Assignment 3

Steve Hawley

December 12, 2018

## Section 0: Load the data and packages

library(tidyverse) #for data management  
library(foreign) #for importing spss data  
library(car) #for leven's test  
library(sjstats) #for eta squared  
library(effects) #for adjusted means in ancova  
library(BaylorEdPsych) #for logisitic regression effect size  
#library(QuantPsyc) # For Beta values -- conflicts with dplyr  
#library(ppcor) #for pr values -- conflicts with dplyr  
  
# import the spss file. Label the factors  
data<-read.spss("Data+for+assignments.sav",to.data.frame = TRUE, add.undeclared.levels = "no")  
# take a look at the data. See what the fields are  
head(data)

## code gender age marital educ smoke work polaff  
## 1 146 female 35 Married 21 Never Smoked Full-Time Democrat  
## 2 147 male 37 Married 18 Never Smoked Full-Time Independent  
## 3 148 female 43 Married 22 Never Smoked Full-Time Independent  
## 4 149 male 35 Married 20 Quit Smoking Full-Time Democrat  
## 5 150 female 35 Married 20 Quit Smoking Part-Time Independent  
## 6 151 female 35 Never Married 19 Never Smoked Part-Time Independent  
## depress exer eat satcurwt satwt18 health  
## 1 Sometimes Rarely Often 6 5 7  
## 2 Sometimes Sometimes Often 3 3 7  
## 3 Rarely Routinely Routinely 9 7 9  
## 4 Sometimes Routinely Sometimes 9 9 10  
## 5 Rarely Routinely Routinely 9 4 10  
## 6 Rarely Routinely Routinely 10 10 9  
## qolcur qol18  
## 1 generally satisfied, pleased generally satisfied, pleased  
## 2 generally satisfied, pleased generally satisfied, pleased  
## 3 very happy most of time very happy most of the time  
## 4 very happy most of time generally satisfied, pleased  
## 5 very happy most of time generally satisfied, pleased  
## 6 very happy most of time very happy most of the time  
## gift winter life confid  
## 1 invest it with brokerage firm beachfront condo in Hawaii 87 56  
## 2 invest it with brokerage firm beachfront condo in Hawaii 83 65  
## 3 invest it with brokerage firm beachfront condo in Hawaii 97 68  
## 4 invest it with brokerage firm beachfront condo in Hawaii 87 65  
## 5 invest it with brokerage firm beachfront condo in Hawaii 108 72  
## 6 invest it with brokerage firm beachfront condo in Hawaii 112 80

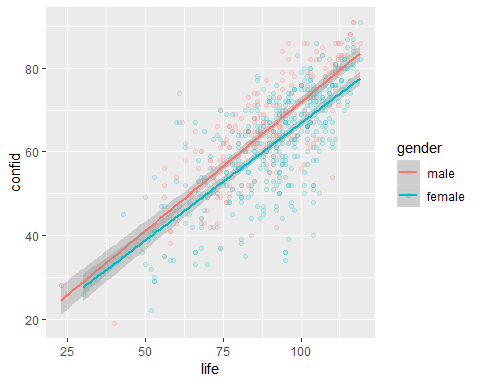
#inspect the data types  
str(data)

