

```

rm( list=ls() )

setwd( '~/Dropbox/Today/School/EDUC_767/Teaching/Code/' )

dfData <- read.table( 'regress.dat', sep='\t', header=TRUE )

head( dfData )

# subjno - the unique subject identifier
# timedrs - number of visits to health professionals
# phyheal - number of physical health symptoms
# menheal - number of mental health symptoms
# stress - stress from acute life changes

tvDescriptive <- function( x ){
  z <- as.numeric( x )
  N <- sum( !is.na( x ) )
  NMiss <- sum( is.na( x ) )
  Mean <- mean( z )
  Median <- median( z )
  P25 <- quantile( z, .25 )
  P75 <- quantile( z, .75 )
  SD <- sd( z )
  Min <- min( z )
  Max <- max( z )
  SE <- SD / N^.5
  CV <- SD / Mean
  LCL <- Mean - SE * qt( .975, N - 1 )
  UCL <- Mean + SE * qt( .975, N - 1 )

  M3 <- sum( ( z - Mean )^3 ) # This is the 3rd moment about the mean
  Skew <- M3 / ( ( N - 1 ) * SD^3 )

  M4 <- sum( ( z - Mean )^4 ) # This is the 4th moment about the mean
  Kurtosis <- M4 / ( ( N - 1 ) * SD^4 )

  return( c( N=N, Missing=NMiss, Mean=Mean, Median=Median, 'P25'=P25, 'P75'=P75,
    SD=SD, SE=SE, Min=Min, Max=Max, CV=CV, Skew=Skew,
    Kurtosis=Kurtosis, LCL=LCL, UCL=UCL ) )
}

t( apply( dfData[ , 2:5 ], 2, tvDescriptive ) )

cor( dfData[ , 2:5 ] )

# this library is useful for creating nice looking plots
library( GGally )
ggpairs( dfData[ , 2:5 ] )

# this library is useful for diagnosing our regression model
library( car )

# Our VIF
diag( solve( cor( dfData[ , c( 'phyheal', 'menheal', 'stress' ) ] ) ) )

# Fit the model
summary( Fit1 <- lm( timedrs ~ phyheal + menheal + stress, data=dfData ) )

vif( Fit1 )

dfStuff <- data.frame( timedrs=dfData$timedrs, fitted.values=Fit1$fitted.values, residuals=Fit1$residuals
)

library( ggplot2 )

# Observed vs Expected scatter plot
ggplot(dfStuff, aes( x=fitted.values, y=timedrs ) ) +
  geom_point(shape=19) + # Use filled circles
  geom_smooth(method=lm, # Add linear regression line

```

```

se=FALSE)      # Don't add shaded confidence region

# Residual Plot
ggplot(dfStuff, aes( x=fitted.values, y=residuals ) ) +
  geom_point(shape=19) +      # Use filled circles
  geom_smooth(method=lm,      # Add linear regression line
              se=FALSE)      # Don't add shaded confidence region

# Standardize the residuals
sResiduals <- rstandard( Fit1 )
hist( sResiduals, freq=F, las=1, xlab='Standardized Residuals', main='Histogram of Residuals', col='navy'
)
curve( dnorm, add=T, col='red', lwd=3 )

plot( Fit1 )

d <- apply( dfData[ , 2:5 ], 2, FUN=function( x ) x - mean( x ) )

summary( lm( timedrs ~ phyheal + menheal + stress, data=data.frame( d ) ) )

```