IST772– Problem Set 8

Sathish Kumar Rajendiran

Attribution statement: 1. I did this homework by myself, with help from the book and the professor.

# Chapter 8, Exercise 2

*The data sets package in R contains a small data set called swiss that contains n = 47 observations of socio-economic indicators for each of 47 French-speaking provinces of Switzerland at about 1888. Use “?swiss” to display help about the data set.*

?swiss   
# Swiss Fertility and Socioeconomic Indicators (1888) Data  
# Standardized fertility measure and socio-economic indicators for each of 47 French-speaking provinces of Switzerland at about 1888.  
  
# A data frame with 47 observations on 6 variables, each of which is in percent, i.e., in [0, 100].  
#   
# [,1] Fertility Ig, ‘common standardized fertility measure’  
# [,2] Agriculture % of males involved in agriculture as occupation  
# [,3] Examination % draftees receiving highest mark on army examination  
# [,4] Education % education beyond primary school for draftees.  
# [,5] Catholic % ‘catholic’ (as opposed to ‘protestant’).  
# [,6] Infant.Mortality live births who live less than 1 year.  
# All variables but ‘Fertility’ give proportions of the population.  
  
# summary(swiss)  
  
dim(swiss)

## [1] 47 6

# View(swiss)  
# define variables  
Fertility <- swiss$Fertility  
Education <- swiss$Education  
Examination <- swiss$Examination  
Catholic <- swiss$Catholic  
Infant <- swiss$Infant  
Agriculture <- swiss$Agriculture

*Create and interpret a bivariate correlation matrix using cor(swiss) keeping in mind the idea that you will be trying to predict the Fertility variable. Which other variable might be the single best predictor? (1 pt)*

# 1) Question 1: BiVariate normality/Correlation matrix

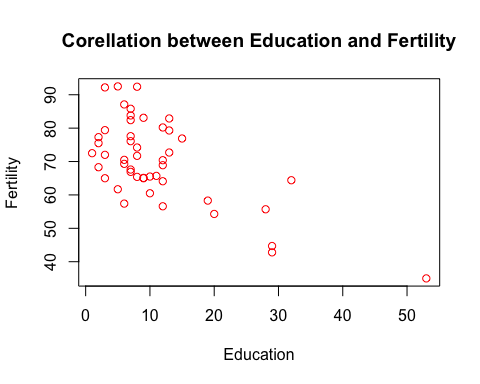
* cor(swiss)
  + Based on the correlation matrix, We find that Education variable has higher (negative) impact on the Feritility (dependent) variable.
  + Please find the table below,

#correlation on swiss dataset  
cor(swiss)

## Fertility Agriculture Examination Education Catholic  
## Fertility 1.0000000 0.35307918 -0.6458827 -0.66378886 0.4636847  
## Agriculture 0.3530792 1.00000000 -0.6865422 -0.63952252 0.4010951  
## Examination -0.6458827 -0.68654221 1.0000000 0.69841530 -0.5727418  
## Education -0.6637889 -0.63952252 0.6984153 1.00000000 -0.1538589  
## Catholic 0.4636847 0.40109505 -0.5727418 -0.15385892 1.0000000  
## Infant.Mortality 0.4165560 -0.06085861 -0.1140216 -0.09932185 0.1754959  
## Infant.Mortality  
## Fertility 0.41655603  
## Agriculture -0.06085861  
## Examination -0.11402160  
## Education -0.09932185  
## Catholic 0.17549591  
## Infant.Mortality 1.00000000

\* Based on our selection, Please find the correlation plot between the predictor (education) and outcome/dependent variable "feritility" below  
\* It clearly shows that , lesser the Education , stronger the fertility rate is.

