Homework 4

Ben Chu

May 2, 2018

loading in packages/data and then doing a little cleaning

library(tidyverse)  
library(car)  
library(psych)  
library(forecast)  
library(purrr)  
library(lmtest)  
load("C:/Users/Branly Mclanbry/Downloads/heat.RData")  
  
hw1 <- heat %>%  
 janitor::clean\_names()%>%  
 mutate(date.fac = factor(date))  
  
var <- list(hw1$nviolent,hw1$rape,hw1$rob)  
names <- c("crime","rape","robbery")

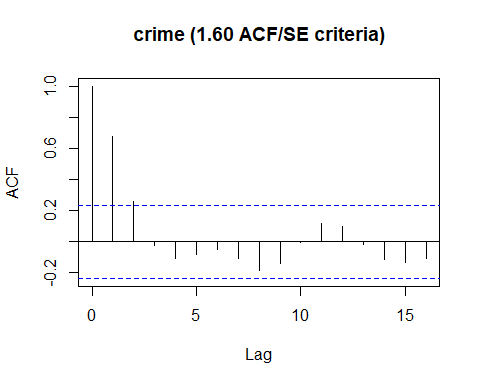
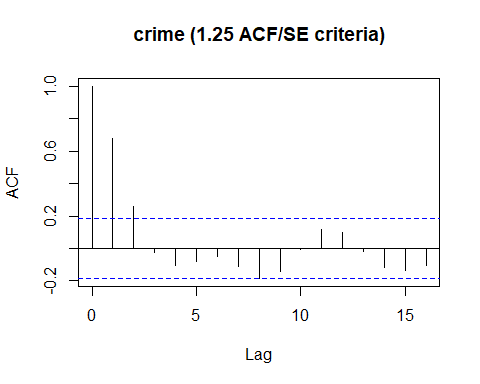
writing a function and then just mapping it across the three variables

dur\_pan\_box <- function(x,name){  
 mod <- lm(x ~ year + temp + age, hw1)  
   
 print(name)  
 print(summary(mod))  
 dwtest(mod)  
   
 x79 <- acf(residuals(mod), ci = .79, main = paste(name, "(1.25 ACF/SE criteria)"))  
 x89 <- acf(residuals(mod), ci = .89, main = paste(name, "(1.60 ACF/SE criteria)"))  
 print(x79)  
 print(x89)  
 test<- lapply(1:6,function(x){  
 Box.test(residuals(mod),lag = x, type = "Ljung-Box")  
 })  
   
 print(test)  
 }

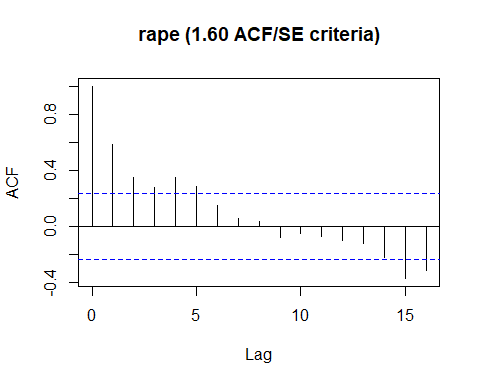
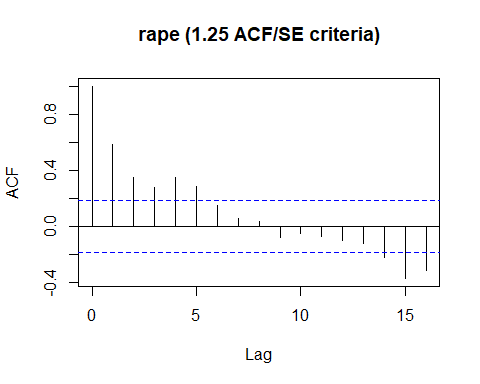
# 1a

walk2(var,names,dur\_pan\_box)

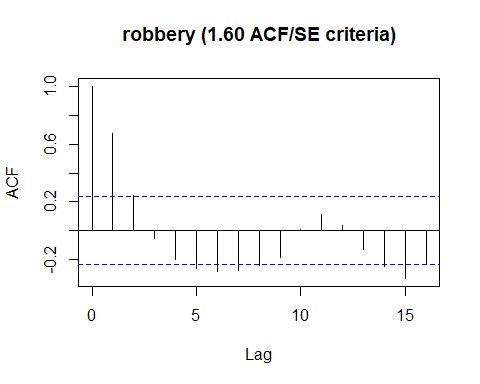
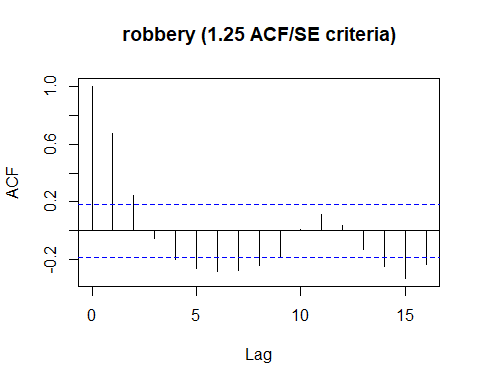
## [1] "crime"  
##   
## Call:  
## lm(formula = x ~ year + temp + age, data = hw1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -270.43 -68.90 14.47 75.45 220.23   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2615.767 1650.801 -1.585 0.121   
## year 29.231 1.735 16.845 <2e-16 \*\*\*  
## temp -12.575 28.674 -0.439 0.663   
## age 109.701 7.917 13.856 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 122.4 on 42 degrees of freedom  
## Multiple R-squared: 0.9625, Adjusted R-squared: 0.9599   
## F-statistic: 359.7 on 3 and 42 DF, p-value: < 2.2e-16



##   
## Autocorrelations of series 'residuals(mod)', by lag  
##   
## 0 1 2 3 4 5 6 7 8 9   
## 1.000 0.678 0.257 -0.027 -0.105 -0.081 -0.053 -0.108 -0.183 -0.143   
## 10 11 12 13 14 15 16   
## -0.007 0.119 0.099 -0.019 -0.116 -0.135 -0.105   
##   
## Autocorrelations of series 'residuals(mod)', by lag  
##   
## 0 1 2 3 4 5 6 7 8 9   
## 1.000 0.678 0.257 -0.027 -0.105 -0.081 -0.053 -0.108 -0.183 -0.143   
## 10 11 12 13 14 15 16   
## -0.007 0.119 0.099 -0.019 -0.116 -0.135 -0.105   
## [[1]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 22.575, df = 1, p-value = 2.021e-06  
##   
##   
## [[2]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 25.892, df = 2, p-value = 2.386e-06  
##   
##   
## [[3]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 25.929, df = 3, p-value = 9.87e-06  
##   
##   
## [[4]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 26.51, df = 4, p-value = 2.497e-05  
##   
##   
## [[5]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 26.864, df = 5, p-value = 6.064e-05  
##   
##   
## [[6]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 27.016, df = 6, p-value = 0.0001438  
##   
##   
## [1] "rape"  
##   
## Call:  
## lm(formula = x ~ year + temp + age, data = hw1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.9092 -2.4625 -0.4771 1.7185 6.4265   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -208.68254 39.89504 -5.231 5.02e-06 \*\*\*  
## year 0.73464 0.04194 17.518 < 2e-16 \*\*\*  
## temp 2.73791 0.69296 3.951 0.000292 \*\*\*  
## age 0.90992 0.19133 4.756 2.34e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.959 on 42 degrees of freedom  
## Multiple R-squared: 0.9468, Adjusted R-squared: 0.943   
## F-statistic: 249.2 on 3 and 42 DF, p-value: < 2.2e-16



