R Code for Validating the MPI

The MPI is primarily a normative index. As the MPI handbook argues,

"Statistical approaches are relevant for multidimensional poverty measures, but Chapter 6 argued, value judgements also constitute a fundamental prior element. Thus, information on relationships between indicators is used to improve rather than determine measurement design. For example, if indicators are very highly associated in a particular dataset, that is not sufficient grounds to mechanically drop either indicator; both may be retained for other reasons-for example if the sequence of their reduction over time differs, or if both are important in policy terms. So the normative decision may be to retain both indicators, with or without adjustments to their weights, but the analysis of redundancy will have clarified their justification and treatment." (Ch 7, p. 16)

Nonetheless, there are some quantitative checks that can be useful in assessing the index.

The handbook suggests using Cramer's V to measure correlation, as well as their own measure of redundancy.Both consist of similar terms that are based on a two-by-two contingency table. For indicators j and j', we can then define the following terms:

Let $\mathbb{p}_{11}^{jj'}$ be the percentage of the population experiencing both deprivations.

Let $\mathbb{p}_{10}^{jj'}$ be the percentage of the population deprived in j but not j'.

Let $\mathbb{p}_{01}^{jj'}$ be the percentage of the population deprived in j' but not j.

Let $\mathbb{p}_{00}^{jj'}$ be the percentage of the population deprived in neither.

The percentages deprived in each indicator are written as:

Let $\mathbb{p}_{+1}^{j'}$ be the percentage of the population deprived in j'.

Let $\mathbb{p}_{+0}^{j'}$ be the percentage of the population not deprived in j'.

Let $\mathbb{p}_{1+}^{j}$ be the percentage of the population deprived in j.

Let $\mathbb{p}_{0+}^{j}$ be the percentage of the population not deprived in j.

Cramer's V can then be written:

$$Cramer's\ V = \frac{(\mathbb{p}_{11}^{jj'} \times \mathbb{p}_{00}^{jj'}) - (\mathbb{p}_{10}^{jj'} \times \mathbb{p}_{00}^{jj'})}{\sqrt{\mathbb{p}_{+1}^{j'} \times \mathbb{p}_{+0}^{j'} \times \mathbb{p}_{1+}^{j} \times \mathbb{p}_{0+}^{j}}}$$

The redundancy measure is as follows:

$$R^o = \frac{\mathbb{p}_{11}^{jj'}}{\min(\mathbb{p}_{+1}^{j'}, \mathbb{p}_{1+}^{j})}, 0 \leq R^o \leq 1$$

Loading in required libraries and data

```
library(purrr)
library(feather)
library(tidyverse)
library(gtools)
library(stargazer)

mpi_data_15 = read_feather("C:/Users/natek/Documents/mpi_data_15.feather")
```

Applying Cramer's V and the Redundancy measure

```
combos <- c("income_poverty", "fam_emp", "health", "education", "overcrowd", "computer_internet",
            "hcost_dep", "lang_dep", "disability")

c_matrix <- combinations(n=length(combos), r=2, v=combos) %>% as_data_frame()

cramers_v <- function(x_name, y_name){
  x <- mpi_data_15[[x_name]]
  y <- mpi_data_15[[y_name]]
  t1 <- table(x,y)
  n <- nrow(mpi_data_15)
  numerator <- ((t1[1,1]/n)*(t1[2,2]/n)) - ((t1[1,2]/n)*(t1[2,1]/n))
  denominator <- sqrt(mean(x)*mean(y)*(1-mean(x)*(1-mean(y))))
  cv <- numerator/denominator
  cv
}

calc_redundancy <- function(x_name, y_name){
  x <- mpi_data_15[[x_name]]
  y <- mpi_data_15[[y_name]]
  t1 = table(x,y)
  redundancy = (t1[2,2]/nrow(mpi_data_15))/
    (min(mean(x, na.rm = T),
         mean(y, na.rm = T),
         mean(x[mpi_data_15$mpi_poor == 1], na.rm =T),
         mean(y[mpi_data_15$mpi_poor == 1], na.rm =T)
    ))
  redundancy
}

chi_sq_p <- function(x_name,y_name){
  x <- mpi_data_15[[x_name]]
  y <- mpi_data_15[[y_name]]
  temp <- chisq.test(x, y)
  temp$p.value
}

c_matrix$cramers_V <- map2(c_matrix$V1, c_matrix$V2, cramers_v)
c_matrix$redundancy_coefficient <- map2(c_matrix$V1, c_matrix$V2, calc_redundancy)
c_matrix$p_value <- map2(c_matrix$V1, c_matrix$V2, chi_sq_p)
```