#Correlation plot  
plot(Education, Fertility  
 ,main="Corellation between Education and Fertility"  
 ,col="red")

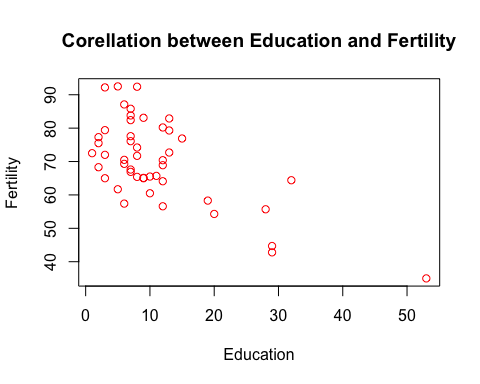


# Please find more details from the R code below,

cor(swiss)

## Fertility Agriculture Examination Education Catholic  
## Fertility 1.0000000 0.35307918 -0.6458827 -0.66378886 0.4636847  
## Agriculture 0.3530792 1.00000000 -0.6865422 -0.63952252 0.4010951  
## Examination -0.6458827 -0.68654221 1.0000000 0.69841530 -0.5727418  
## Education -0.6637889 -0.63952252 0.6984153 1.00000000 -0.1538589  
## Catholic 0.4636847 0.40109505 -0.5727418 -0.15385892 1.0000000  
## Infant.Mortality 0.4165560 -0.06085861 -0.1140216 -0.09932185 0.1754959  
## Infant.Mortality  
## Fertility 0.41655603  
## Agriculture -0.06085861  
## Examination -0.11402160  
## Education -0.09932185  
## Catholic 0.17549591  
## Infant.Mortality 1.00000000

#Correlation plot  
plot(Education, Fertility  
 ,main="Corellation between Education and Fertility"  
 ,col="red")

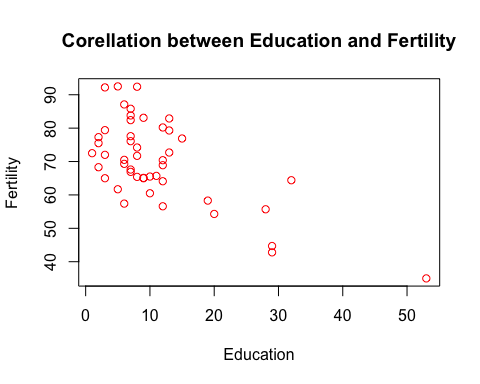


# 2) Question 2: Most Correlated

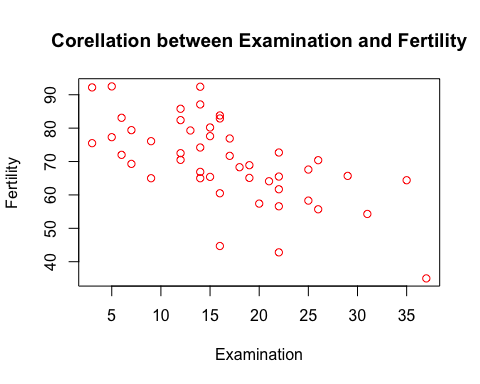
* Scatter Plot/Correlation between variables,Based on the correlation matrix
  + Education variable has higher (negative) impact on the Feritility (dependent) variable.
  + Examination does have negative impact on the Fertility variable.However, thre are possible outliers that impacts the baseline curve between these two variables
  + Catholic has two extreme cases. Either strong/greater Catholic values increases Fertility rate or lowest/much less Catholic increases Fertility rate. In between, there very few entries suggesting either possible outliers or defines thats how the equation is.
  + Infant mortality and Agriculture both favors postive impact. Howver, with more outlier/spread data - these two may impact the prediction accuracy. In other words - may induce more residuals/R-Squared values.

# Please find more details from the Plots below,

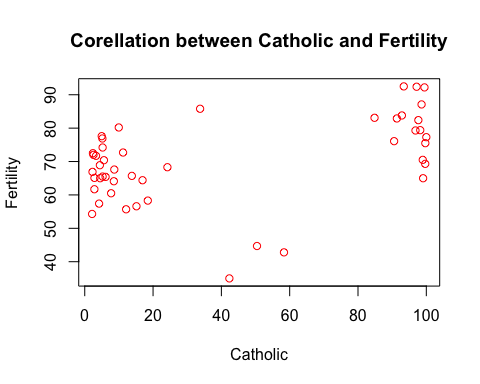
# Correlation plot  
# Corellation between Education and Fertility  
plot(Education, Fertility  
 ,main="Corellation between Education and Fertility"  
 ,col="red")