##   
## Autocorrelations of series 'residuals(mod)', by lag  
##   
## 0 1 2 3 4 5 6 7 8 9   
## 1.000 0.583 0.348 0.274 0.348 0.288 0.147 0.056 0.031 -0.083   
## 10 11 12 13 14 15 16   
## -0.052 -0.077 -0.102 -0.125 -0.223 -0.373 -0.316   
##   
## Autocorrelations of series 'residuals(mod)', by lag  
##   
## 0 1 2 3 4 5 6 7 8 9   
## 1.000 0.583 0.348 0.274 0.348 0.288 0.147 0.056 0.031 -0.083   
## 10 11 12 13 14 15 16   
## -0.052 -0.077 -0.102 -0.125 -0.223 -0.373 -0.316   
## [[1]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 16.666, df = 1, p-value = 4.458e-05  
##   
##   
## [[2]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 22.752, df = 2, p-value = 1.147e-05  
##   
##   
## [[3]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 26.615, df = 3, p-value = 7.089e-06  
##   
##   
## [[4]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 32.981, df = 4, p-value = 1.205e-06  
##   
##   
## [[5]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 37.436, df = 5, p-value = 4.896e-07  
##   
##   
## [[6]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 38.627, df = 6, p-value = 8.47e-07  
##   
##   
## [1] "robbery"  
##   
## Call:  
## lm(formula = x ~ year + temp + age, data = hw1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -30.742 -14.025 -0.746 12.239 36.961   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -988.3235 244.8099 -4.037 0.000225 \*\*\*  
## year 4.8053 0.2573 18.673 < 2e-16 \*\*\*  
## temp 9.6406 4.2522 2.267 0.028588 \*   
## age 10.0275 1.1741 8.541 9.98e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18.16 on 42 degrees of freedom  
## Multiple R-squared: 0.9582, Adjusted R-squared: 0.9552   
## F-statistic: 320.6 on 3 and 42 DF, p-value: < 2.2e-16



##   
## Autocorrelations of series 'residuals(mod)', by lag  
##   
## 0 1 2 3 4 5 6 7 8 9   
## 1.000 0.674 0.243 -0.055 -0.200 -0.264 -0.283 -0.273 -0.239 -0.182   
## 10 11 12 13 14 15 16   
## 0.012 0.111 0.033 -0.132 -0.245 -0.330 -0.231   
##   
## Autocorrelations of series 'residuals(mod)', by lag  
##   
## 0 1 2 3 4 5 6 7 8 9   
## 1.000 0.674 0.243 -0.055 -0.200 -0.264 -0.283 -0.273 -0.239 -0.182   
## 10 11 12 13 14 15 16   
## 0.012 0.111 0.033 -0.132 -0.245 -0.330 -0.231   
## [[1]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 22.289, df = 1, p-value = 2.346e-06  
##   
##   
## [[2]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 25.251, df = 2, p-value = 3.288e-06  
##   
##   
## [[3]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 25.407, df = 3, p-value = 1.269e-05  
##   
##   
## [[4]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 27.51, df = 4, p-value = 1.567e-05  
##   
##   
## [[5]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 31.265, df = 5, p-value = 8.305e-06  
##   
##   
## [[6]]  
##   
## Box-Ljung test  
##   
## data: residuals(mod)  
## X-squared = 35.69, df = 6, p-value = 3.167e-06

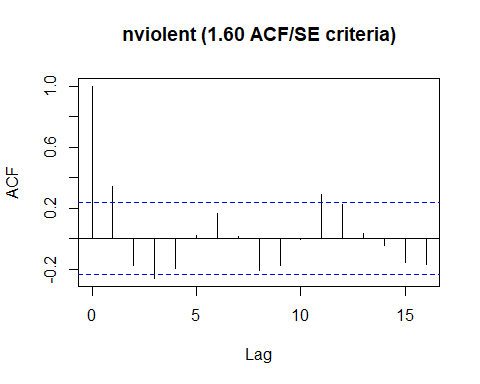
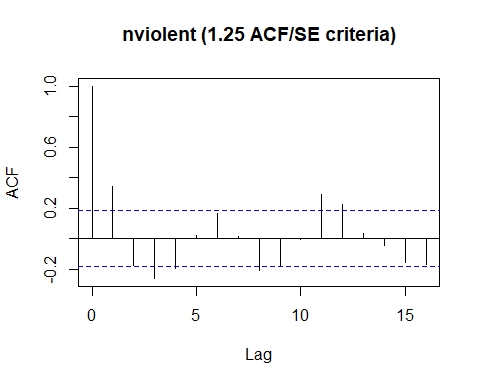
Writing an ARIMA function

pred <- select(hw1,temp,age,year)  
arimar <- function(x){  
 for(i in 1:3){  
 armod <- Arima(hw1[x], xreg = pred, order = c(i,0,0))  
   
 acf(residuals(armod), ci = .79, main = paste(x, "(1.25 ACF/SE criteria)"))  
 acf(residuals(armod), ci = .89, main = paste(x, "(1.60 ACF/SE criteria)"))  
   
 sapply(1:6, function(x){  
 print(Box.test(residuals(armod),lag = x, type = "Ljung-Box"))  
 })  
 print("t values")  
 print(as.list((armod$coef)/sqrt(diag(armod$var.coef))))  
 print("p values")  
 print(as.list((1-pnorm(abs(armod$coef)/sqrt(diag(armod$var.coef))))\*2))  
 print(armod)  
 }  
}

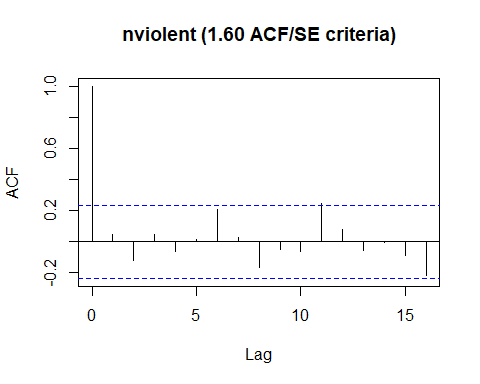
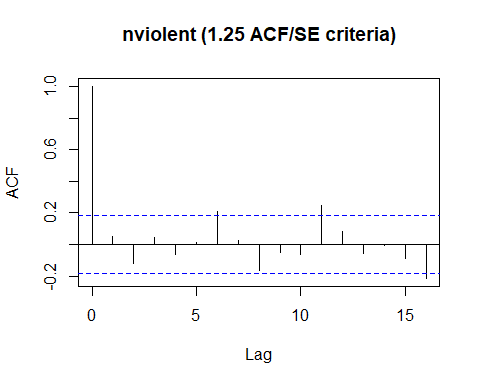
# 1b

It looks like 2 autocorrelative terms were used for robbery and property crimer, while 1 autocorrelative term was used for rape.

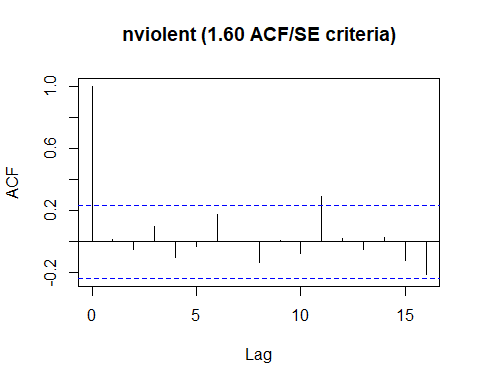
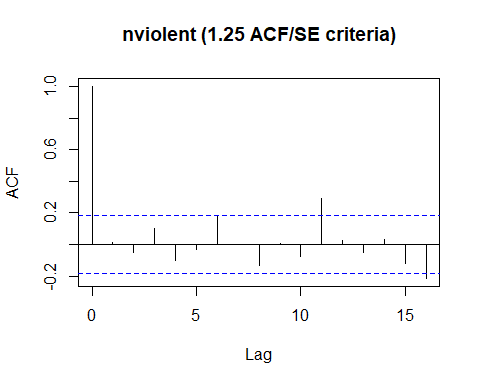
map(c("nviolent","rape","rob"),arimar)



