Teaching Statistics with Real World Data from IPUMS

These exercises cover basic statistical concepts, guiding students through real world examples, using R and census microdata from the IPUMS International database. See other IPUMS resources for instructions and a tutorial about accessing IPUMS data, and for a guide to the IPUMS R package for reading IPUMS data extracts into R.

Exercise: Probability

Topics covered:
- Probability of an event
- Probability of union and intersection of two events
- Disjoint and independent events
- Conditional probability

Required dataset: IPUMS-International

Required variables:
1. COUNTRY
2. YEAR
3. RELATE (relationship to household head)
4. OWNERSHIP (ownership of housing unit)
5. INTERNET (internet access)

[The only preselected variables that are needed in this exercise are COUNTRY and YEAR. Make sure to remove all of the other preselected variables by unchecking the blue boxes next to them. This will reduce the size of your data file and also make it easier to view the data in R.]

RECOMMENDED SAMPLE:
Fiji [2014]

SAMPLE SELECTION INSTRUCTIONS:
[Note that both the variables – OWNERSHIP and INTERNET are household variables. We assume that head of the household represents a household. Hence, we record the OWNERSHIP and INTERNET values for the head of a household.]

1. Limit sample to 10,000 households.

2. In the "extract request" page, which comes after you, click "create data extract", first click on "select cases", choose "RELATE" and submit. Then, select "include only those persons meeting case selection criteria" and "1 Head".
3. Next select "customize sample size" and type 10 in the box under households and Fiji 2014. *[Note that the sample size is in 1000s]*

NOTES:
- Round off all probabilities to the nearest two decimal places.
- Let us define event A as the event of having ownership of housing unit and event B as the event of having internet access.

• SECTION I

1. Describe the sampling method used to collect the customized samples. . [*Hint: Check help in "customize sample sizes"*]

2. What is the universe for the two variables – OWNERSHIP and INETRNET?

3. What does the values INTERNET = 0 and OWNERSHIP = 0 signify? Should these observations be included or excluded in calculation of probability?

4. What does the values INTERNET = 9 and OWNERSHIP = 9 signify? Should these observations be included or excluded in calculation of probability?

• SECTION II

1. What is the probability that a household has ownership of housing unit?

2. What is the probability that a household does not have ownership of housing unit?

3. What is the probability that a household has an internet connection?

4. What is the probability that a household does not have an internet connection?

• SECTION III

1. Are the variables OWNERSHIP and INTERNET disjoint (mutually exclusive)? Why/ why not?

2. Are the variables OWNERSHIP and INTERNET independent? Why/ why not?

3. What is the probability that a random household from Brazil has internet access given that household owns housing unit?

- **SECTION IV**

1. What is the probability that a random household from Brazil does not own housing unit and has internet access?

2. What is the probability that a random household from Brazil owns housing unit and does not have internet access?

3. What is the probability that a random household from Brazil does not own housing unit and does not have internet access?

- **SECTION V**

1. What is the probability that a randomly selected household from Brazil owns housing unit or has internet access?

2. What is the probability that a randomly selected household from Brazil does not own housing unit or internet access?

3. What is the probability that a randomly selected household from Brazil owns housing unit or does not have internet access?

4. What is the probability that a randomly selected household from Brazil does not own housing unit or does not have internet access?