Calibration

July 10, 2009

This is an example of 'calib' function using calibration and adjustment for nonresponse (with response homogeneity groups).

Creates the population data frame (4 variables, 'state', 'region', 'income' and 'sex'; 'state' has 2 categories 'nc' and 'sc'; 'region' has 3 categories 1,2,3; 'income' and 'sex' are randomnly generated):

```
> data = rbind(matrix(rep("nc", 165), 165, 1, byrow = TRUE),
+     matrix(rep("sc", 70), 70, 1, byrow = TRUE))
> data = cbind.data.frame(data, c(rep(1, 100), rep(2,
+     50), rep(3, 15), rep(1, 30), rep(2, 40)),
+     1000 * runif(235))
> sex = runif(nrow(data))
> for (i in 1:length(sex)) if (sex[i] < 0.3) sex[i] = 1 else sex[i] = 2
> data = cbind.data.frame(data, sex)
> names(data) = c("state", "region", "income", "sex")
```

Computes the population stratum sizes:

```
> table(data$state)
```

Not run:

nc sc

165 70

We select a stratified sample. The 'state' variable is used as a stratification variable. The sample stratum sizes are 25 and 10, respectively. The method is 'srswor' (equal probability, without replacement).

```
> s = strata(data, c("state"), size = c(25, 10),
+     method = "srswor")
```

Obtains the observed data:

```
> s = getdata(data, s)
```

The 'status' variable is used in the 'rhg_strata' function. Adds the 'status' column to s (1 - sample respondent, 0 otherwise); it is randomnly generated:

```
> status = runif(nrow(s))
> for (i in 1:length(status)) if (status[i] < 0.3) status[i] = 0 else status[i] = 1
> s = cbind.data.frame(s, status)
```

Computes the response homeogeneity groups using the 'region' variable:

```
> s = rhg_strata(s, selection = "region")
```

Selects only the sample respondents:

```
> sr = s[s$status == 1, ]
```

Creates the population data frame of sex and region indicators:

```
> X = matrix(0, nrow = nrow(data), ncol = 5)
> for (i in 1:nrow(data)) {
+     if (data$sex[i] == 1)
+         X[i, 1] = 1
+     if (data$sex[i] == 2)
+         X[i, 2] = 1
+     if (data$region[i] == 1)
+         X[i, 3] = 1
+     if (data$region[i] == 2)
+         X[i, 4] = 1
+     if (data$region[i] == 3)
+         X[i, 5] = 1
+ }
```

Computes the population totals for each sex and region:

```
> total = c(t(rep(1, nrow(data))) %*% X)
```

Creates the sample data frame of sex and region indicators:

```
> Xs = matrix(0, nrow = nrow(sr), ncol = 5)
> for (i in 1:nrow(sr)) {
+     if (sr$sex[i] == 1)
+         Xs[i, 1] = 1
+     if (sr$sex[i] == 2)
+         Xs[i, 2] = 1
+     if (sr$region[i] == 1)
+         Xs[i, 3] = 1
```

```
+     if (sr$region[i] == 2)
+         Xs[i, 4] = 1
+     if (sr$region[i] == 3)
+         Xs[i, 5] = 1
+ }
```

Computes the initial weights using the inclusion and response probabilities:

```
> d = 1/(sr$Prob * sr$prob_resp)
```

Computes the g-weights:

```
> g = calib(Xs, d, total, method = "linear")
```

Checks the calibration:

```
> checkcalibration(Xs, d, total, g)
```