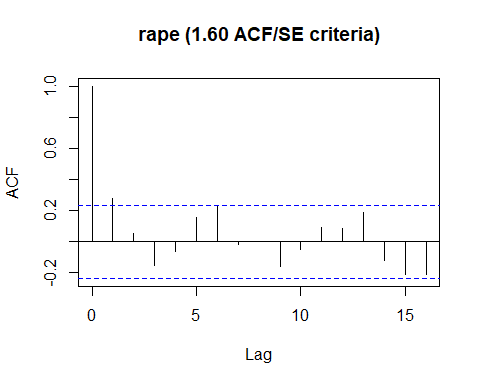
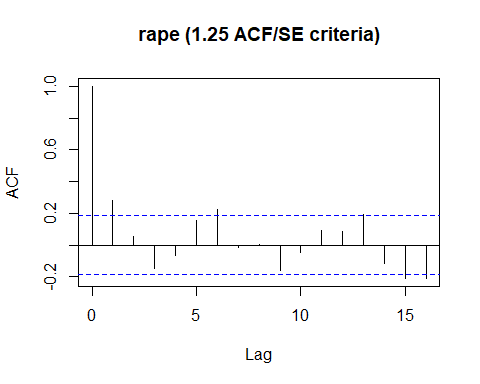
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 5.7457, df = 1, p-value = 0.01653  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 7.2621, df = 2, p-value = 0.02649  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 10.739, df = 3, p-value = 0.01322  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 12.75, df = 4, p-value = 0.01256  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 12.779, df = 5, p-value = 0.02554  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 14.311, df = 6, p-value = 0.02635  
##   
## [1] "t values"  
## $ar1  
## [1] 7.171522  
##   
## $intercept  
## [1] -3.239159  
##   
## $temp  
## [1] 0.3579092  
##   
## $age  
## [1] 4.012761  
##   
## $year  
## [1] 7.541735  
##   
## [1] "p values"  
## $ar1  
## [1] 7.41629e-13  
##   
## $intercept  
## [1] 0.001198827  
##   
## $temp  
## [1] 0.7204112  
##   
## $age  
## [1] 6.001276e-05  
##   
## $year  
## [1] 4.640732e-14  
##   
## Series: hw1[x]   
## Regression with ARIMA(1,0,0) errors   
##   
## Coefficients:  
## ar1 intercept temp age year  
## 0.8371 -3378.521 6.0592 92.0464 29.9810  
## s.e. 0.1167 1043.024 16.9295 22.9384 3.9753  
##   
## sigma^2 estimated as 6649: log likelihood=-265.68  
## AIC=543.36 AICc=545.51 BIC=554.33



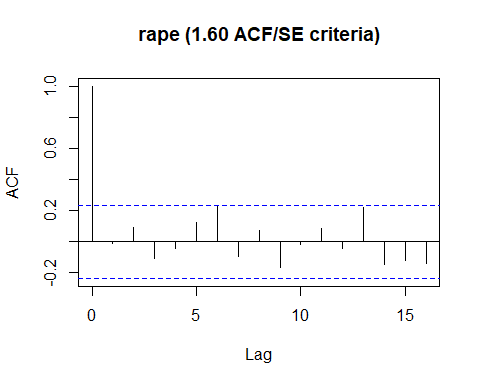
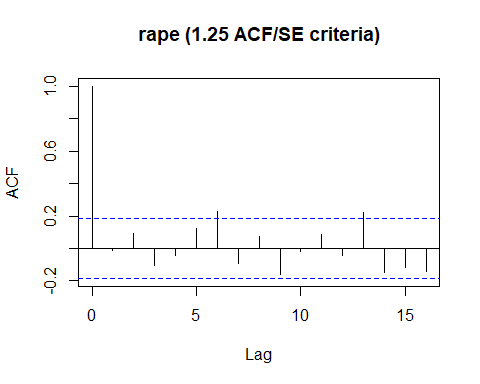
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.11993, df = 1, p-value = 0.7291  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.86502, df = 2, p-value = 0.6489  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.97954, df = 3, p-value = 0.8062  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.1671, df = 4, p-value = 0.8835  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.1744, df = 5, p-value = 0.9473  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 3.5797, df = 6, p-value = 0.7333  
##   
## [1] "t values"  
## $ar1  
## [1] 8.434694  
##   
## $ar2  
## [1] -3.723028  
##   
## $intercept  
## [1] -4.074437  
##   
## $temp  
## [1] -0.1874068  
##   
## $age  
## [1] 7.185045  
##   
## $year  
## [1] 11.76834  
##   
## [1] "p values"  
## $ar1  
## [1] 0  
##   
## $ar2  
## [1] 0.000196848  
##   
## $intercept  
## [1] 4.61257e-05  
##   
## $temp  
## [1] 0.8513417  
##   
## $age  
## [1] 6.71907e-13  
##   
## $year  
## [1] 0  
##   
## Series: hw1[x]   
## Regression with ARIMA(2,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 intercept temp age year  
## 1.1763 -0.5086 -3044.6082 -2.3656 101.0238 29.6711  
## s.e. 0.1395 0.1366 747.2463 12.6226 14.0603 2.5213  
##   
## sigma^2 estimated as 5263: log likelihood=-259.9  
## AIC=533.79 AICc=536.74 BIC=546.59



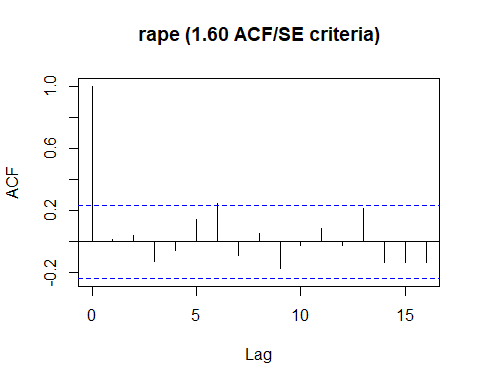
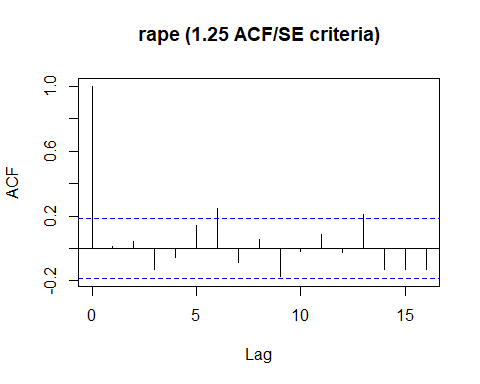
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.012435, df = 1, p-value = 0.9112  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.134, df = 2, p-value = 0.9352  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.65119, df = 3, p-value = 0.8846  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.1738, df = 4, p-value = 0.8824  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.2334, df = 5, p-value = 0.9416  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 2.9688, df = 6, p-value = 0.8128  
##   
## [1] "t values"  
## $ar1  
## [1] 7.600998  
##   
## $ar2  
## [1] -3.538438  
##   
## $ar3  
## [1] 2.154948  
##   
## $intercept  
## [1] -2.57443  
##   
## $temp  
## [1] 0.1187351  
##   
## $age  
## [1] 2.043786  
##   
## $year  
## [1] 5.437173  
##   
## [1] "p values"  
## $ar1  
## [1] 2.930989e-14  
##   
## $ar2  
## [1] 0.0004025016  
##   
## $ar3  
## [1] 0.03116591  
##   
## $intercept  
## [1] 0.01004054  
##   
## $temp  
## [1] 0.9054852  
##   
## $age  
## [1] 0.04097473  
##   
## $year  
## [1] 5.413252e-08  
##   
## Series: hw1[x]   
## Regression with ARIMA(3,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 ar3 intercept temp age year  
## 1.4792 -0.8890 0.321 -2480.8220 1.1904 62.7580 30.5219  
## s.e. 0.1946 0.2512 0.149 963.6393 10.0259 30.7068 5.6136  
##   
## sigma^2 estimated as 5007: log likelihood=-258.69  
## AIC=533.37 AICc=537.27 BIC=548