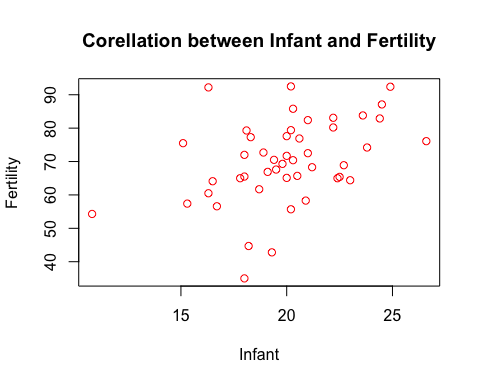
# arrows(min(Education),Fertility[which.min(Education)],min(Education),min(Education)\*56)  
  
# Corellation between Examination and Fertility  
  
plot(Examination, Fertility  
 ,main="Corellation between Examination and Fertility"  
 ,col="red")



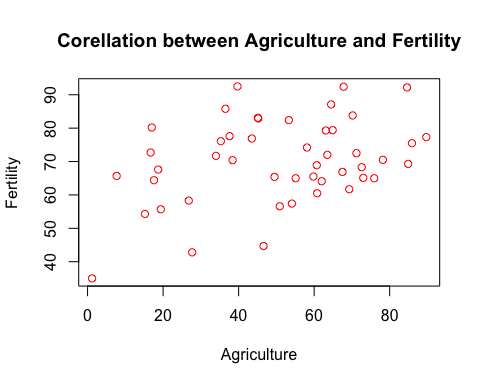
# Corellation between Catholic and Fertility  
plot(Catholic, Fertility  
 ,main="Corellation between Catholic and Fertility"  
 ,col="red")



# Corellation between Infant and Fertility  
plot(Infant, Fertility  
 ,main="Corellation between Infant and Fertility"  
 ,col="red")



# Corellation between Agriculture and Fertility  
plot(Agriculture, Fertility  
 ,main="Corellation between Agriculture and Fertility"  
 ,col="red")



options(scipen=999) # turn-off scientific notation like 1e+48  
# install.packages("dlookr") # note: requires version 3.5.2 or higher  
library(dlookr)