## 'data.frame': 701 obs. of 20 variables:  
## $ code : num 146 147 148 149 150 151 152 153 154 155 ...  
## $ gender : Factor w/ 2 levels "male","female": 2 1 2 1 2 2 2 NA 1 1 ...  
## $ age : num 35 37 43 35 35 35 43 45 52 69 ...  
## $ marital : Factor w/ 6 levels "Never Married",..: 2 2 2 2 2 1 2 2 2 2 ...  
## $ educ : num 21 18 22 20 20 19 18 18 18 15 ...  
## $ smoke : Factor w/ 3 levels "Never Smoked",..: 1 1 1 2 2 1 1 1 1 3 ...  
## $ work : Factor w/ 3 levels "Unemployed","Part-Time",..: 3 3 3 3 2 2 2 3 2 1 ...  
## $ polaff : Factor w/ 3 levels "Republican","Democrat",..: 2 3 3 2 3 3 1 1 2 1 ...  
## $ depress : Factor w/ 4 levels "Rarely","Sometimes",..: 2 2 1 2 1 1 1 1 1 3 ...  
## $ exer : Factor w/ 4 levels "Rarely","Sometimes",..: 1 2 4 4 4 4 2 2 2 2 ...  
## $ eat : Factor w/ 4 levels "Rarely","Sometimes",..: 3 3 4 2 4 4 4 4 4 2 ...  
## $ satcurwt: num 6 3 9 9 9 10 2 7 7 6 ...  
## ..- attr(\*, "value.labels")= Named chr "10" "1"  
## .. ..- attr(\*, "names")= chr "Very Satisfied" "Very Dissatisfied"  
## $ satwt18 : num 5 3 7 9 4 10 8 9 9 9 ...  
## ..- attr(\*, "value.labels")= Named chr "10" "1"  
## .. ..- attr(\*, "names")= chr "Very Satisfied" "Very Dissatisfied"  
## $ health : num 7 7 9 10 10 9 9 8 8 5 ...  
## ..- attr(\*, "value.labels")= Named chr "10" "1"  
## .. ..- attr(\*, "names")= chr "Very Healthy" "Very Sick"  
## $ qolcur : Factor w/ 6 levels "very dissatisfied, unhappy most of time",..: 4 4 5 5 5 5 5 5 5 3 ...  
## $ qol18 : Factor w/ 6 levels "very dissatisfied, unhappy most of time",..: 4 4 5 4 4 5 4 4 3 5 ...  
## $ gift : Factor w/ 4 levels "invest it with brokerage firm",..: 1 1 1 1 1 1 2 1 2 1 ...  
## $ winter : Factor w/ 4 levels "beachfront condo in Hawaii",..: 1 1 1 1 1 1 1 1 3 4 ...  
## $ life : num 87 83 97 87 108 112 102 103 107 56 ...  
## $ confid : num 56 65 68 65 72 80 73 77 83 35 ...  
## - attr(\*, "variable.labels")= Named chr "subject's identification number" "gender" "subject's age" "marital status" ...  
## ..- attr(\*, "names")= chr "code" "gender" "age" "marital" ...  
## - attr(\*, "codepage")= int 1252

## Section 1: ANCOVA

#### **Research Question:** Is there a difference in confidence between males and females when we control for overall life satisfaction?

#reference: http://faculty.missouri.edu/huangf/data/quantf/ancova\_in\_r\_handout.pdf  
  
#subset the data for analysis  
anc<-data %>%   
 select(gender, life, confid) %>%   
 drop\_na(gender, life, confid)  
  
#plot to look at homogeneity of regression  
ggplot(anc, aes(x=life, y=confid, color=gender)) +  
 geom\_point(alpha = 0.15) +  
 geom\_smooth(method = "lm")



#slopes look reasonably parellel   
  
#Check descriptives  
anc %>%   
 group\_by(gender) %>%   
 summarize(mean = mean(confid),  
 sd = sd(confid),  
 n = n())

## # A tibble: 2 x 4  
## gender mean sd n  
## <fct> <dbl> <dbl> <int>  
## 1 male 64.6 13.0 245  
## 2 female 61.7 12.8 411

t.test(life~gender, data = anc)

##   
## Welch Two Sample t-test  
##   
## data: life by gender  
## t = -1.8918, df = 491.45, p-value = 0.05911  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -5.3629651 0.1015519  
## sample estimates:  
## mean in group male mean in group female   
## 88.16735 90.79805

#t test is n.s. genders are not different on covariate   
  
leveneTest(anc$confid,anc$gender, center = mean)

