4.14443	4e-05	***
## educ9	1404.54961	443.9629	2365.13631	2.86582	0.00435	**
## minority	-725.83311	-1801.29851	349.6323	-1.32278	0.18656	

```
##
## All continuous predictors are mean-centered.
```

## 5

arcsine

```
h5 <- TOP2003 %>% janitor::clean_names()
h5.1 <- lm(quiz2 ~ current, data = h5)
summary(h5.1)
```

```
##
## Call:
## lm(formula = quiz2 ~ current, data = h5)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-3.9928	-1.4414	0.4889	1.0205	2.8138

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	3.009	2.897	1.039	0.310
## current	4.719	3.434	1.374	0.183

```
##
## Residual standard error: 1.758 on 23 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared: 0.07586, Adjusted R-squared: 0.03568
## F-statistic: 1.888 on 1 and 23 DF, p-value: 0.1827
```

```
h5.2 <- lm(quiz2 ~ current + condit, data = h5)
summary(h5.2)
```

```
##
## Call:
## lm(formula = quiz2 ~ current + condit, data = h5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.181 -0.761  0.120  1.050  2.200
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.7690     2.7559   1.368   0.1852
## current       6.0999     3.3096   1.843   0.0788 .
## condit       -1.3690     0.6919  -1.979   0.0605 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.657 on 22 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.2155, Adjusted R-squared:  0.1442
## F-statistic: 3.021 on 2 and 22 DF,  p-value: 0.06929
```

```
h5.3 <- lm(quiz2 ~ current * condit, data = h5)
summary(h5.3)
```

```
##
## Call:
## lm(formula = quiz2 ~ current * condit, data = h5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0566 -1.1405  0.3545  1.1656  2.2841
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.378     9.284   1.226   0.234
## current        -2.840    10.930  -0.260   0.798
## condit         -7.927     7.668  -1.034   0.313
## current:condit   7.658     8.917   0.859   0.400
##
## Residual standard error: 1.667 on 21 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.2421, Adjusted R-squared:  0.1338
## F-statistic: 2.236 on 3 and 21 DF,  p-value: 0.1139
```

## hw6

### Analysis

```
cor.dat <- cor(grants)
round(cor.dat,2)
```

```
##          SUBMIT QUALITY UNIVERS MONEY
## SUBMIT    1.00   -0.80   -0.60 -0.24
## QUALITY  -0.80    1.00    0.72  0.45
## UNIVERS  -0.60    0.72    1.00  0.56
## MONEY    -0.24    0.45    0.56  1.00
```

```
pwr.MRC(-.80,-.60,-.24,.72,.45,.56,110)
```

```
## [1] "Sample size is 110"
## [1] "Power R2 = 1"
## [1] "Power b1 = 1"
## [1] "Power b2 = 0.374"
## [1] "Power b3 = 0.805"
```

```
pwr.MRC_all(-.80,-.60,-.24,.72,.45,.56,320)
```

```
## [1] "Sample size is 320"
## [1] "Power R2 = 1"
## [1] "Power b1 = 1"
## [1] "Power b2 = 0.8189"
## [1] "Power b3 = 0.998"
## [1] "Proportion Rejecting None = 0"
## [1] "Proportion Rejecting One = 0.0014"
## [1] "Proportion Rejecting Two = 0.1803"
## [1] "Power ALL (Proportion Rejecting All) = 0.8183"
```

```
pwr.MRC_all(-.80,-.60,-.24,.72,.45,.56,130)
```

```
## [1] "Sample size is 130"
## [1] "Power R2 = 1"
## [1] "Power b1 = 1"
## [1] "Power b2 = 0.427"
## [1] "Power b3 = 0.8641"
## [1] "Proportion Rejecting None = 0"
## [1] "Proportion Rejecting One = 0.1105"
## [1] "Proportion Rejecting Two = 0.4879"
## [1] "Power ALL (Proportion Rejecting All) = 0.4016"
```