## Loading required package: mice

##   
## Attaching package: 'mice'

## The following object is masked from 'package:stats':  
##   
## filter

## The following objects are masked from 'package:base':  
##   
## cbind, rbind

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

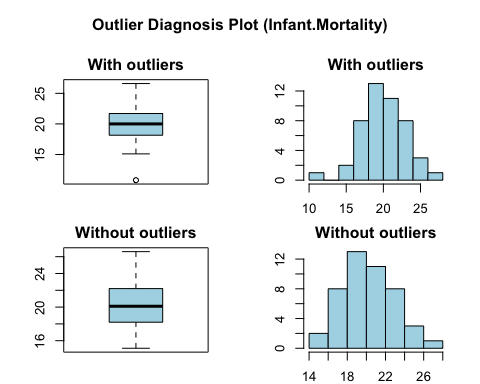
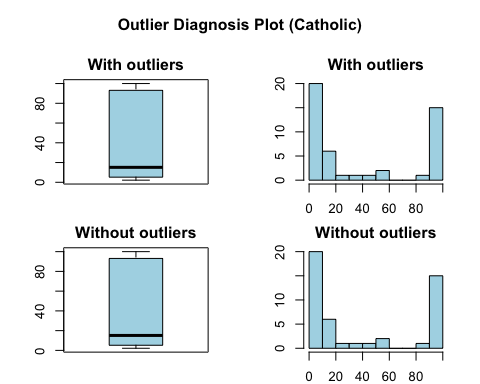
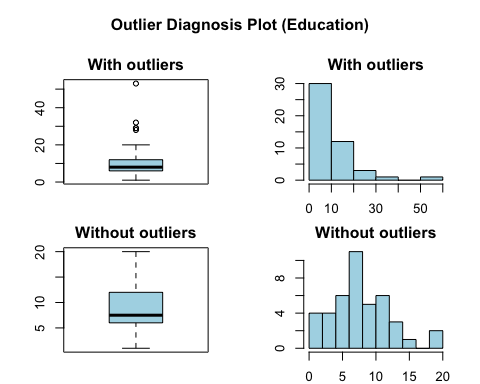
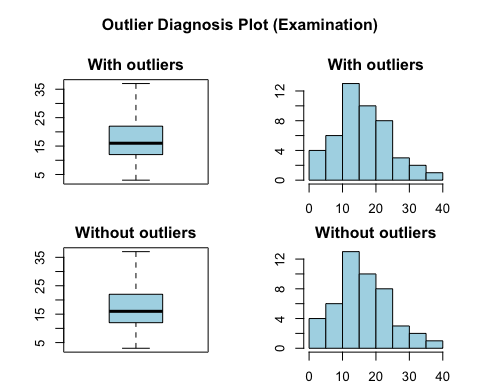
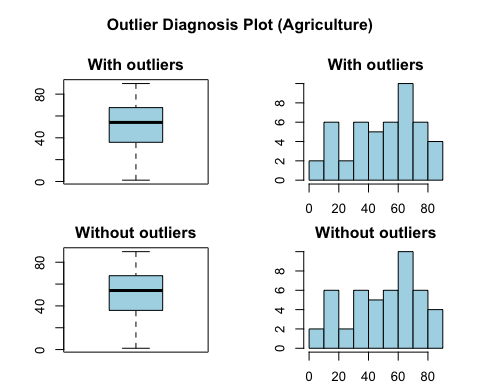
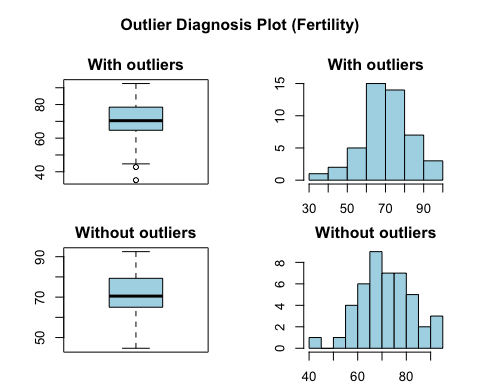
##   
## Attaching package: 'dlookr'

## The following object is masked from 'package:base':  
##   
## transform

diagnose\_outlier(swiss)

## variables outliers\_cnt outliers\_ratio outliers\_mean with\_mean  
## 1 Fertility 2 4.255319 38.9 70.14255  
## 2 Agriculture 0 0.000000 NaN 50.65957  
## 3 Examination 0 0.000000 NaN 16.48936  
## 4 Education 5 10.638298 34.2 10.97872  
## 5 Catholic 0 0.000000 NaN 41.14383  
## 6 Infant.Mortality 1 2.127660 10.8 19.94255  
## without\_mean  
## 1 71.531111  
## 2 50.659574  
## 3 16.489362  
## 4 8.214286  
## 5 41.143830  
## 6 20.141304

plot\_outlier(swiss)



# Chapter 8, Exercise 3

*Run a multiple regression analysis on the swiss data with lm(), using Fertility as the dependent variable and Education and Agriculture as the predictors. (1 pt) Check the diagnostics. (1 pt) Say whether or not the overall R‐squared was significant. If it was significant, report the value and say in your own words whether it seems like a strong result or not. Review the significance tests on the coefficients (B‐weights). For each one that was significant, report its value and say in your own words whether it seems like a strong result or not. (1 pt)*