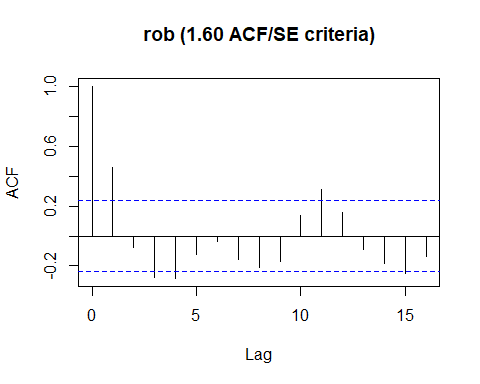
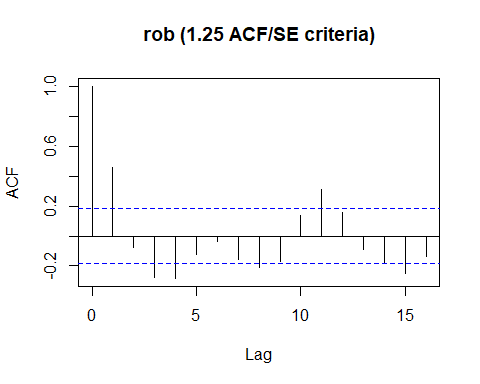
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 3.8039, df = 1, p-value = 0.05113  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 3.9338, df = 2, p-value = 0.1399  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 5.0864, df = 3, p-value = 0.1656  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 5.3079, df = 4, p-value = 0.2571  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 6.5857, df = 5, p-value = 0.2533  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 9.3773, df = 6, p-value = 0.1534  
##   
## [1] "t values"  
## $ar1  
## [1] 16.35108  
##   
## $intercept  
## [1] -2.47034  
##   
## $temp  
## [1] 0.812345  
##   
## $age  
## [1] 1.410406  
##   
## $year  
## [1] 5.631148  
##   
## [1] "p values"  
## $ar1  
## [1] 0  
##   
## $intercept  
## [1] 0.01349847  
##   
## $temp  
## [1] 0.4165937  
##   
## $age  
## [1] 0.1584199  
##   
## $year  
## [1] 1.790137e-08  
##   
## Series: hw1[x]   
## Regression with ARIMA(1,0,0) errors   
##   
## Coefficients:  
## ar1 intercept temp age year  
## 0.9178 -51.2869 0.2430 0.5631 0.6645  
## s.e. 0.0561 20.7611 0.2991 0.3993 0.1180  
##   
## sigma^2 estimated as 2.252: log likelihood=-82.22  
## AIC=176.43 AICc=178.59 BIC=187.41



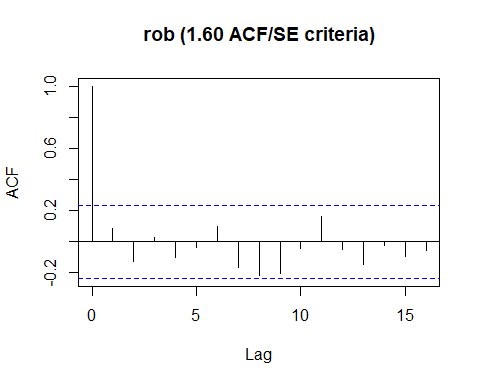
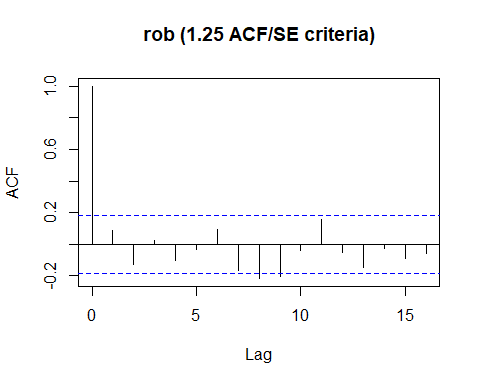
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.0075147, df = 1, p-value = 0.9309  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.43729, df = 2, p-value = 0.8036  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.0294, df = 3, p-value = 0.7941  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.1272, df = 4, p-value = 0.8899  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.9591, df = 5, p-value = 0.8548  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 4.8657, df = 6, p-value = 0.5611  
##   
## [1] "t values"  
## $ar1  
## [1] 7.704022  
##   
## $ar2  
## [1] -2.271188  
##   
## $intercept  
## [1] -2.1366  
##   
## $temp  
## [1] 0.6916311  
##   
## $age  
## [1] 0.5409008  
##   
## $year  
## [1] 5.348854  
##   
## [1] "p values"  
## $ar1  
## [1] 1.310063e-14  
##   
## $ar2  
## [1] 0.02313561  
##   
## $intercept  
## [1] 0.03263051  
##   
## $temp  
## [1] 0.489169  
##   
## $age  
## [1] 0.588576  
##   
## $year  
## [1] 8.851309e-08  
##   
## Series: hw1[x]   
## Regression with ARIMA(2,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 intercept temp age year  
## 1.2548 -0.3486 -43.1601 0.1686 0.2817 0.6968  
## s.e. 0.1629 0.1535 20.2004 0.2437 0.5209 0.1303  
##   
## sigma^2 estimated as 2.055: log likelihood=-79.75  
## AIC=173.51 AICc=176.46 BIC=186.31



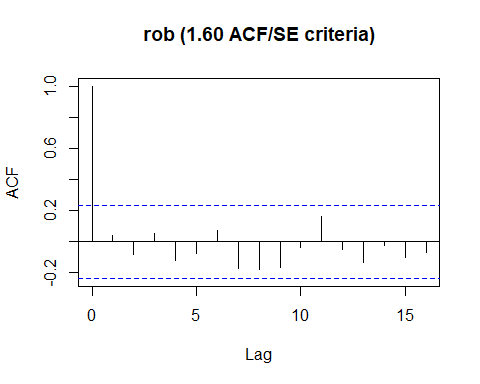
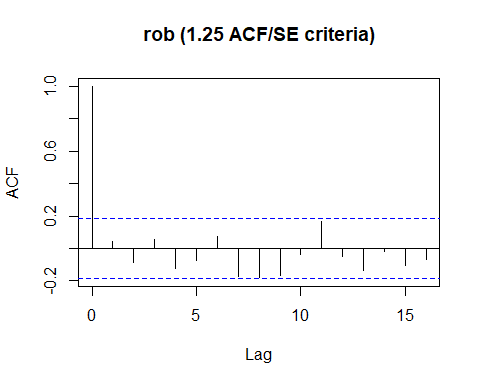
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.0099142, df = 1, p-value = 0.9207  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.099213, df = 2, p-value = 0.9516  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.95973, df = 3, p-value = 0.811  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.1129, df = 4, p-value = 0.8922  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 2.2169, df = 5, p-value = 0.8184  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 5.5929, df = 6, p-value = 0.4703  
##   
## [1] "t values"  
## $ar1  
## [1] 7.560722  
##   
## $ar2  
## [1] -1.102142  
##   
## $ar3  
## [1] -0.4318153  
##   
## $intercept  
## [1] -1.995539  
##   
## $temp  
## [1] 0.753798  
##   
## $age  
## [1] 0.2622117  
##   
## $year  
## [1] 5.077909  
##   
## [1] "p values"  
## $ar1  
## [1] 4.019007e-14  
##   
## $ar2  
## [1] 0.2704001  
##   
## $ar3  
## [1] 0.6658756  
##   
## $intercept  
## [1] 0.04598413  
##   
## $temp  
## [1] 0.4509706  
##   
## $age  
## [1] 0.7931582  
##   
## $year  
## [1] 3.816111e-07  
##   
## Series: hw1[x]   
## Regression with ARIMA(3,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 ar3 intercept temp age year  
## 1.2487 -0.2670 -0.0759 -42.0824 0.1894 0.1693 0.7000  
## s.e. 0.1652 0.2422 0.1757 21.0882 0.2513 0.6458 0.1378  
##   
## sigma^2 estimated as 2.093: log likelihood=-79.66  
## AIC=175.32 AICc=179.21 BIC=189.95