## Levene's Test for Homogeneity of Variance (center = mean)  
## Df F value Pr(>F)  
## group 1 0.0686 0.7934  
## 654

#levene's is n.s. therefore assumption of homogeneity of variance is met  
  
#test to look for equality of slopes  
mod1 <- aov(confid~gender + life + gender:life, data = anc)  
summary(mod1)

## Df Sum Sq Mean Sq F value Pr(>F)   
## gender 1 1290 1290 18.902 1.6e-05 \*\*\*  
## life 1 64231 64231 941.329 < 2e-16 \*\*\*  
## gender:life 1 127 127 1.855 0.174   
## Residuals 652 44489 68   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

EtaSq(mod1)

## Eta^2 Partial Eta^2  
## gender 0.011710762 0.028174269  
## life 0.583196174 0.590793828  
## gender:life 0.001149316 0.002837165

#interaction of gender:life is n.s. so the slope across groups is not different  
  
mod2 <- aov(confid~life + gender, data = anc)  
summary(mod2)

## Df Sum Sq Mean Sq F value Pr(>F)   
## life 1 62522 62522 915.1 < 2e-16 \*\*\*  
## gender 1 2999 2999 43.9 7.25e-11 \*\*\*  
## Residuals 653 44616 68   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

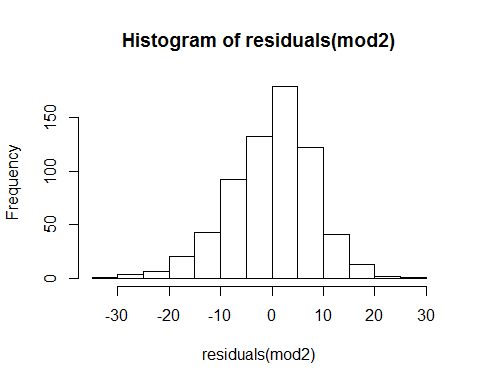
#effect size. Don't need to report results for covariate   
EtaSq(mod2)

## Eta^2 Partial Eta^2  
## life 0.56767656 0.58356732  
## gender 0.02723038 0.06298613

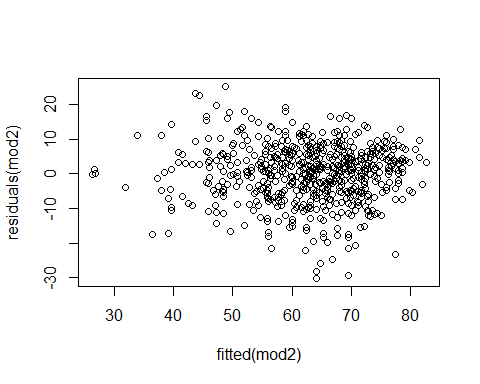
#adjusted means  
effect("gender", mod2)

##   
## gender effect  
## gender  
## male female   
## 65.60183 61.16923

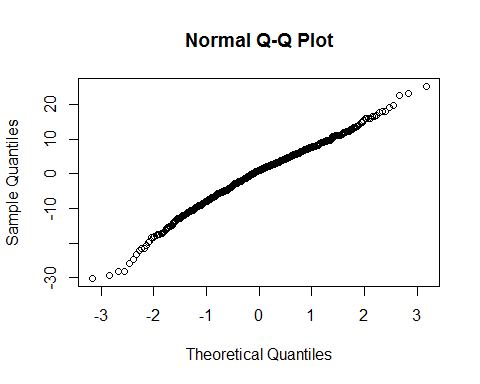
#CHECKING ASSUMPTIONS  
  
hist(residuals(mod2))



#histogram of residuals in normal  
  
plot(fitted(mod2), residuals(mod2))



qqnorm(residuals(mod2))



##########Alternative approach###########  
mod1 <- lm(data = anc, confid~life + gender + life:gender)  
Anova(mod1, type = "II")