Reformatting the data to make tables

```
c_matrix$cramers_V <- round(as.numeric(c_matrix$cramers_V), 2)
c_matrix$red <- round(as.numeric(c_matrix$redundancy_coefficient), 2)
c_matrix$p <- round(as.numeric(c_matrix$p_value), 8)


col1 <- c("-", c_matrix$cramers_V[1:8])
col2 <- c("", "-", c_matrix$cramers_V[9:15])
col3 <- c("", "","-", c_matrix$cramers_V[16:21])
col4 <- c("", "","", "-", c_matrix$cramers_V[22:26])
col5 <- c("", "", "", "", "-", c_matrix$cramers_V[27:30])
col6 <- c("", "", "", "", "", "-", c_matrix$cramers_V[31:33])
```

```r
col7 <- c("", "", "", "", "", "", "-", c_matrix$cramers_V[34:35])
col8 <- c("", "", "", "", "", "", "", "-", c_matrix$cramers_V[36])
col9 <- c("", "", "", "", "", "", "",  "", "-")


v_table <- cbind(col1, col2, col3, col4, col5, col6, col7, col8, col9)
colnames(v_table) <- c("Internet", "Disability", "Education", "Employment", "Housing", "Health",
                       "Income", "Language", "Overcrowding")
rownames(v_table) <- c("Internet", "Disability", "Education", "Employment", "Housing", "Health",
                       "Income", "Language", "Overcrowding")

col1 <- c("-", c_matrix$red[1:8])
col2 <- c("", "-", c_matrix$red[9:15])
col3 <- c("", "","-", c_matrix$red[16:21])
col4 <- c("", "","", "-", c_matrix$red[22:26])
col5 <- c("", "", "", "", "-", c_matrix$red[27:30])
col6 <- c("", "", "", "", "", "-", c_matrix$red[31:33])
col7 <- c("", "", "", "", "", "", "-", c_matrix$red[34:35])
col8 <- c("", "", "", "", "", "", "", "-", c_matrix$red[36])
col9 <- c("", "", "", "", "", "", "",  "", "-")


r_table <- cbind(col1, col2, col3, col4, col5, col6, col7, col8, col9)
colnames(r_table) <- c("Internet", "Disability", "Education", "Employment", "Housing", "Health",
                       "Income", "Language", "Overcrowding")
rownames(r_table) <- c("Internet", "Disability", "Education", "Employment", "Housing", "Health",
                       "Income", "Language", "Overcrowding")

col1 <- c("-", c_matrix$p[1:8])
col2 <- c("", "-", c_matrix$p[9:15])
col3 <- c("", "","-", c_matrix$p[16:21])
col4 <- c("", "","", "-", c_matrix$p[22:26])
col5 <- c("", "", "", "", "-", c_matrix$p[27:30])
col6 <- c("", "", "", "", "", "-", c_matrix$p[31:33])
col7 <- c("", "", "", "", "", "", "-", c_matrix$p[34:35])
col8 <- c("", "", "", "", "", "", "", "-", c_matrix$p[36])
col9 <- c("", "", "", "", "", "", "",  "", "-")


p_table <- cbind(col1, col2, col3, col4, col5, col6, col7, col8, col9)
colnames(p_table) <- c("Internet", "Disability", "Education", "Employment", "Housing", "Health",
                       "Income", "Language", "Overcrowding")
rownames(p_table) <- c("Internet", "Disability", "Education", "Employment", "Housing", "Health",
                       "Income", "Language", "Overcrowding")
```

Cramer's V

```r
print(v_table, quote = "FALSE")
```

```
##              Internet Disability Education Employment Housing Health
## Internet     -
## Disability   0.1      -
## Education    0.18     0.14       -
## Employment   0.15     0.22       0.14      -
```

```
## Housing      0.1      0.02        0.06       0.1        -
## Health       0.07     -0.04       0.14       0.06       0.08    -
## Income       0.1      0.12        0.12       0.34       0.42    0.13
## Language     0.11     0           0.14       0.01       0.09    0.1
## Overcrowding -0.06    0.16        0.09       0.39       -0.07   0.1
##              Income Language Overcrowding
## Internet
## Disability
## Education
## Employment
## Housing
## Health
## Income       -
## Language     0.07   -
## Overcrowding 0.42   -0.01    -
```

Redundancy

```r
print(r_table, quote = "FALSE")
```

```
##              Internet Disability Education Employment Housing Health
## Internet     -
## Disability   0.27     -
## Education    0.33     0.27        -
## Employment   0.29     0.39        0.25       -
## Housing      0.26     0.19        0.22       0.26       -
## Health       0.22     0.04        0.24       0.17       0.26    -
## Income       0.28     0.35        0.31       0.57       0.57    0.35
## Language     0.34     0.06        0.32       0.13       0.32    0.23
## Overcrowding 0.05     0.22        0.2        0.6        0.05    0.19
##              Income Language Overcrowding
## Internet
## Disability
## Education
## Employment
## Housing
## Health
## Income       -
## Language     0.3    -
## Overcrowding 0.81   0.05     -
```

Chi Square

```r
print(p_table, quote = "FALSE")
```

```
##              Internet Disability Education Employment Housing Health
## Internet     -
## Disability   0        -
## Education    0        0           -
## Employment   0        0           0          -
## Housing      0        0           0          0          -
## Health       0        0           0          0          0       -
## Income       0        0           0          0          0       0
## Language     0        0.00022257 0          0          0       0
## Overcrowding 0        0           0          0          0       0
##              Income Language Overcrowding
```

```
## Internet
## Disability
## Education
## Employment
## Housing
## Health
## Income        -
## Language     0      -
## Overcrowding 0      0        -
```

Bootstrapping the MPI

```r
df_boot <- mpi_data_15 %>% select(mpi_poor, PERWT)

df_length <- nrow(mpi_data_15)

nboot <- 1000
boot_mean <- vector(length = nboot)

for (i in 1:nboot) {
  df_boot[sample(nrow(df_boot), size = df_length, replace = TRUE), ]
  boot_mean[i] = sum(df_boot$mpi_poor*df_boot$PERWT, na.rm = TRUE)/sum(df_boot$PERWT)
}


bootstrap_mean <- mean(boot_mean)
ci <- quantile(boot_mean, c(.025, .975))
```