##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 10.416, df = 1, p-value = 0.001249  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 10.723, df = 2, p-value = 0.004694  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 14.67, df = 3, p-value = 0.002121  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 18.881, df = 4, p-value = 0.0008296  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 19.716, df = 5, p-value = 0.001413  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 19.787, df = 6, p-value = 0.003022  
##   
## [1] "t values"  
## $ar1  
## [1] 7.706336  
##   
## $intercept  
## [1] -3.686318  
##   
## $temp  
## [1] 1.088114  
##   
## $age  
## [1] 4.36057  
##   
## $year  
## [1] 10.46833  
##   
## [1] "p values"  
## $ar1  
## [1] 1.287859e-14  
##   
## $intercept  
## [1] 0.0002275224  
##   
## $temp  
## [1] 0.2765449  
##   
## $age  
## [1] 1.29724e-05  
##   
## $year  
## [1] 0  
##   
## Series: hw1[x]   
## Regression with ARIMA(1,0,0) errors   
##   
## Coefficients:  
## ar1 intercept temp age year  
## 0.7462 -592.6452 2.9317 9.6231 4.8001  
## s.e. 0.0968 160.7689 2.6943 2.2069 0.4585  
##   
## sigma^2 estimated as 156.9: log likelihood=-179.31  
## AIC=370.62 AICc=372.78 BIC=381.6



##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.36588, df = 1, p-value = 0.5453  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.1972, df = 2, p-value = 0.5496  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.2319, df = 3, p-value = 0.7454  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.755, df = 4, p-value = 0.7807  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.8186, df = 5, p-value = 0.8736  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 2.3224, df = 6, p-value = 0.8878  
##   
## [1] "t values"  
## $ar1  
## [1] 10.79236  
##   
## $ar2  
## [1] -5.579143  
##   
## $intercept  
## [1] -5.208961  
##   
## $temp  
## [1] 0.7315349  
##   
## $age  
## [1] 6.583023  
##   
## $year  
## [1] 15.97634  
##   
## [1] "p values"  
## $ar1  
## [1] 0  
##   
## $ar2  
## [1] 2.417068e-08  
##   
## $intercept  
## [1] 1.899007e-07  
##   
## $temp  
## [1] 0.4644525  
##   
## $age  
## [1] 4.609779e-11  
##   
## $year  
## [1] 0  
##   
## Series: hw1[x]   
## Regression with ARIMA(2,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 intercept temp age year  
## 1.2434 -0.6284 -502.2360 1.1819 9.7624 4.9028  
## s.e. 0.1152 0.1126 96.4177 1.6157 1.4830 0.3069  
##   
## sigma^2 estimated as 96.86: log likelihood=-168.18  
## AIC=350.36 AICc=353.31 BIC=363.16



##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.089926, df = 1, p-value = 0.7643  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.4504, df = 2, p-value = 0.7984  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 0.60769, df = 3, p-value = 0.8947  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.3913, df = 4, p-value = 0.8457  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.698, df = 5, p-value = 0.8892  
##   
##   
## Box-Ljung test  
##   
## data: residuals(armod)  
## X-squared = 1.9844, df = 6, p-value = 0.9211  
##   
## [1] "t values"  
## $ar1  
## [1] 8.789139  
##   
## $ar2  
## [1] -3.440682  
##   
## $ar3  
## [1] 0.6591355  
##   
## $intercept  
## [1] -5.250555  
##   
## $temp  
## [1] 0.6950779  
##   
## $age  
## [1] 6.094596  
##   
## $year  
## [1] 15.09128  
##   
## [1] "p values"  
## $ar1  
## [1] 0  
##   
## $ar2  
## [1] 0.0005802501  
##   
## $ar3  
## [1] 0.5098088  
##   
## $intercept  
## [1] 1.516416e-07  
##   
## $temp  
## [1] 0.4870065  
##   
## $age  
## [1] 1.097142e-09  
##   
## $year  
## [1] 0  
##   
## Series: hw1[x]   
## Regression with ARIMA(3,0,0) errors   
##   
## Coefficients:  
## ar1 ar2 ar3 intercept temp age year  
## 1.3055 -0.7523 0.0989 -493.8748 1.0749 9.6350 4.9142  
## s.e. 0.1485 0.2186 0.1501 94.0614 1.5464 1.5809 0.3256  
##   
## sigma^2 estimated as 98.35: log likelihood=-167.97  
## AIC=351.93 AICc=355.82 BIC=366.56

## [[1]]  
## NULL  
##   
## [[2]]  
## NULL  
##   
## [[3]]  
## NULL

# 1c

Anderson and his colleagues utilized Box-Ljung/chi-square tests for autocorrelations/fit. It looks like they used six lags for residuals.

|  |  |  |
| --- | --- | --- |
| Box Ljung | Chi Square | DF |
| Violence(1) | 5.75\* | 1 |
|  | 7.26\* | 2 |
|  | 10.74\* | 3 |
|  | 12.75\* | 4 |
|  | 12.78\* | 5 |
|  | 14.31\* | 6 |
| Violence(2) | 0.12 | 1 |
|  | 0.86 | 2 |
|  | 0.98 | 3 |
|  | 1.17 | 4 |
|  | 1.17 | 5 |
|  | 3.58 | 6 |
| Violence(3) | 0.01 | 1 |
|  | 0.13 | 2 |
|  | 0.65 | 3 |
|  | 1.17 | 4 |
|  | 1.23 | 5 |
|  | 2.98 | 6 |
| Rape(1) | 3.80 | 1 |
|  | 3.94 | 2 |
|  | 5.08 | 3 |
|  | 5.31 | 4 |
|  | 6.58 | 5 |
|  | 9.34 | 6 |
| Rape(2) | 0.01 | 1 |
|  | 0.44 | 2 |
|  | 1.03 | 3 |
|  | 1.13 | 4 |
|  | 1.956 | 5 |
|  | 4.97 | 6 |
| Rape(3) | 0.01 | 1 |
|  | 0.01 | 2 |
|  | 0.96 | 3 |
|  | 1.11 | 4 |
|  | 2.22 | 5 |
|  | 5.60 | 6 |
| Rob(1) | 10.41\* | 1 |
|  | 10.72\* | 2 |
|  | 14.67\* | 3 |
|  | 18.81\* | 4 |
|  | 19.71\* | 5 |
|  | 19.75\* | 6 |
| Rob(2) | 0.35 | 1 |
|  | 1.19 | 2 |
|  | 1.23 | 3 |
|  | 1.76 | 4 |
|  | 1.82 | 5 |
|  | 2.32 | 6 |
| Rob(3) | 0.08 | 1 |
|  | 0.45 | 2 |
|  | 0.60 | 3 |
|  | 1.39 | 4 |
|  | 1.68 | 5 |
|  | 1.98 | 6 |

Note \* p < .05

# 1d

A model predicting violence from temperature, year, and age indicated issues with autocorrelation. Two autoregressive terms were used in the Arima Model.Year (*b* = 29.67, *p* < .001) and age (*b* = 101.02, *p* < .001) were significant while temperature (*b* = -2.36, *p* = .85) were not

A model predicting rape from temperature, year, and age indicated issues with autocorrelation. One autoregressive terms were used in the Arima Model.Year (*b* = 0.66, *p* < .001) was significant but age (*b* = 0.56, *p* = .16) and temperature (*b* = 0.24, *p* = .41) were not

A model predicting Robbery from temperature, year, and age indicated issues with autocorrelation. Two autoregressive terms were used in the Arima Model.Year (*b* = 4.90, *p* < .001) and age (*b* = 9.76, *p* < .001) but temperature (*b* = 1.18, *p* = .46) were not.