# 3) Question 3: Linear Regression Analysis

* lm() | Dependent variable (Fertility) & Education and Agriculture as Predictors/Independent variables
  + lm(Fertility ~ Education + Agriculture,data=swiss) is stored on a variable “swiss.lm”
  + above code runs a regression analysis on Fertility as the dependent variable and Education and Agriculture as the predictors from “swiss” dataset.
* Diagnostics:
  + From the lm(formula = Fertility ~ Education + Agriculture, data = swiss); this formula predicts “Fertility” values from Education and Agriculture combined from the dataset “swiss”
  + As stated in earlier sections, Education and Agriculture had different impact on the fertility effect.
  + Procedure lm() is similar to aov() procedure and expects the dependent variable (variable, in question for prediction) first. Followed by, independent/prdictor variable. It can have number of predictor variables follows after “~” symbol. “.” after “~” means - it expects to include all the remaining variables from the dataset. In this case, we are only trying to predict “fertility” rate based on “Education” & “Agriculture” variable.
  + In addition, “data=swiss” implies what is the sample/population the prediction is run against from the observations it contains.
  + Successul execution of the lm() procedure provides results as shown below.
    - call
    - Residuals
    - Coefficients
    - Significant codes
    - Residual standard error with degrees of freedom
    - Multiple R-Squared
    - F-Static and p-value
* Results:
  + Once the lm() procedure executes successfuly, it returns various data points as an outcome. The first two lines defines the model, we wanted.
    - lm(formula = Fertility ~ Education + Agriculture, data = swiss)
  + Next, Summary of residuals that gives an overview of errors of prediction.
    - With min as -17.3072 and max of 20.5291 shows, larger spread between -ve to positive almost spreading equally on both sides, with Median almost 0 ( -0.9443).
    - It seems the residuals are symmetrically distributed.
  + Coefficients shows the key results.
    - Intercept in high above Y-axis at 84.08
    - Slopes for Education variable is -0.96276 and Agriculture is -0.06648 are way off from the intercept or or B-weights. It defines the orientation of the best fitting hyperplane.
    - Std.Errors around the estimates of slope and intercept shows the estimated sampling distribution around these point estimates.
    - t-value shows the student’s t-test of the null hypothesis test that each estimated coefficients is equal to zero.
    - three asterik (\*) indicates the significance level of alpha at p < 0.001.
    - With above p-value; Education has the strong coefficient value as 0.0000071 far less than the test of significance (p < 0.001); Where as Agriculture has p-value as 0.411. This shows that , we can ignore the “Agriculture”
    - Multiple R-Squared value is at 0.4492 - which means ~45% variability in Fertility is based on these predictor variables. its comparatively better for a bivariate predictor model.
    - The standard error of the residuals is shown as 9.479 on 44 degrees of freedom. Starting with 47 observations, two degrees of freedom is lost for calculating the slopes (coeffients on Education and Agriculture or B-weights) and one degree of freedom is lost for calculating the Y-intercept.
    - Finally, F-Statistic (2,44) with 17.95 with p-value 0.000002 tests the null hypothesis that R-Squared value is equal to 0. Clearly, based on these results we can reject the null hypothesis.

# Please find more details from the R code below,

options(scipen=999) # turn-off scientific notation like 1e+48  
swiss.lm <- lm(Fertility ~ Education + Agriculture,data=swiss)  
summary(swiss.lm)

##   
## Call:  
## lm(formula = Fertility ~ Education + Agriculture, data = swiss)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.3072 -6.6157 -0.9443 8.7028 20.5291   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 84.08005 5.78180 14.542 < 0.0000000000000002 \*\*\*  
## Education -0.96276 0.18906 -5.092 0.0000071 \*\*\*  
## Agriculture -0.06648 0.08005 -0.830 0.411   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.479 on 44 degrees of freedom  
## Multiple R-squared: 0.4492, Adjusted R-squared: 0.4242   
## F-statistic: 17.95 on 2 and 44 DF, p-value: 0.000002

# Chapter 8, Exercise 4

*Using the results of the analysis from Exercise 2, construct a prediction equation for Fertility using all three of the coefficients from the analysis (the intercept along with the two B‐weights). If you observed a province with scores for Education of 8 and Agriculture of 35, what would you predict the Fertility rate to be (you can use your equation or the predict function)? Show your calculation and the resulting value of Fertility. (1 pt)*

# 4) Question 4: Predict New value

* New value of predictor variable (Fertility) is calculated in linear formula as below
  + NewFertility = 84.08005 + (-0.96276 \* Education) + (-0.06648 \* Agriculture)
  + result is ~74 at scores for Education of 8 and Agriculture of 35 with above resulting values.