## Anova Table (Type II tests)  
##   
## Response: confid  
## Sum Sq Df F value Pr(>F)   
## life 64231 1 941.3289 < 2.2e-16 \*\*\*  
## gender 2999 1 43.9522 7.062e-11 \*\*\*  
## life:gender 127 1 1.8551 0.1737   
## Residuals 44489 652   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Interaction is not significant, so the slope across groups is not different  
mod2 <- lm(data = anc, confid~life + gender)  
Anova(mod2, type = "II")

## Anova Table (Type II tests)  
##   
## Response: confid  
## Sum Sq Df F value Pr(>F)   
## life 64231 1 940.098 < 2.2e-16 \*\*\*  
## gender 2999 1 43.895 7.251e-11 \*\*\*  
## Residuals 44616 653   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##### Results (APA style)

To investigate whether gender has any effect on confidence, a one-way ANCOVA was conducted. The independant variable, gender, included two levels, male and female. The dependant variable was a confidence score from a self reported assessment and the covariate was a life satisfaction score. A preliminary analysis evaluating the homogeneity-of-slopes assumption indicated that the relationship between the covariate and the dependent variable did not differe significantly as a function of the independent variable, F(1,652) = 1.855, MSE = 127, p = 0.174, partial eta\_squared = 0.003. The ANCOVA was significant, F(1,653) = 43.9, MSE = 2999, p <0.01, partial eta\_squared = 0.06. The strength of relationship between gender and confidence was small, as assessed by partial eta\_squared, with the gender accounting for 6% of the variance in the dependent variable, holding life score constant. The means of the confidence scores adjusted for initial differences in life satisfaction scores were ordered in the following way: males (M = 65.60) and females (M = 61.17). Follow up tests were not conducted as gender contained only two factors.

## Section 2: Logistic Regression

#### **Research Question:** To what extent does the age and working status predict the politcal affiliation of a participant?

#subset the data for analysis -- DV = polaff ("independent" factor filtered out); cat IV = work; contin IV = age  
lr <- data %>%   
 select(work, age, polaff) %>%   
 filter(polaff=="Republican" | polaff=="Democrat") %>%   
 drop\_na(work, age)  
#dropping the empty factor  
lr$polaff <- factor(lr$polaff)  
  
#let's have a look to see if there's any issues  
summary(lr)

## work age polaff   
## Unemployed: 56 Min. :15.00 Republican:134   
## Part-Time : 81 1st Qu.:29.00 Democrat :221   
## Full-Time :218 Median :38.00   
## Mean :39.32   
## 3rd Qu.:47.00   
## Max. :95.00

#nothing out of the ordinary  
  
#set up the model and get the output  
model <- glm(polaff~work + age,  
 family = binomial(link = 'logit'),  
 data = lr)  
summary(model)

##   
## Call:  
## glm(formula = polaff ~ work + age, family = binomial(link = "logit"),   
## data = lr)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7991 -1.3104 0.8571 0.9832 1.2341   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.431671 0.437420 -0.987 0.3237   
## workPart-Time 0.538633 0.370211 1.455 0.1457   
## workFull-Time 0.140460 0.311137 0.451 0.6517   
## age 0.018703 0.008184 2.285 0.0223 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 470.59 on 354 degrees of freedom  
## Residual deviance: 462.97 on 351 degrees of freedom  
## AIC: 470.97  
##   
## Number of Fisher Scoring iterations: 4

modelx <- glm(polaff~work,  
 family = binomial(link = 'logit'),  
 data = lr)  
summary(modelx)

##   
## Call:  
## glm(formula = polaff ~ work, family = binomial(link = "logit"),   
## data = lr)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5333 -1.3639 0.8592 1.0017 1.0284   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.36101 0.27163 1.329 0.184  
## workPart-Time 0.44546 0.36282 1.228 0.220  
## workFull-Time 0.06744 0.30493 0.221 0.825  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 470.59 on 354 degrees of freedom  
## Residual deviance: 468.39 on 352 degrees of freedom  
## AIC: 474.39  
##   
## Number of Fisher Scoring iterations: 4

