STAT 360: Computational Statistics and Data Analysis

Load Relevant R Libraries

```
library(rmarkdown); library(knitr); library(moments);
library(scatterplot3d); library(corrplot); library(pso)
library(psych); library(GPArotation); library(lavaan)
```

##SEM

```
data_matrix_csv <- read.csv("C:/Users/User/OneDrive - University of St. Thomas/Classes/STAT360/STAT360 F
life_expectancy <- as.matrix(data_matrix_csv)

col_names_unedited <- c(
  "Life expectancy at birth, total (years)",
  "CO2 emissions (metric tons per capita)",
  "Access to electricity (% of population)",
  "Current health expenditure (% of GDP)",
  "Out-of-pocket expenditure (% of current health expenditure)",
  "Domestic private health expenditure per capita, PPP (current international $)",
  "Domestic general government health expenditure per capita, PPP (current international $)",
  "Renewable internal freshwater resources per capita (cubic meters)",
  "Prevalence of HIV, total (% of population ages 15-49)",
  "Unemployment, total (% of total labor force) (national estimate)",
  "Government Effectiveness: Estimate",
  "Income share held by highest 10%",
  "Prevalence of current tobacco use (% of adults)",
  "Total alcohol consumption per capita (liters of pure alcohol, projected estimates, 15+ years of age)
  "Political Stability and Absence of Violence/Terrorism: Estimate",
  "Population density (people per sq. km of land area)"
)

# Better for displaying
col_names <- c(
  "Life_expectancy",
  "CO2_emissions",
  "Electricity",
  "Health_expenditure",
  "Out_of_pocket",
  "Private_health_expenditure",
  "Govt_health_expenditure",
  "Freshwater_resources",
  "HIV_prevalence",
  "Unemployment",
  "Govt_effectiveness",
  "Income_share",
  "Tobacco_use",
  "Alcohol_consumption",
  "Political_stability",
  "Population_density"
)
```

```r
colnames(life_expectancy) <- col_names
```

```r
SIG <- cov(life_expectancy, use = "pairwise.complete.obs")
EQN <- '
        # Measurement Model (Factor Definition)
        Healthcare_Spending =~ Private_health_expenditure + Health_expenditure
        SocioEconomicHealth =~ HIV_prevalence + Unemployment

        #structural model
        Life_expectancy ~ SocioEconomicHealth + Healthcare_Spending + CO2_emissions + Electricity + Gov
        Political_stability + Political_stability + Freshwater_resources + Out_of_pocket +  Govt_effect

'
MOD <- sem(EQN, data = scale(life_expectancy), sample.nobs = 217)
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: some estimated ov
## variances are negative
```

```r
MOD
```

```
## lavaan 0.6.16 ended normally after 109 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        21
##
##                                                   Used       Total
##   Number of observations                            81         217
##
## Model Test User Model:
##
##   Test statistic                               172.425
##   Degrees of freedom                                39
##   P-value (Chi-square)                           0.000
```

```r
parameterEstimates(MOD)[1:4,]
```

PRACTICAL SIGNIFICANCE

```
##                   lhs op                        rhs  est    se     z pvalue
## 1 Healthcare_Spending =~ Private_health_expenditure 1.00 0.000    NA     NA
## 2 Healthcare_Spending =~         Health_expenditure 0.14 0.321 0.436  0.663
## 3 SocioEconomicHealth =~             HIV_prevalence 1.00 0.000    NA     NA
## 4 SocioEconomicHealth =~               Unemployment 0.26 0.143 1.815  0.070
##   ci.lower ci.upper
## 1    1.000     1.00
## 2   -0.490     0.77
## 3    1.000     1.00
## 4   -0.021     0.54
```

The `Healthcare_Spending` factor had a stronger effect on `Private_health_expenditure` than `Health_expenditure` because of its regression coefficient of `1.00` compared to `0.14`.

The `SocioEconomicHealth` factor had a stronger effect on `Hiv_prevalence` than `Unemployment` because of its regression coefficient of `1.00` compared to `0.26`.

```r
parameterEstimates(MOD)[5:15,c("lhs","op", "rhs", "est", "z")]
```

STATISTICAL SIGNIFICANCE

```
##                lhs op                      rhs    est      z
## 5  Life_expectancy  ~      SocioEconomicHealth -0.138 -1.896
## 6  Life_expectancy  ~      Healthcare_Spending  0.012  0.375
## 7  Life_expectancy  ~            CO2_emissions -0.021 -0.436
## 8  Life_expectancy  ~              Electricity  0.609  8.383
## 9  Life_expectancy  ~ Govt_health_expenditure  0.169  2.630
## 10 Life_expectancy  ~              Tobacco_use  0.087  1.967
## 11 Life_expectancy  ~      Alcohol_consumption -0.107 -2.084
## 12 Life_expectancy  ~      Political_stability  0.232  3.063
## 13 Life_expectancy  ~     Freshwater_resources -0.015 -0.542
## 14 Life_expectancy  ~            Out_of_pocket -0.203 -3.672
## 15 Life_expectancy  ~       Govt_effectiveness  0.104  1.084
```

`Electricity` is the most significant predictor of `Life_expectancy` as it has the z-score with the highest magnitude (`8.383`). It is statistically significant because the p-value (`0.000`) is below the common alpha level of `0.05`. There is positive relationship, such that for a 1-unit increase in a country's `Electricity` we would expect a `0.609` unit increase in `Life_expectancy` according to our model.