# Please find more details from the R code below,

# Predict manually the "feritility value from above variables "  
NewFertility = 84.08005 + (-0.96276 \* Education) + (-0.06648 \* Agriculture)  
NewFertility <- 84.08005 + (-0.96276 \* 8) + (-0.06648 \* 35)  
NewFertility

## [1] 74.05117

# Chapter 8, Exercise 5

*Run a multiple regression analysis on the swiss data with lmBF(), using Fertility as the dependent variable and Education and Agriculture as the predictors. (1 pt) Interpret the resulting Bayes factor in terms of the odds in favor of the alternative hypothesis (be sure to note the hypotheses being compared). Do these results strengthen or weaken your conclusion to exercise 2? (1 pt)*

# 5) Question 5: Bayesian Approach to Multiple Regression Analysis

* lmBF() | Dependent variable (Fertility) & Education and Agriculture as Predictors/Independent variables
  + lmBF(Fertility ~ Education + Agriculture, data = swiss) is stored on a variable “swissOutBF”
  + above code runs a bayesian regression analysis on Fertility as the dependent variable and Education and Agriculture as the predictors from “swiss” dataset.
* Results:
  + Once the lmBF() procedure executes successfuly, it returns Bayes factor analyis as an outcome. The first two lines defines the model, we wanted.
    - [1] Education + Agriculture : 8927.474 ±0%
    - It shows that the odds are overwhelmingly in favor of alternate hypothesis at 8927.474 ±0% ratio, in the sense that a model containing Education and Agriculture as predictors is hugley favored over a model that only contains the Y-intercept.
    - i.e. Y-Intercept only model means that we are forcing all of the B-weights on the predictors to be 0
    - Bayes factor is an odds ratio showing the likelihood of the alternative hypothesis (in this case a model with nonzero wieghts for the predictors) divided by the liklihood of the null model (in this case intercept- only model)
  + Bayes factor type: BFlinearModel, JZS
  + Comparison to lm(): NHST vs Alternative Hypothesis
    - Earlier, linear regression analysis proved that p-value is much lower than the alpha value, hence we rejected the Null hypothesis
    - Bayesian factor also, favored alternate hypothesis at 8927.474 ±0% ratio. So, it does strenghten our exercise 2.

# Please find more details from the R code below,

options(scipen=999) # turn-off scientific notation like 1e+48  
# install.packages("BayesFactor")  
library("BayesFactor")

## Loading required package: coda

## Loading required package: Matrix

## \*\*\*\*\*\*\*\*\*\*\*\*  
## Welcome to BayesFactor 0.9.12-4.2. If you have questions, please contact Richard Morey (richarddmorey@gmail.com).  
##   
## Type BFManual() to open the manual.  
## \*\*\*\*\*\*\*\*\*\*\*\*

swissOutBF <- lmBF(Fertility ~ Education + Agriculture, data = swiss)  
summary(swissOutBF)

## Bayes factor analysis  
## --------------  
## [1] Education + Agriculture : 8927.474 ±0%  
##   
## Against denominator:  
## Intercept only   
## ---  
## Bayes factor type: BFlinearModel, JZS

# Chapter 8, Exercise 6

*Run lmBF() with the same model as for Exercise 4, but with the options posterior=TRUE and iterations=10000. (1 pt) Interpret the resulting information about the coefficients. (1 pt)*

# 6) Question 6: Bayesian Approach to Multiple Regression Analysis with Posterior distribution

* lmBF() | Dependent variable (Fertility) & Education and Agriculture as Predictors/Independent variables with Posterior = True and Iterations =10000
  + lmBF(Fertility ~ Education + Agriculture, data = swiss, posterior=TRUE, iterations=10000) is stored on a variable “swissOutMCMC”
  + above code runs a bayesian regression analysis on Fertility as the dependent variable and Education and Agriculture as the predictors from “swiss” dataset with Posterior distribution having 10000 iterations
* Results:
  + Once the lmBF() procedure executes successfuly, summary results shares several similarities with conventional analysis along with striking differences
    - Mean B-weights of differences of Education (-0.8898) and Agriculture(-0.0621) are similar to what the conventional analysis provided
    - 95% HDI shows that there is a overlap between these two B-weights
    - 95% HDI of Agriculture overlaps with 0, providing evidence that the population value of that B-weight dies not credibly differ from 0
    - mean of the the R-Squarded values is 0.3888445.Slighlty lesser than the Adjusted R-Squared value of conventional analysis.
    - 95% HDI for the posterior distribution of R-Squaed values is between 0.08113708 (2.5%) and 0.5994509 (97.5%)
    - In addition, -0.06089 of Agriculture shows that the valus is much lower than the conventional analysis claiming that this variable is not a strong predictor.