#is there a significant reduction in error by adding the predictors of work and age (i.e., Null deviance vs. Residual deviance)?  
chidiff <- model$null.deviance - model$deviance  
dfdiff <- model$df.null - model$df.residual  
chidiff

## [1] 7.629628

dfdiff

## [1] 3

pchisq(chidiff,dfdiff, lower.tail = F)

## [1] 0.05431966

#difference is n.s. (though trending towards significance) therefore predictors do not improve model  
  
#Effect size  
PseudoR2(model)

## McFadden Adj.McFadden Cox.Snell Nagelkerke   
## 0.016212738 -0.005036973 0.021262603 0.028953866   
## McKelvey.Zavoina Effron Count Adj.Count   
## 0.028752016 0.021058273 0.600000000 -0.059701493   
## AIC Corrected.AIC   
## 470.965022449 471.079308163

#checking which group is the lower group  
table(lr$polaff)

##   
## Republican Democrat   
## 134 221

#look at the % correct from the model  
#fitted values give probability of being Republican  
correct <- model$fitted.values  
#Thresholding values into bins where 0.5 is equal split between groups  
binarycorrect <- ifelse(correct >0.5,1,0)  
binarycorrect <- factor(binarycorrect,  
 levels = c(0,1),  
 labels = c("Republican pred", "Democrat pred"))  
  
table(lr$polaff, binarycorrect)

## binarycorrect  
## Republican pred Democrat pred  
## Republican 6 128  
## Democrat 14 207

#from the table, 6 republicans predicted correctly, 14 incorrectly  
#207 Democrats predicted correctly, 128 predicted incorrectly  
  
#To get probabiliy of correctly guessing using model:  
#Republican  
6/(6+128)\*100

## [1] 4.477612

#Democrat  
207/(14+207)\*100

## [1] 93.66516

#Overall  
(6+207)/nrow(lr)\*100

## [1] 60

#to get Odds Ratios  
exp(model$coefficients)

## (Intercept) workPart-Time workFull-Time age   
## 0.649423 1.713663 1.150803 1.018879

##### Results (APA style)

A logistic regression was conducted to explore whether age and working status are related to political affiliation. The model had a non-significant fit X2(3) = 7.63, p = 0.054, Nagelkerke R2 = 0.029. The model correctly classifies 60% of participants in the sample. The classification of Democrats (93.7%) was much better than Republicans (4.5%). Of two predictor variables considered in this analysis, only age was significantly related to political affiliation. Specifically, with each additional year in age, there was approximately a 2% increased chance of a Democrat political affiliation.

## Section 3: Hierarchical Multiple Regression

#### **Research Question:** After controlling for demographic variables (age and education), does life satisfaction of the participant predict their confidence?

#For this section I followed this tutorial:  
#https://www.youtube.com/watch?v=zFEP-lJ1LD0&feature=youtu.be  
  
#subset the data for this analysis  
hmr<- data %>%   
 select(age,educ,life,confid) %>%   
 drop\_na(age,educ,life,confid)  
  
#Let's take a look for anythign weird  
summary(hmr)

## age educ life confid   
## Min. :15.00 Min. : 7.00 Min. : 23.00 Min. :19.00   
## 1st Qu.:29.00 1st Qu.:14.00 1st Qu.: 80.00 1st Qu.:55.00   
## Median :36.00 Median :16.00 Median : 92.50 Median :64.00   
## Mean :37.92 Mean :16.55 Mean : 90.03 Mean :63.04   
## 3rd Qu.:46.00 3rd Qu.:19.00 3rd Qu.:103.00 3rd Qu.:72.00   
## Max. :95.00 Max. :30.00 Max. :119.00 Max. :91.00

#running the final model to get assumption checks  
output <- lm(confid~life + educ + age, data=hmr)  
summary(output)

