**R Lesson 1 - Introduction to Statistics**

**References**

Black - Chapter 1 Introduction to Statistics (pp. 3-49)  
Verzani - Chapter 2 Univariate Data (pp. 28-29) and Chapter 6 Populations (pp. 219-220)  
Lander - Chapter 4 Basics of R (pp. 35-51) and Chapter 6 Reading Data into R (pp. 73-81)  
Stowell - Chapter 3Preparing and Manipulating Your Data (pp. 44)

**Data set:** [home\_prices.csvView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178709/download?wrap=1)

**Description:** This data file is derived from a random sample of home resale records maintained by realtors. There are 117 observations and eight variables:

* 1. PRICE = Selling price ($hundreds)
  2. SQFT = Square feet of living space
  3. YEAR = Year of construction (year)
  4. BATHS = Number of bathrooms
  5. FEATS = Number out of 11 features (dishwasher, refrigerator, microwave, disposal, washer, intercom, skylight(s), compactor, dryer, handicap fit, cable TV access)
  6. NBR = Located in northeast sector of city (YES) or not (NO)
  7. CORNER = Corner location (YES) or not (NO)
  8. TAX = Annual taxes ($)

**Exercises:**

1. What are the measurement levels of each of the eight variables?
2. Should any variable have its values changed to better reflect its true nature?
3. For the variable PRICE, select a simple random sample of size 12 from the file. Save this sample in a vector named SRS. Print the values in SRS and compute the mean value. (For random sampling, see Lander pp.187-188, Stowell pp. 44 and Verzani pp. 219-220.)
4. For the variable PRICE, select a systematic sample of twelve observations. Start with the seventh observation and pick every 10th observation thereafter (i.e. 7, 17, 27,..). You should end with the 117th observation. Save the sample in a vector named SS. Print the values SRS and compute the mean value. (For systematic sampling, refer to page 283 of Lander for seq(from,to,by=) and Verzani pp. 28-29. Use this to define a vector of indices for selection.)
5. Examine the printed values and mean values obtained from the two sampling procedures. Do you see a difference? (Try the commands summary(SRS) and summary(SS).
6. Create boxplots for SRS and SS using boxplot(). How do the two samples compare?

**Running into Trouble?** Check out these solutions to help guide you along.

* [Lesson\_01\_Solution.pdfPreview the documentView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178786/download?wrap=1)
* [Lesson\_01\_Code\_Solution.rView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178761/download?wrap=1)