# Please find more details from the R code below,

options(scipen=999) # turn-off scientific notation like 1e+48  
# install.packages("BayesFactor")  
library("BayesFactor")  
  
swissOutMCMC <- lmBF(Fertility ~ Education + Agriculture, data = swiss, posterior=TRUE, iterations=10000)  
summary(swissOutMCMC)

##   
## Iterations = 1:10000  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 10000   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE Time-series SE  
## mu 70.14258 1.43057 0.0143057 0.0147149  
## Education -0.88769 0.19592 0.0019592 0.0021243  
## Agriculture -0.06177 0.07924 0.0007924 0.0007924  
## sig2 95.71618 21.25900 0.2125900 0.2390655  
## g 0.85332 7.89901 0.0789901 0.0789901  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50% 75% 97.5%  
## mu 67.30063 69.2107 70.15434 71.07480 72.96180  
## Education -1.26875 -1.0177 -0.89097 -0.75518 -0.49369  
## Agriculture -0.21860 -0.1148 -0.06127 -0.00909 0.09432  
## sig2 62.90084 80.8504 92.74942 107.60234 144.76240  
## g 0.06648 0.1768 0.32490 0.66622 3.79018

rsqList <- 1 - (swissOutMCMC[,"sig2"] / var(swiss$Fertility))  
MeanrsqList <- mean(rsqList)  
MeanrsqList

## [1] 0.3866018

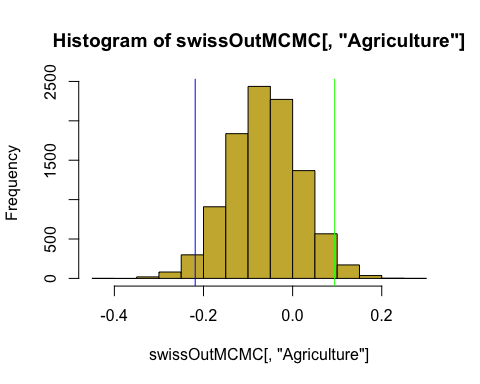
quantile(rsqList,c(0.025))

## 2.5%   
## 0.0722886

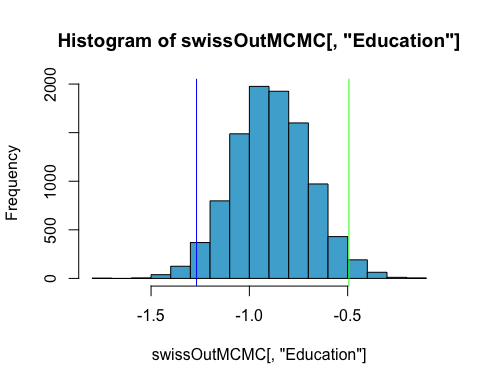
quantile(rsqList,c(0.975))

## 97.5%   
## 0.5968993

# histogram of 95% HDI mean differences on Agriculture | overlaps with 0  
hist(swissOutMCMC[,"Agriculture"],col="#CBB43D")  
abline(v=quantile(swissOutMCMC[,"Agriculture"],c(0.025)),col="blue")  
abline(v=quantile(swissOutMCMC[,"Agriculture"],c(0.975)),col="green")



# histogram of 95% HDI mean differences on Education | does not overlaps with 0  
hist(swissOutMCMC[,"Education"],col="#4DAFD4")  
abline(v=quantile(swissOutMCMC[,"Education"],c(0.025)),col="blue")  
abline(v=quantile(swissOutMCMC[,"Education"],c(0.975)),col="green")



# Chapter 8, Exercise 7

*Run install.packages() and library() for the “car” package. The car package is “companion to applied regression” rather than more data about automobiles. Read the help file for the vif() procedure and then look up more information online about how to interpret the results. Then write down in your own words a “rule of thumb” for interpreting vif. (1 pt)*

# 7) Question 7: Vif() procedure

* In multiple regression (Chapter @ref(linear-regression)), two or more predictor variables might be correlated with each other. This situation is referred as collinearity.There is an extreme situation, called multicollinearity, where collinearity exists between three or more variables even if no pair of variables has a particularly high correlation. This means that there is redundancy between predictor variables. In the presence of multicollinearity, the solution of the regression model becomes unstable.
* For a given predictor (p), multicollinearity can assessed by computing a score called the variance inflation factor (or VIF), which measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model.
* The smallest possible value of VIF is one (absence of multicollinearity). As a rule of thumb, a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity (James et al. 2014).
* When faced to multicollinearity, the concerned variables should be removed, since the presence of multicollinearity implies that the information that this variable provides about the response is redundant in the presence of the other variables (James et al. 2014,P. Bruce and Bruce (2017)).

