Example analysis

Peter Konings & Niko Speybroeck

April 13, 2012

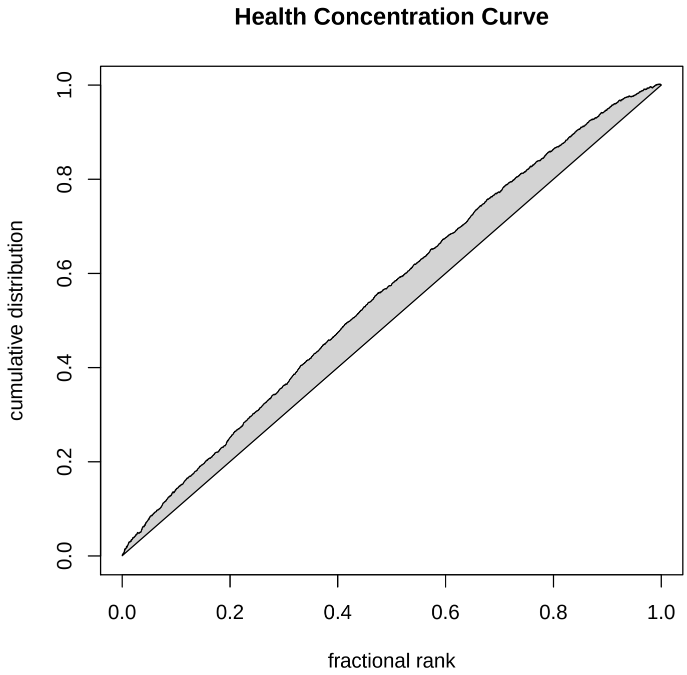
1 Relative Concentration Index

We use a Demographic Health Survey dataset from Kenya (2003) for illustrating this new R-package. Start R, and load the decomp package (which includes the kenya dataset). The function RCI calculates the concentration index for stunting and its confidence interval.

```
> library(decomp)
> data(kenya)
> kenya.stunting.CI <- RCI(kenya$wealth, kenya$stunting, wt = kenya$popweight)
> summary(kenya.stunting.CI)

Call:
RCI(x = kenya$wealth, y = kenya$stunting, wt = kenya$popweight)

Health Concentration Index:
 -0.1153306

Variance:
 0.0001241775

95% Confidence Interval:
 -0.1371715 -0.09348977
```

The next command will return a concentration plot.

```
> plot(kenya.stunting.CI)
```



Since the stunting variable contains both negative and positive values, the calculated RCI is not bounded between $[-1, 1]$. This can be corrected in a number of ways, using the function `correctSign`. One way is to impute a value of 0 for all negative values.

```
> kenya$stunting.corrected <- correctedValue(correctSign(kenya$stunting, shift=FALSE))
> kenya$stunting.corrected <- correctedValue(correctSign(kenya$stunting, shift=FALSE))
> kenya.correctedstunting.CI <- RCI(kenya$wealth, kenya$stunting.corrected, wt = kenya$pop
> summary(kenya.correctedstunting.CI)

Call:
RCI(x = kenya$wealth, y = kenya$stunting.corrected, wt = kenya$popweight)

Health Concentration Index:
 -0.08680988

Variance:
 5.265099e-05

95% Confidence Interval:
 -0.1010316 -0.07258818
```

2 Decomposition

We now fit a linear model for stunting with sex, place of residence (urban/rural) and mother's education as covariates. Other generalized linear models (*e.g.*, a logistic regression) are allowed.

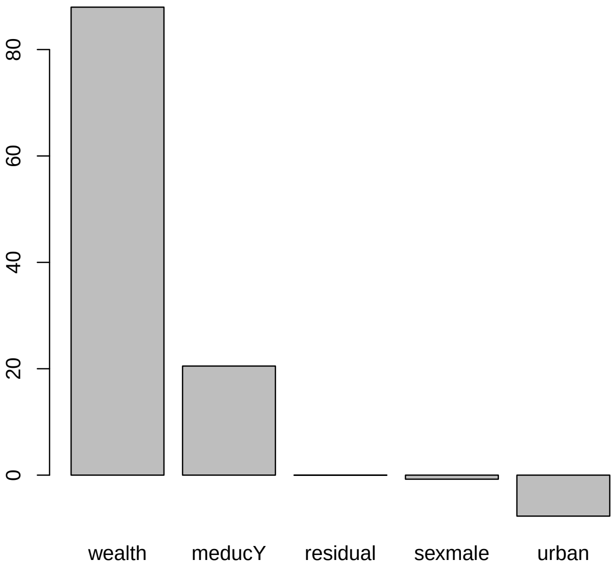
```
> kenya.stunting.lm <- lm(stunting ~ meducY + sex + wealth + urban, weights = popweight, d
```