##   
## Call:  
## lm(formula = confid ~ life + educ + age, data = hmr)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31.340 -4.501 0.681 5.548 28.403   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.67131 2.40714 4.433 1.1e-05 \*\*\*  
## life 0.58312 0.02039 28.592 < 2e-16 \*\*\*  
## educ -0.14549 0.10121 -1.437 0.1511   
## age 0.06010 0.02609 2.303 0.0216 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.457 on 622 degrees of freedom  
## Multiple R-squared: 0.576, Adjusted R-squared: 0.574   
## F-statistic: 281.7 on 3 and 622 DF, p-value: < 2.2e-16

####DEALING WITH OUTLIERS####  
#mahalanobis  
mahal <- mahalanobis(hmr,  
 colMeans(hmr),  
 cov(hmr))  
  
cutoff <- qchisq(1-.001, ncol(hmr))  
cutoff

## [1] 18.46683

ncol(hmr)#df

## [1] 4

#how many outliers are there?  
badmahal <- as.numeric(mahal>cutoff)  
table(badmahal)

## badmahal  
## 0 1   
## 623 3

#leverage  
k = 3  
leverage <- hatvalues(output)  
cutleverage <- (2\*k+2)/nrow(hmr)  
cutleverage

## [1] 0.01277955

badleverage <- as.numeric(leverage>cutleverage)  
table(badleverage)

## badleverage  
## 0 1   
## 572 54

#54 people exceed the leverage cutoff  
  
#cooks  
cooks <- cooks.distance(output)  
cutcooks <- 4/(nrow(hmr)-k-1)  
cutcooks

## [1] 0.006430868

badcooks <- as.numeric(cooks>cutcooks)  
table(badcooks)

## badcooks  
## 0 1   
## 594 32

#32 people exceed the cook's cutoff score  
  
##Total outliers -- general rule of thumb: people with 2 or more outlier indicators should be excluded  
totalout <- badmahal + badleverage + badcooks  
table(totalout)

## totalout  
## 0 1 2   
## 552 59 15

#15 people have 2 indicators. We should exclude them.  
hmr\_xo <- subset(hmr, totalout<2)  
  
#rerun analysis without outliers  
output <- lm(confid~life + educ + age, data=hmr\_xo)  
summary(output)

##   
## Call:  
## lm(formula = confid ~ life + educ + age, data = hmr\_xo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31.3781 -4.5215 0.7498 5.3604 27.9956   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.92226 2.40529 4.957 9.32e-07 \*\*\*  
## life 0.57136 0.02069 27.622 < 2e-16 \*\*\*  
## educ -0.19017 0.10268 -1.852 0.06448 .   
## age 0.08115 0.02669 3.041 0.00246 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.22 on 607 degrees of freedom  
## Multiple R-squared: 0.5689, Adjusted R-squared: 0.5668   
## F-statistic: 267.1 on 3 and 607 DF, p-value: < 2.2e-16

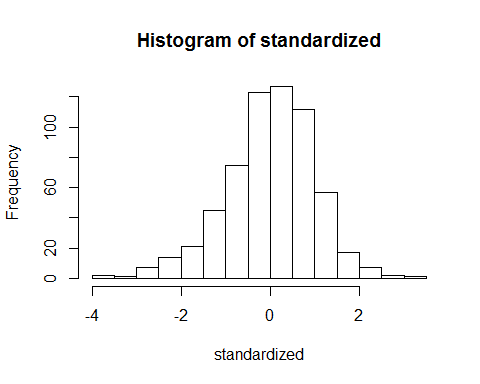
#additivity. Check to see if IVs are correlated.   
correl <- cor(hmr\_xo, use = "pairwise.complete.obs")  
correl

## age educ life confid  
## age 1.00000000 0.1012580 0.08092957 0.1373523  
## educ 0.10125798 1.0000000 0.19500812 0.1043848  
## life 0.08092957 0.1950081 1.00000000 0.7487151  
## confid 0.13735230 0.1043848 0.74871507 1.0000000

