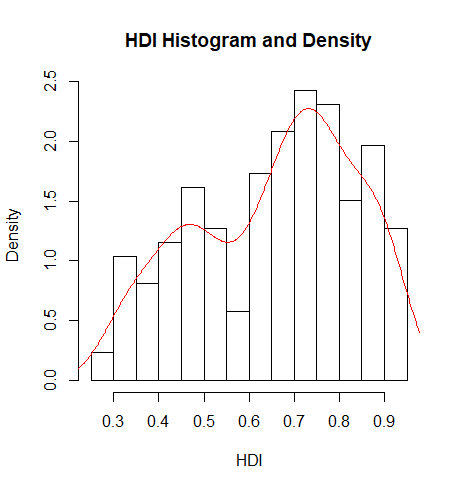
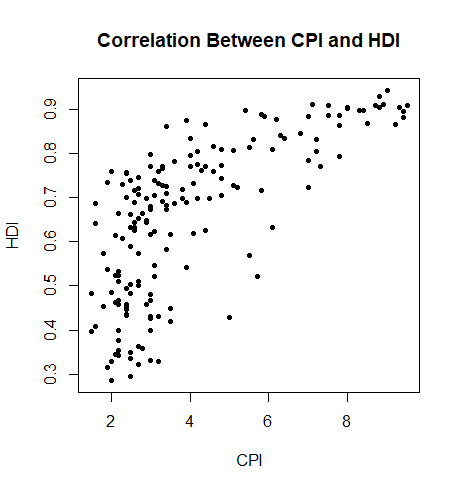
Human Development Index Analysis

**1. What are the X and Y variables in your analysis? Explain your reasoning for this determination.**

In my analysis the X (independent) variable is CPI (Corruption Perception Index). The Y (dependent) variable is HDI (Human Development Index). My reasoning for this is that countries with high degrees of corruption are likely to steal or squander their resources, negatively affecting the quality of life of their populace for the benefit of a few. Conversely, countries with low corruption are better environments for the economy and human development to flourish due to efficient and equitable distribution of resources.

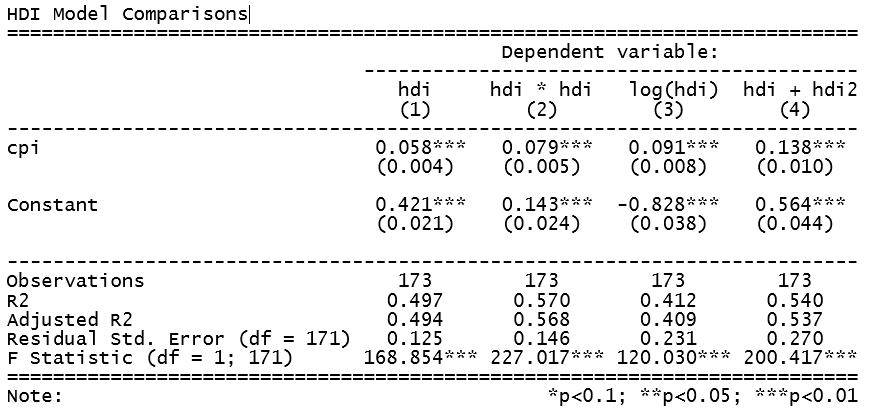
The alternative analysis is that a higher HDI leads to lower CPI and vice versa, however the cause and effect relationship is not as strong in this direction.

**2. What is the precise nature of the relationship (linear, quadratic, logarithmic, exponential, additive/multiplicative, ...) between HDI and CPI? Present appropriate evidence to support your conclusion.**



Looking at the plot of HDI against CPI, we can see that the datapoints are largely vertical between 2 and 4 CPI, before curving towards the right side of the plot. The histogram of HDI shows a non-normal distribution, with a right-skew and a minor and major peak in the frequency. Both of these indicate non-linearity. The shape created by the points in the plot appears to resemble a logarithmic function.

**3. Run appropriately specified regression model(s) to test the relationshi4 - p between CPI and HDI. Present the results of your analysis in a nicely formatted table.**



Of the transformations of HDI, it appears that the exponential function (2) of HDI results in the best predictive relationship, based upon the adjusted R-squared score.

**4. Provide a precise interpretation (in terms of marginal effects) of the strength of the relationship between CPI and HDI.**

ggpredict(mExponential,"cpi")

pred <- ggpredict(mExponential,"cpi")

plot(pred)

# Predicted values of hdi \* hdi

# x = cpi

x | Predicted | SE | 95% CI

------------------------------------

1 | 0.22 | 0.02 | [0.18, 0.26]

2 | 0.30 | 0.02 | [0.27, 0.33]

3 | 0.38 | 0.01 | [0.36, 0.40]

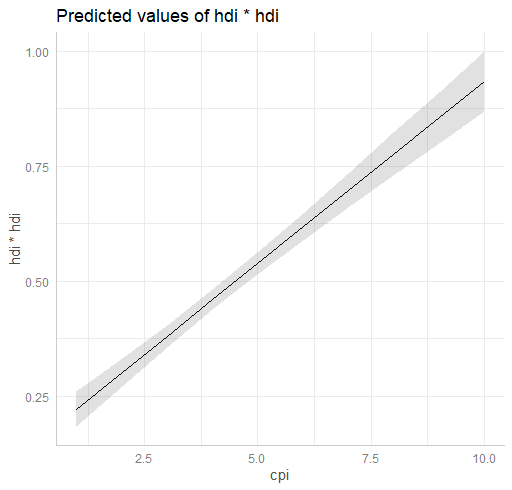
4 | 0.46 | 0.01 | [0.44, 0.48]

6 | 0.62 | 0.02 | [0.59, 0.65]

7 | 0.70 | 0.02 | [0.66, 0.73]

8 | 0.78 | 0.02 | [0.73, 0.82]

10 | 0.93 | 0.03 | [0.87, 1.00]



**5. How will readers know that your analysis is reliable and trustworthy?**

#Question 5

# Shapiro-Wilk's test of multivariate normality

shapiro.test(mExponential$res)

Shapiro-Wilk normality test

data: mExponential$res

W = 0.98733, p-value = 0.1218

Shows that residuals may be normally distributed.

# Bartlett's test of homoskedasticity

bartlett.test(list(mExponential$res, mExponential$fit))

Bartlett test of homogeneity of variances

data: list(mExponential$res, mExponential$fit)

Bartlett's K-squared = 3.4311, df = 1, p-value = 0.06398

Shows that data may have equal variances.

# Durbin-Watson test of autocorrelation

dwtest(mExponential)

Durbin-Watson test

data: mExponential

DW = 1.8845, p-value = 0.22

alternative hypothesis: true autocorrelation is greater than 0

Shows that data is not autocorrelated.