Assessment instructions.

Please read all instructions carefully before commencing the analysis.

General guidance

- Create an R-script in your GitHub branch and name it after yourself e.g. ***Rie.R***
- Include the answers to questions as comments in your R script.
- For all numeric answers include a point estimate and 95% confidence intervals.
- This mid assessment is not designed to produce a pass/fail result but rather get a general sense of respondents' ability to work with household consumption-expenditure survey data working in the R environment.
- The data files to be used for the assessment are in CSV (.csv) format.

Background

We are interested in introducing fortified wheat flour in a country. We have national household consumption & expenditure survey data from the country, which includes information on the weight, value and frequency of foods consumed over the previous 7 days.

Questions

We would like to use this data to help us assess the suitability of wheat flour for fortification in this country context. To do this, we would like to know over the last 7 days, on average:

1. What proportion of households consumed wheat flour?

2. What was the monetary value and weight (in grams) of wheat flour consumed by households?

3. How many days did households consume wheat flour?

4. What proportion of household expenditure on wheat flour was accounted for by purchases from markets and what proportion was accounted for by own production?

5. Provide a box plot of the price paid for 100 grams of wheat flour when purchased from the market at 1) urban level, and 2) rural level?

Notes on the data to use for the assessment:

Data to be used for the assessment is stored in 2 datasets:

- Data set A is in a long format and lists the expenditure in quantity (in weight), measurement unit (in weight), monetary value (Rupee) on each food item, both for food consumed from own production and purchased at the market, as well as the number of days each food was consumed per week, alongside other variables.
- Data set B contains information on the sample weights, demographic profile and geographic location of households.

To answer the above questions, you will need to:

- Set the data up for analysis of survey data, noting that the survey is based on a stratified two-stage cluster design, where: First stage: Enumeration Areas (EA) are drawn from Census

files Second stage: in each EA selected, a primary sampling unit is chosen. The survey is stratified by urban and rural households. A survey sampling weight is provided.

- Merge the two data sets (noting one dataset is in long format and one is in wide format) and create variables necessary for analysis.
- For weight conversion: 1 kg of foods= 1,000 grams. 1 litre of all liquids=1,000 grams.
- All currency is in Rupees.

Codebook

Data File A

psu	Primary Sampling Unit
hhno	Household Number
foodcode	Numeric food code
foodname	Food name
home_q	Quantity of food consumed from own production
home_u	Measurement Unit for quantity of food consumed from own production, 1 = kg, 2 = g, 3 = L, 4 = mL
home_val	Value of food consumed from own production in Rupees
purchase_q	Quantity of food consumed from own the market
purchase_u	Measurement unit for quantity of food consumed from market, 1 = kg, 2 = g, 3 = L, 4 = mL
purchase_val	Value of food consumed from market in Rupees
daysconsumed	Number of days food consumed out of 7

Data File B

psu	Primary Sampling Unit
hhno	Household number
n_member	Number of household members
hhsize	Household size
hh_sample_weight	Household survey sample weight
headname	Name of head of household
hhldsex	Sex of head of household
dist	District code
urbrur	Urban/Rural location of household
vdcmun	Municipality code
tole	Tole/village number
Ward	Ward identifier
ea	Enumeration area
fmlymemb	Family members
memb5	Family members under 5 years
totmemb	Total household members