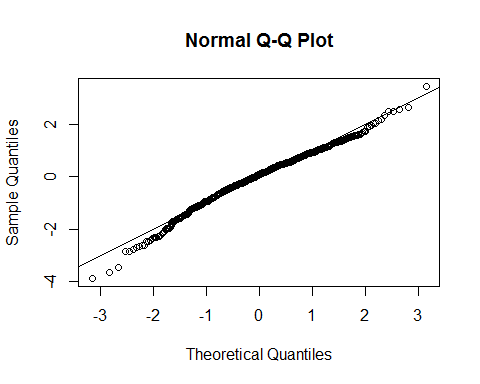
summary(output, correlation = T)

##   
## Call:  
## lm(formula = confid ~ life + educ + age, data = hmr\_xo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31.3781 -4.5215 0.7498 5.3604 27.9956   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.92226 2.40529 4.957 9.32e-07 \*\*\*  
## life 0.57136 0.02069 27.622 < 2e-16 \*\*\*  
## educ -0.19017 0.10268 -1.852 0.06448 .   
## age 0.08115 0.02669 3.041 0.00246 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.22 on 607 degrees of freedom  
## Multiple R-squared: 0.5689, Adjusted R-squared: 0.5668   
## F-statistic: 267.1 on 3 and 607 DF, p-value: < 2.2e-16  
##   
## Correlation of Coefficients:  
## (Intercept) life educ   
## life -0.62   
## educ -0.52 -0.19   
## age -0.31 -0.06 -0.09

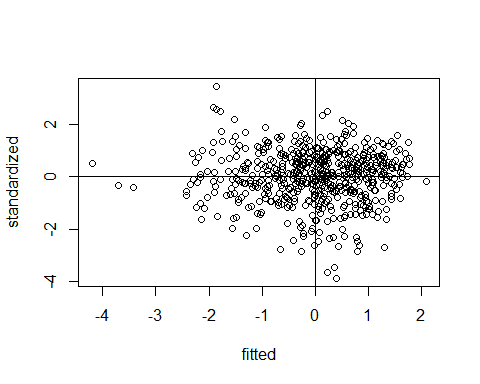
#assumption set up  
standardized <- rstudent(output)  
fitted <- scale(output$fitted.values)  
  
#normality  
hist(standardized)



#linearity  
qqnorm(standardized)  
abline(0,1)



#homogeneity and homoscedasticity  
plot(fitted,standardized)  
abline(0,0)  
abline(v=0)



##First model of just demographics  
model1 <- lm(confid~age + educ, data = hmr\_xo)  
summary(model1)

##   
## Call:  
## lm(formula = confid ~ age + educ, data = hmr\_xo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -35.190 -7.786 1.012 8.223 28.612   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 53.06106 2.83507 18.716 < 2e-16 \*\*\*  
## age 0.12737 0.03998 3.186 0.00152 \*\*   
## educ 0.34413 0.15136 2.274 0.02334 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.34 on 608 degrees of freedom  
## Multiple R-squared: 0.02714, Adjusted R-squared: 0.02394   
## F-statistic: 8.48 on 2 and 608 DF, p-value: 0.0002332

#to get Beta values  
QuantPsyc::lm.beta(model1)

## age educ   
## 0.12809590 0.09141406

#to get pr values  
partials <- ppcor::pcor(hmr\_xo[,c(1,2,4)], method = "pearson")  
partials$estimate^2

## age educ confid  
## age 1.000000000 0.007785274 0.016419226  
## educ 0.007785274 1.000000000 0.008429886  
## confid 0.016419226 0.008429886 1.000000000

#adding in the next IV to the model   
model2 <- lm(confid~age + educ + life, data = hmr\_xo)  
summary(model2)