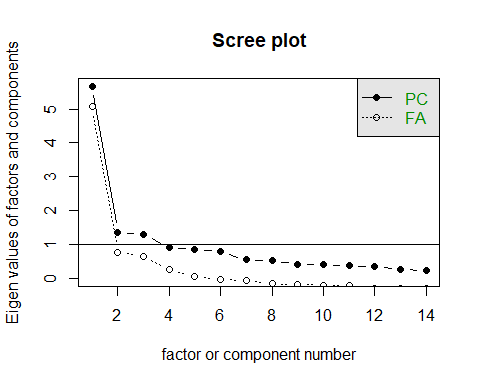
ACF and Pankrantz criteria were used, but ultimately ignored because there was a complete lack of autocorrelation assumptions. I just looked at the Box-Ljung until it hit over .05 for the lags and called it good. This led me to the ARIMA models andI found which autoregressive terms that were the *best*. Year seems to be significant for all models but age fluctuates and temperature is not significant at all.

# 2a

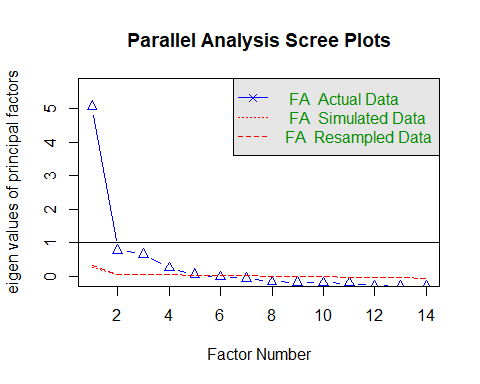
The added proportional values total about 55% which suggests a little more than half of the variance can be explained with the four factors. The first factor is a(1-3) while the second factor is a(21-23) and the third factor a(12,16-18 and 24) while the fourth factor is a(14-16).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | Communality |
| q5a2 | **0.86** | 0.20 | 0.18 | 0.15 | 1.3 |
| q5a1 | **0.73** | 0.23 | 0.22 | 0.17 | 1.5 |
| q5a3 | **0.70** | 0.17 | 0.20 | 0.15 | 1.4 |
| q5a23 | 0.20 | **0.82** | 0.14 | 0.18 | 1.3 |
| q5a21 | 0.23 | **0.75** | 0.17 | 0.22 | 1.5 |
| q5a22 | 0.22 | **0.67** | 0.16 | 0.31 | 1.8 |
| q5a18 | 0.15 | 0.11 | **0.80** | 0.15 | 1.2 |
| q5a6 | 0.25 | 0.15 | **0.59** | 0.15 | 1.6 |
| q5a17 | 0.34 | 0.18 | **0.46** | 0.34 | 3.2 |
| q5a24 | 0.07 | 0.07 | **0.35** | 0.16 | 1.6 |
| q5a12 | 0.21 | 0.15 | **0.21** | 0.17 | 3.7 |
| q5a16 | 0.19 | 0.27 | 0.23 | **0.72** | 1.7 |
| q5a14 | 0.19 | 0.32 | 0.16 | **0.56** | 2.1 |
| q5a15 | 0.12 | 0.14 | 0.31 | **0.55** | 1.8 |
| Eigenvalue | 2.23 | 2.10 | 1.74 | 1.60 |  |
| % explained | 16 | 15 | 12 | 11 |  |

load("C:/Users/Branly Mclanbry/Downloads/schumm.RData")  
scree(schumm)



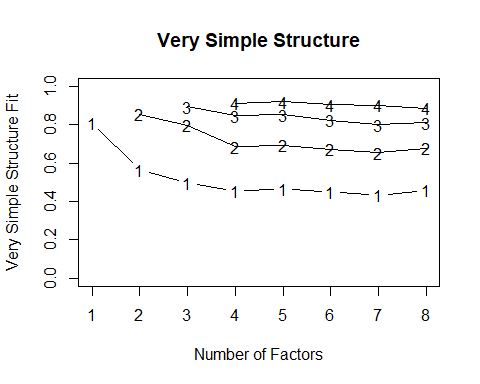
fa.parallel(schumm, fa = "fa")



## Parallel analysis suggests that the number of factors = 5 and the number of components = NA

vss(schumm, rotate = "varimax", fm = "fa")

## factor method not specified correctly, minimum residual (unweighted least squares used  
## factor method not specified correctly, minimum residual (unweighted least squares used  
## factor method not specified correctly, minimum residual (unweighted least squares used  
## factor method not specified correctly, minimum residual (unweighted least squares used  
## factor method not specified correctly, minimum residual (unweighted least squares used  
## factor method not specified correctly, minimum residual (unweighted least squares used  
## factor method not specified correctly, minimum residual (unweighted least squares used  
## factor method not specified correctly, minimum residual (unweighted least squares used



##   
## Very Simple Structure  
## Call: vss(x = schumm, rotate = "varimax", fm = "fa")  
## VSS complexity 1 achieves a maximimum of 0.81 with 1 factors  
## VSS complexity 2 achieves a maximimum of 0.85 with 2 factors  
##   
## The Velicer MAP achieves a minimum of 0.03 with 3 factors   
## BIC achieves a minimum of -77.08 with 7 factors  
## Sample Size adjusted BIC achieves a minimum of -32.59 with 7 factors  
##   
## Statistics by number of factors   
## vss1 vss2 map dof chisq prob sqresid fit RMSEA BIC SABIC  
## 1 0.81 0.00 0.034 77 13010 0.0e+00 7.4 0.81 0.152 12325.29 12570  
## 2 0.57 0.85 0.033 64 7445 0.0e+00 5.7 0.85 0.126 6875.82 7079  
## 3 0.50 0.80 0.031 52 2469 0.0e+00 4.2 0.89 0.080 2006.32 2172  
## 4 0.45 0.69 0.036 41 563 5.3e-93 3.5 0.91 0.042 198.64 329  
## 5 0.47 0.69 0.053 31 276 4.2e-41 2.7 0.93 0.033 0.19 99  
## 6 0.45 0.67 0.075 22 152 2.0e-21 2.7 0.93 0.028 -43.67 26  
## 7 0.43 0.65 0.095 14 47 1.6e-05 2.4 0.94 0.018 -77.08 -33  
## 8 0.46 0.68 0.132 7 22 2.9e-03 2.2 0.95 0.017 -40.59 -18  
## complex eChisq SRMR eCRMS eBIC  
## 1 1.0 11062.7 0.0912 0.0991 10378  
## 2 1.4 5956.0 0.0669 0.0798 5387  
## 3 1.6 1580.7 0.0345 0.0456 1118  
## 4 1.8 328.9 0.0157 0.0234 -36  
## 5 1.7 140.2 0.0103 0.0176 -136  
## 6 1.8 50.4 0.0062 0.0125 -145  
## 7 1.9 13.1 0.0031 0.0080 -111  
## 8 1.9 6.8 0.0023 0.0081 -56

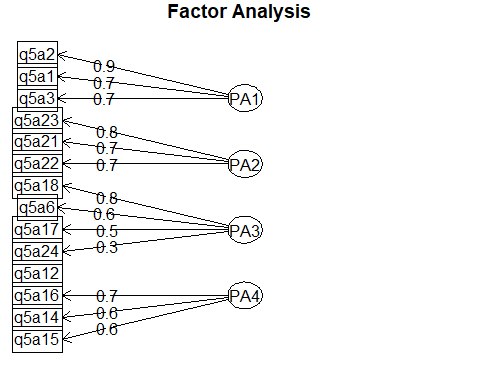
yy<-fa(schumm, nfactors = 4, rotate = "varimax", fm = "pa")  
fa.sort(yy)