# Please find more details from the R code below,

options(scipen=999) # turn-off scientific notation like 1e+48  
install.packages(pkgs=c("car"),repos = "http://cran.us.r-project.org")

##   
## The downloaded binary packages are in  
## /var/folders/\_z/ltmjkt4156b37rsk7cgvj7l80000gn/T//Rtmpv2NzrC/downloaded\_packages

library(car)

## Loading required package: carData

#“companion to applied regression”  
  
  
?vif() #Variance Inflation Factors

# Chapter 8, Exercise 8

*Run vif() on the results of the model from Exercise 3. (1 pt) Interpret the results. Then run a model that predicts Fertility from all five of the predictors in swiss. Run vif() on those results and interpret what you find. (1 pt)*

# 8) Question 8: vif() on conventional model with 2 predictors | from Exercise 3

* vif(swiss.lm) | Variance inflation factor analysis on a linear model with predictor variables Education and Agriculture
  + vif factor on Education and Agriculture variable is 1.692016
  + Values are in range of 1. Based on the rule of thumb the values are far less that 5 or 10. Showing not possible multi collinearity in the model
* vif(swiss.lmAll) | Variance inflation factor analysis on a linear model with all variables
  + predictor variables Infant.Mortality (1.107542), Catholic (1.937160) are far better variance inflation factors with values less than 2.
  + predictor variables Agriculture just above (2.284129)
  + predictor variables Education just below 3 (2.774943)
  + predictor variables Examination well above 3 (3.675420)
  + It is possible that when the variables are combined, Examination variable may lead towards multi-collinearity becoming a reduntant variable.

# Please find more details from the R code below,

summary(swiss.lm)

##   
## Call:  
## lm(formula = Fertility ~ Education + Agriculture, data = swiss)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.3072 -6.6157 -0.9443 8.7028 20.5291   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 84.08005 5.78180 14.542 < 0.0000000000000002 \*\*\*  
## Education -0.96276 0.18906 -5.092 0.0000071 \*\*\*  
## Agriculture -0.06648 0.08005 -0.830 0.411   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.479 on 44 degrees of freedom  
## Multiple R-squared: 0.4492, Adjusted R-squared: 0.4242   
## F-statistic: 17.95 on 2 and 44 DF, p-value: 0.000002

vif(swiss.lm) # vif() procedure on execise 3 linear regression model

## Education Agriculture   
## 1.692016 1.692016

# Run multiple linear regression with all variables  
options(scipen=999) # turn-off scientific notation like 1e+48  
swiss.lmAll <- lm(Fertility ~ .,data=swiss)   
summary(swiss.lmAll)

##   
## Call:  
## lm(formula = Fertility ~ ., data = swiss)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.2743 -5.2617 0.5032 4.1198 15.3213   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 66.91518 10.70604 6.250 0.000000191 \*\*\*  
## Agriculture -0.17211 0.07030 -2.448 0.01873 \*   
## Examination -0.25801 0.25388 -1.016 0.31546   
## Education -0.87094 0.18303 -4.758 0.000024306 \*\*\*  
## Catholic 0.10412 0.03526 2.953 0.00519 \*\*   
## Infant.Mortality 1.07705 0.38172 2.822 0.00734 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.165 on 41 degrees of freedom  
## Multiple R-squared: 0.7067, Adjusted R-squared: 0.671   
## F-statistic: 19.76 on 5 and 41 DF, p-value: 0.0000000005594

vif(swiss.lmAll) # vif() procedure on all variables from swiss.lmAll model

## Agriculture Examination Education Catholic   
## 2.284129 3.675420 2.774943 1.937160   
## Infant.Mortality   
## 1.107542