Next, we calculate the overall RCI and the contribution of the individual covariates with the `contribution` command:

```
> contribution(kenya.stunting.lm, kenya$wealth)

Overall CI: -0.02054074
(based on a corrected value)

Decomposition:
         Contribution (%) Corrected
residual     -0.005950229        no
meducY       20.502829983        no
sexmale      -0.764719310        no
wealth       87.971215576        no
urban        -7.703376020        no
```

It is possible to display a barplot of the relative contributions as well:

```
> plot(contribution(kenya.stunting.lm, kenya$wealth))
```

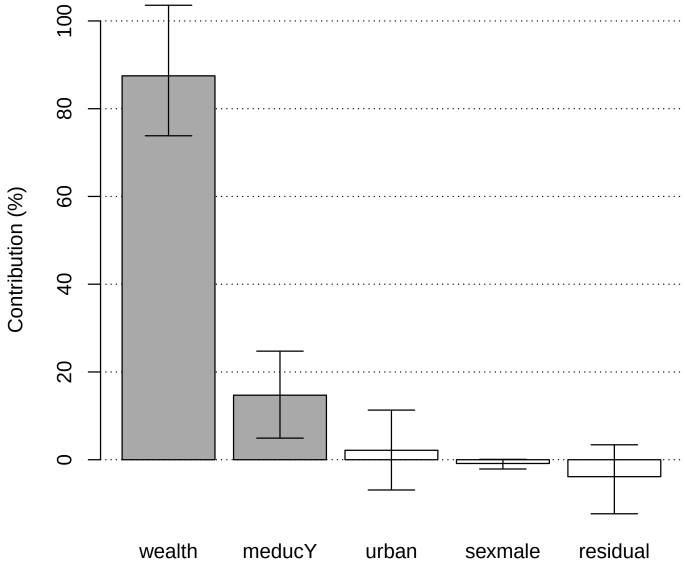


An analytical solution for the sampling variability and the standard errors for the contributions in the decomposition analysis is not available. However, statistical inference on the decomposition results can be obtained through bootstrapping. The following code shows how to conduct a bootstrap assuming a random sample, ingnoring survey design characteristics. For more advanced complex multi-stage sampling designs, the bootstrap needs to be adjusted, as was done by van Doorslaer. The example is therefore purely illustrative.

```
> # set a random seed to make the analysis reproducible
> set.seed(12345)
> model1 <- lm(stunting ~ meducY + sex + wealth + urban, data = kenya)
> nobs <- length(model1$fitted.values)
> B <- 1000
> results <- matrix(nrow = B, ncol = length(model1$coefficients))
> for (b in 1:B)
+ {
+          bootstrapsample <- sample(nobs, nobs, replace = TRUE)
+          bootstrapdata <- model1$model[bootstrapsample,]
+          row.names(bootstrapdata) <- 1:nrow(bootstrapdata)
+          bootstrap.lm <- lm(stunting ~ meducY + sex + wealth + urban, data = bootstrapda
+          results[b,] <- contribution(bootstrap.lm, bootstrap.lm$model$wealth)$contributi
+ }
> results <- data.frame(results)
> names(results) <- c('residual', names(model1$coefficients)[-1])
> resultstable <- apply(results, 2 , quantile, probs = c(0.05, 0.5, 0.95))
> # resultstable contains the contribution and associated confidence intervals
> resultstable

      residual    meducY     sexmale    wealth     urban
5%  -12.321109  4.924441 -2.11495419  73.83186 -6.895522
50%  -3.860404 14.707484 -0.85615194  87.49508  2.150305
95%   3.401915 24.742063  0.07833726 103.57449 11.309509
```

We can plot these results as follows:

```
> library(gplots)
> resultstable <- resultstable[,order(resultstable[2,], decreasing = TRUE)]
> significant <- (resultstable[1,] / resultstable[3,]) > 0
> color <- ifelse(significant, 'dark gray', 'white')
> barplot2(resultstable[2,], col = color,       plot.grid = TRUE, plot.ci = TRUE,
+                 ci.l = resultstable[1,], ci.u = resultstable[3,], ylab = 'Contribution (
```



Acknowledgments

This paper resulted from cooperation initiated through meetings organized by the WHO European Office for Investment for Health and Development, Venice, Italy.