## Factor Analysis using method = pa  
## Call: fa(r = schumm, nfactors = 4, rotate = "varimax", fm = "pa")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## PA1 PA2 PA3 PA4 h2 u2 com  
## q5a2 0.86 0.20 0.18 0.15 0.83 0.17 1.3  
## q5a1 0.73 0.23 0.22 0.17 0.66 0.34 1.5  
## q5a3 0.70 0.17 0.20 0.15 0.58 0.42 1.4  
## q5a23 0.20 0.82 0.14 0.18 0.77 0.23 1.3  
## q5a21 0.23 0.75 0.17 0.22 0.69 0.31 1.5  
## q5a22 0.22 0.67 0.16 0.31 0.61 0.39 1.8  
## q5a18 0.15 0.11 0.80 0.15 0.69 0.31 1.2  
## q5a6 0.25 0.15 0.59 0.15 0.46 0.54 1.6  
## q5a17 0.34 0.18 0.46 0.34 0.48 0.52 3.2  
## q5a24 0.07 0.07 0.35 0.16 0.15 0.85 1.6  
## q5a12 0.21 0.15 0.21 0.17 0.14 0.86 3.7  
## q5a16 0.19 0.27 0.23 0.72 0.68 0.32 1.7  
## q5a14 0.19 0.32 0.16 0.56 0.48 0.52 2.1  
## q5a15 0.12 0.14 0.31 0.55 0.43 0.57 1.8  
##   
## PA1 PA2 PA3 PA4  
## SS loadings 2.23 2.10 1.74 1.60  
## Proportion Var 0.16 0.15 0.12 0.11  
## Cumulative Var 0.16 0.31 0.43 0.55  
## Proportion Explained 0.29 0.27 0.23 0.21  
## Cumulative Proportion 0.29 0.57 0.79 1.00  
##   
## Mean item complexity = 1.8  
## Test of the hypothesis that 4 factors are sufficient.  
##   
## The degrees of freedom for the null model are 91 and the objective function was 6.04 with Chi Square of 44076.49  
## The degrees of freedom for the model are 41 and the objective function was 0.08   
##   
## The root mean square of the residuals (RMSR) is 0.02   
## The df corrected root mean square of the residuals is 0.02   
##   
## The harmonic number of observations is 7308 with the empirical chi square 328.96 with prob < 1.3e-46   
## The total number of observations was 7308 with Likelihood Chi Square = 563.63 with prob < 4.8e-93   
##   
## Tucker Lewis Index of factoring reliability = 0.974  
## RMSEA index = 0.042 and the 90 % confidence intervals are 0.039 0.045  
## BIC = 198.86  
## Fit based upon off diagonal values = 1  
## Measures of factor score adequacy   
## PA1 PA2 PA3 PA4  
## Correlation of (regression) scores with factors 0.92 0.90 0.85 0.81  
## Multiple R square of scores with factors 0.84 0.80 0.71 0.66  
## Minimum correlation of possible factor scores 0.68 0.61 0.43 0.32

fa.sort(yy$loadings)

##   
## Loadings:  
## PA1 PA2 PA3 PA4   
## q5a2 0.860 0.200 0.181 0.148  
## q5a1 0.727 0.234 0.220 0.171  
## q5a3 0.701 0.172 0.197 0.152  
## q5a23 0.195 0.824 0.143 0.177  
## q5a21 0.227 0.750 0.174 0.216  
## q5a22 0.220 0.667 0.162 0.306  
## q5a18 0.149 0.114 0.797 0.150  
## q5a6 0.249 0.148 0.594 0.150  
## q5a17 0.343 0.183 0.457 0.343  
## q5a24 0.346 0.156  
## q5a12 0.209 0.154 0.215 0.174  
## q5a16 0.187 0.266 0.235 0.720  
## q5a14 0.187 0.318 0.165 0.559  
## q5a15 0.120 0.142 0.310 0.550  
##   
## PA1 PA2 PA3 PA4  
## SS loadings 2.232 2.100 1.738 1.597  
## Proportion Var 0.159 0.150 0.124 0.114  
## Cumulative Var 0.159 0.309 0.434 0.548

fa.diagram(yy)

 #2b  
I suppose the editors could have added communalities, total percentage explained and individual contributions, uniqueness, correlations of regressions, multiple R squared, minimum correlations, mean item complexity. I might have added a diagram to explain the factor loadings so it’s easy to read, kinda like the graph I have provided.

# 2c

Frankly, four factors seems to work well. running it with five factors does not increase the amount explained that much (4%).

yyz<-fa(schumm, nfactors = 5, rotate = "varimax", fm = "pa")

## maximum iteration exceeded

fa.sort(yyz)

## Factor Analysis using method = pa  
## Call: fa(r = schumm, nfactors = 5, rotate = "varimax", fm = "pa")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## PA1 PA2 PA3 PA4 PA5 h2 u2 com  
## q5a2 0.86 0.21 0.17 0.15 0.07 0.84 0.16 1.3  
## q5a1 0.71 0.23 0.19 0.16 0.24 0.67 0.33 1.8  
## q5a3 0.70 0.18 0.19 0.16 0.05 0.59 0.41 1.4  
## q5a23 0.19 0.82 0.13 0.18 0.08 0.77 0.23 1.3  
## q5a21 0.22 0.75 0.16 0.22 0.08 0.69 0.31 1.5  
## q5a22 0.21 0.67 0.15 0.30 0.12 0.61 0.39 1.8  
## q5a18 0.15 0.12 0.83 0.16 0.03 0.75 0.25 1.2  
## q5a6 0.25 0.15 0.56 0.16 0.13 0.44 0.56 1.8  
## q5a17 0.34 0.19 0.44 0.35 0.09 0.48 0.52 3.4  
## q5a24 0.06 0.07 0.33 0.15 0.14 0.16 0.84 2.0  
## q5a16 0.18 0.27 0.21 0.72 0.09 0.68 0.32 1.6  
## q5a15 0.12 0.14 0.29 0.56 0.07 0.43 0.57 1.8  
## q5a14 0.18 0.32 0.15 0.55 0.12 0.47 0.53 2.1  
## q5a12 0.15 0.12 0.17 0.13 0.60 0.44 0.56 1.5  
##   
## PA1 PA2 PA3 PA4 PA5  
## SS loadings 2.17 2.10 1.66 1.60 0.52  
## Proportion Var 0.16 0.15 0.12 0.11 0.04  
## Cumulative Var 0.16 0.31 0.42 0.54 0.57  
## Proportion Explained 0.27 0.26 0.21 0.20 0.06  
## Cumulative Proportion 0.27 0.53 0.74 0.94 1.00  
##   
## Mean item complexity = 1.8  
## Test of the hypothesis that 5 factors are sufficient.  
##   
## The degrees of freedom for the null model are 91 and the objective function was 6.04 with Chi Square of 44076.49  
## The degrees of freedom for the model are 31 and the objective function was 0.04   
##   
## The root mean square of the residuals (RMSR) is 0.01   
## The df corrected root mean square of the residuals is 0.02   
##   
## The harmonic number of observations is 7308 with the empirical chi square 145.56 with prob < 9.2e-17   
## The total number of observations was 7308 with Likelihood Chi Square = 284.28 with prob < 1e-42   
##   
## Tucker Lewis Index of factoring reliability = 0.983  
## RMSEA index = 0.033 and the 90 % confidence intervals are 0.03 0.037  
## BIC = 8.49  
## Fit based upon off diagonal values = 1  
## Measures of factor score adequacy   
## PA1 PA2 PA3 PA4  
## Correlation of (regression) scores with factors 0.92 0.90 0.87 0.81  
## Multiple R square of scores with factors 0.84 0.80 0.75 0.66  
## Minimum correlation of possible factor scores 0.69 0.61 0.50 0.33  
## PA5  
## Correlation of (regression) scores with factors 0.64  
## Multiple R square of scores with factors 0.41  
## Minimum correlation of possible factor scores -0.19

fa.sort(yyz$loadings)