##   
## Call:  
## lm(formula = confid ~ age + educ + life, data = hmr\_xo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31.3781 -4.5215 0.7498 5.3604 27.9956   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11.92226 2.40529 4.957 9.32e-07 \*\*\*  
## age 0.08115 0.02669 3.041 0.00246 \*\*   
## educ -0.19017 0.10268 -1.852 0.06448 .   
## life 0.57136 0.02069 27.622 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.22 on 607 degrees of freedom  
## Multiple R-squared: 0.5689, Adjusted R-squared: 0.5668   
## F-statistic: 267.1 on 3 and 607 DF, p-value: < 2.2e-16

#compare models to see if there's a difference  
anova(model1, model2)

## Analysis of Variance Table  
##   
## Model 1: confid ~ age + educ  
## Model 2: confid ~ age + educ + life  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 608 92569   
## 2 607 41016 1 51553 762.95 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

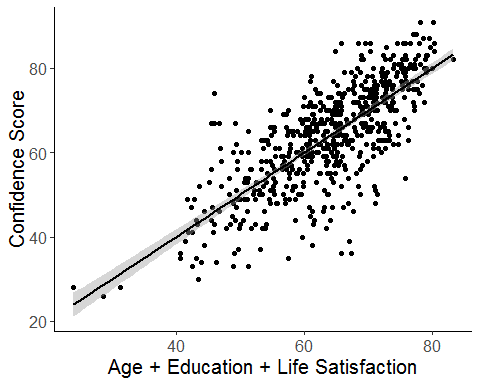
#get beta for new model  
QuantPsyc::lm.beta(model2)

## age educ life   
## 0.08161169 -0.05051766 0.75196163

#get pr values for new model  
partials <- ppcor::pcor(hmr\_xo, method = "pearson")  
partials$estimate^2

## age educ life confid  
## age 1.000000000 0.009161578 0.002498823 0.015004304  
## educ 0.009161578 1.000000000 0.032759979 0.005619942  
## life 0.002498823 0.032759979 1.000000000 0.556917577  
## confid 0.015004304 0.005619942 0.556917577 1.000000000

#plot to show how accurate our model is  
cleanup <- theme(panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank(),  
 panel.background = element\_blank(),  
 axis.line = element\_line(colour = "black"),  
 legend.key = element\_rect(fill = "white"),  
 text = element\_text(size = 15))  
  
fitted <- model2$fitted.values  
  
scatter <- ggplot(hmr\_xo, aes(fitted, confid))  
scatter +   
 cleanup +  
 geom\_point() +  
 geom\_smooth(method = "lm", color = "black") +  
 xlab("Age + Education + Life Satisfaction") +  
 ylab("Confidence Score")

 #####Results (APA style) Age, education and life satisfaction score were used to predict a participant’s confidence score. The data were screened for assumptions, and 15 participants were removed as outliers due to high Mahalanobis, Cook’s and/or leverage scores. Linearity, normality, multicollinearity, homogeneity and homoscedasticity were all met. Age and education were entered first into a hierarchical regression to control for demographic differences in confidence. Overall, this model was significant, indicating that demographics predict a participant’s confidence score, F(2,608) = 8.48, p<0.01, R2 = 0.03. Age was a stronger predictor of confidence, B = 0.12, t(608) = 3.186, p<0.01, pr2 = 0.02, which showed that participants are likely to be more confident as they get older. Education was also positively related to confidence, B = 0.09, t(608) = 2.27, p = 0.02, pr2 = 0.008; therefore, more educated participants tend to be more confident. Next, life satisfaction score was added in a second step to examine its predictive value after controlling for demographic variables. The addition of this variable was significant DELTA\_F(1,607) = 762.95, p < 0.001, DELTA\_R2 = 0.539 (model 2 - model 1). Life satisfaction scores was the highest predictor of confidence score where participants with a higher life satisfaction score also had higher confidence scores B = 0.75, t(607) = 27.62, p<0.001, pr2 = 0.557