`Out_of_pocket` is the second-most significant predictor of `Life_expectancy` as it has the z-score with the second highest magnitude (`3.672`). It is statistically significant because the p-value (`0.000`) is below the common alpha level of `0.05`. There is a negative relationship, such that for a 1-unit increase in in a country's `Out_of_pocket` we would expect a `-0.203` unit decrease in `Life_expectancy` according to our model.

`Political_stability` is the third-most significant predictor of `Life_expectancy` as it has the z-score with the third highest magnitude (`3.063`). It is statistically significant because the p-value (`0.002`) is below the common alpha level of `0.05`. There is a positive relationship, such that for a 1-unit increase in in a country's `Political_stability` we would expect a `0.232` unit positive in `Life_expectancy` according to our model.

`Govt_health_expenditure` is the fourth-most significant predictor of `Life_expectancy` as it has the z-score with the fourth highest magnitude (`2.630`). It is statistically significant because the p-value (`0.009`) is below the common alpha level of `0.05`. There is a positive relationship, such that for a 1-unit increase inin a country's `Govt_health_expenditure` we would expect a `0.169` unit increase in `Life_expectancy` according to our model.

`Alcohol_consumption` is the fifth-most significant predictor of `Life_expectancy` as it has the z-score with the fifth highest magnitude (`2.084`). It is statistically significant because the p-value (`0.037`) is below the common alpha level of `0.05`. There is a negative relationship, such that for a 1-unit increase in in a country's `Alcohol_consumption` we would expect a `0.107` unit decrease in `Life_expectancy` according to our model.

`Tabacoo_use` is the sixth-most significant predictor of `Life_expectancy` as it has the z-score with the sixth highest magnitude (`1.967`). It is statistically significant because the p-value (`0.049`) is below the common alpha level of `0.05`. There is a positive relationship, such that for a 1-unit increase in in a country's `Tabacoo_use` we would expect a `0.087` unit increase in `Life_expectancy` according to our model.

```r
indicies <- fitMeasures(MOD)
indicies
```

Model Fit

```
##                  npar                  fmin                 chisq 
##                21.000                 1.064               172.425 
##                    df                pvalue        baseline.chisq 
##                39.000                 0.000               412.206 
##           baseline.df       baseline.pvalue                   cfi 
##                55.000                 0.000                 0.626 
##                   tli                  nnfi                   rfi 
##                 0.473                 0.473                 0.410 
##                   nfi                  pnfi                   ifi 
##                 0.582                 0.412                 0.642 
##                   rni                  logl     unrestricted.logl 
##                 0.626              -441.813              -355.601 
##                   aic                   bic                ntotal 
##               925.626               975.910                81.000 
##                  bic2                 rmsea        rmsea.ci.lower 
##               909.683                 0.206                 0.175 
##        rmsea.ci.upper        rmsea.ci.level          rmsea.pvalue 
##                 0.237                 0.900                 0.000 
##        rmsea.close.h0 rmsea.notclose.pvalue     rmsea.notclose.h0 
##                 0.050                 1.000                 0.080 
##                   rmr            rmr_nomean                  srmr 
##                 0.191                 0.191                 0.198 
##          srmr_bentler   srmr_bentler_nomean                  crmr 
##                 0.198                 0.198                 0.211 
##           crmr_nomean            srmr_mplus     srmr_mplus_nomean 
##                 0.211                 0.197                 0.197 
##                 cn_05                 cn_01                   gfi 
##                26.636                30.327                 0.899 
##                  agfi                  pgfi                   mfi 
##                 0.729                 0.334                 0.439 
##                  ecvi 
##                 2.647
```

Comparative Fit Indicies

```r
indicies['nfi']
```

```
##       nfi 
## 0.5817025
```

```r
indicies['nnfi']
```

```
##      nnfi 
## 0.4732364
```

```r
indicies['ifi']
```

```
##       ifi
## 0.6424904
```

```r
indicies['cfi']
```

```
##       cfi
## 0.6264768
```

Our Normed Fit Index lies at `0.582`, our Non-normed Fit Index lies at `0.473`, our Incremental Fit Index lies at `0.642`, and our Comparative Fit Index lies at `0.626`. These our below the common threshold (`0.90-0.95`) that is considered a good fit, so our model does not provide an adequate fit when compared to a fully independent model according to these comparative fit indices.

```r
indicies['rmsea']
```

```
##     rmsea
## 0.2055149
```

Our Root Mean Square Error of Approximation lies at `0.206` which is above the common threshold of `0.08`. This means that our model does not provide an adequate fit to the data when compared to a fully saturated model according to this index of fit.

Absolute Fit Indicies

```r
indicies['mfi']
```

```
##      mfi
## 0.438845
```

Our McDonald & Marsh Fit Index lies at `0.439` which is far below the common threshold of `0.90-0.95`. This means our model does not provide an adequate fit to the original covariance structure.

```r
indicies['gfi']
```

```
##       gfi
## 0.8994171
```

```r
indicies['agfi']
```

```
##      agfi
## 0.7291999
```

Our goodness-of-fit index lies at `0.899` and our adjusted goodness-of-fit lies at `0.729`. This means that about `72.9%` of the covariability in the original data matrix can be explained by the structural equation model.