##   
## Loadings:  
## PA1 PA2 PA3 PA4 PA5   
## q5a2 0.862 0.205 0.172 0.154   
## q5a1 0.708 0.229 0.193 0.157 0.241  
## q5a3 0.705 0.178 0.189 0.161   
## q5a23 0.187 0.824 0.131 0.176   
## q5a21 0.221 0.752 0.162 0.216   
## q5a22 0.208 0.666 0.145 0.301 0.120  
## q5a18 0.150 0.120 0.831 0.159   
## q5a6 0.245 0.148 0.565 0.162 0.128  
## q5a17 0.342 0.186 0.439 0.354   
## q5a24 0.327 0.154 0.136  
## q5a16 0.181 0.267 0.214 0.725   
## q5a15 0.118 0.144 0.292 0.556   
## q5a14 0.176 0.318 0.145 0.552 0.125  
## q5a12 0.151 0.123 0.173 0.127 0.596  
##   
## PA1 PA2 PA3 PA4 PA5  
## SS loadings 2.173 2.097 1.659 1.598 0.520  
## Proportion Var 0.155 0.150 0.119 0.114 0.037  
## Cumulative Var 0.155 0.305 0.424 0.538 0.575

# 3a

Looking at age,sex,education category and basevole, we see that rich older females who are college educated and already drinking are more likely to drink wine. There is a good model because of the odds ratio but. The best variables for prediction are probably age and sex.

load("C:/Users/Branly Mclanbry/Downloads/driver.RData")  
mod.3 <- glm(pdrink ~ age + sex + educar + income + basevol, data = driver, family = binomial())  
summary(mod.3)

##   
## Call:  
## glm(formula = pdrink ~ age + sex + educar + income + basevol,   
## family = binomial(), data = driver)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0394 -0.8118 -0.4649 0.9113 2.4771   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.570918 0.484132 -7.376 1.63e-13 \*\*\*  
## age 0.031266 0.006104 5.122 3.03e-07 \*\*\*  
## sex2 1.752883 0.187501 9.349 < 2e-16 \*\*\*  
## educar2 -0.004214 0.333381 -0.013 0.989915   
## educar3 0.658819 0.324038 2.033 0.042037 \*   
## income 0.266754 0.076709 3.477 0.000506 \*\*\*  
## basevol -0.010722 0.003420 -3.135 0.001717 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 919.07 on 688 degrees of freedom  
## Residual deviance: 733.44 on 682 degrees of freedom  
## AIC: 747.44  
##   
## Number of Fisher Scoring iterations: 5

exp(coef(mod.3))

## (Intercept) age sex2 educar2 educar3 income   
## 0.02813001 1.03175956 5.77121939 0.99579482 1.93250943 1.30571915   
## basevol   
## 0.98933555

exp(confint(mod.3))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 0.01061703 0.07100147  
## age 1.01959643 1.04432532  
## sex2 4.01638314 8.38242206  
## educar2 0.52238830 1.93882356  
## educar3 1.03629037 3.70760023  
## income 1.12492483 1.52016329  
## basevol 0.98233480 0.99561139

modchi <- mod.3$null.deviance-mod.3$deviance  
chidf <- mod.3$df.null - mod.3$df.residual  
chi.pr <- 1-pchisq(modchi,chidf)  
chi.pr

## [1] 0

modchi

## [1] 185.6219

table(mod.3$y,fitted(mod.3)>.5)

##   
## FALSE TRUE  
## 0 348 75  
## 1 104 162

mod.age <- glm(pdrink ~ sex + educar + income + basevol, data = driver, family = binomial())  
mod.sex <- glm(pdrink ~ age + educar + income + basevol, data = driver, family = binomial())  
mod.educar <- glm(pdrink ~ age + sex + income + basevol, data = driver, family = binomial())  
mod.income <- glm(pdrink ~ age + sex + educar + basevol, data = driver, family = binomial())  
mod.basevol <- glm(pdrink ~ age + sex + educar + income, data = driver, family = binomial())  
anova(mod.3,mod.age, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: pdrink ~ age + sex + educar + income + basevol  
## Model 2: pdrink ~ sex + educar + income + basevol  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 682 733.44   
## 2 683 760.59 -1 -27.143 1.889e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(mod.3,mod.sex, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: pdrink ~ age + sex + educar + income + basevol  
## Model 2: pdrink ~ age + educar + income + basevol  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 682 733.44   
## 2 683 830.22 -1 -96.772 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(mod.3,mod.educar, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: pdrink ~ age + sex + educar + income + basevol  
## Model 2: pdrink ~ age + sex + income + basevol  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 682 733.44   
## 2 684 745.54 -2 -12.1 0.002358 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(mod.3,mod.income, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: pdrink ~ age + sex + educar + income + basevol  
## Model 2: pdrink ~ age + sex + educar + basevol  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 682 733.44   
## 2 683 745.88 -1 -12.431 0.0004222 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(mod.3,mod.basevol, test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: pdrink ~ age + sex + educar + income + basevol  
## Model 2: pdrink ~ age + sex + educar + income  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 682 733.44   
## 2 683 745.48 -1 -12.037 0.0005216 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# 3b

Looking at age,sex,education category and basevole, we see that rich older females who are college educated and already drinking are more likely to drink wine (chi^2 = 185.62, *p* <.001). This model further breaks down into age (*b* = .03, *p* < .001), females (*b* = 1.75, *p* <.001) education at the college level (*b* = .65, *p* <.05), income (*b* = .27, *p* < .001), and base volume (*b* = -0.01, *p* < .01).

|  |  |  |
| --- | --- | --- |
| Variable | Exp(*b*) | CI |
| Age | 1.03 | [1.02,1.04] |
| Sex (Female) | 5.77 | [4.02,8.38] |
| Education (high school) | 1.00 | [0.52,1.94] |
| Education (College) | 1.93 | [1.04,3.71] |
| Income | 1.31 | [1.13,1.52] |
| Base Volume | 0.99 | [0.98,0.99] |
|  |  |  |

# 4

Teacher expectation is related to achievement (*b* = .71, *95% CI* [0.38,1.11]), and climate (*b* = .84, *95% CI* [0.41,1.11]), while climate is related to achievement (*b* = 0.42, *95% CI*[0.12,0.69]). The indirect effect is significant (*b* = 0.25, *95% CI[0.10,0.95]). One thing is cool is that the c path drops out(*b\* = 0.35, *95% CI* [-0.08,0.95]), so while this does not matter much, it is interesting.

Climate fully mediates the relationship between teacher expectations and achievement.

load("C:/Users/Branly Mclanbry/Downloads/hw4\_med.RData")  
model4 <- function(iv, dv, med, data, samples=50) {  
 model <- paste0(med, " ~ a\*", iv, "  
 ", dv, " ~ b\*", med, " + cp\*", iv, "  
 ind := a\*b  
 c := cp + a\*b")  
 set.seed(1839)  
 out <- lavaan::parameterEstimates(  
 lavaan::sem(model=model, data=data, se="boot", bootstrap=samples),   
 boot.ci.type="bca.simple")  
 out[7,c(6:8)] <- NA  
 out <- out[c(1,2,3,7,8),c(4:10)]  
 rownames(out) <- 1:nrow(out)  
 return(out)  
}  
outcome <- model4("Teacher","Achieve","Climate",hw4\_med)

## Warning in norm.inter(t, adj.alpha): extreme order statistics used as  
## endpoints  
  
## Warning in norm.inter(t, adj.alpha): extreme order statistics used as  
## endpoints  
  
## Warning in norm.inter(t, adj.alpha): extreme order statistics used as  
## endpoints  
  
## Warning in norm.inter(t, adj.alpha): extreme order statistics used as  
## endpoints  
  
## Warning in norm.inter(t, adj.alpha): extreme order statistics used as  
## endpoints

knitr::kable(outcome)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| label | est | se | z | pvalue | ci.lower | ci.upper |
| a | 0.8401378 | 0.1639070 | 5.125697 | 0.0000003 | 0.4101186 | 1.1066844 |
| b | 0.4237239 | 0.1428879 | 2.965428 | 0.0030226 | 0.1198427 | 0.6832211 |
| cp | 0.3517740 | 0.2198672 | 1.599938 | 0.1096123 | -0.0812857 | 0.9532225 |
| ind | 0.3559864 | NA | NA | NA | 0.1022089 | 0.6232454 |
| c | 0.7077605 | 0.1622913 | 4.361049 | 0.0000129 | 0.3839633 | 1.1134424 |