University of Toronto
Summer 2014

STA304/1003 H1F:
Surveys, Sampling, and Observational Data

Data Analysis Assignment # 3

Data analysis assignments are for practice only and do NOT need to be handed in.

The file "baseball.csv" has specifications on 797 baseball players from the rosters of all major league teams in November, 2004. (See Assignment 1 for details about variables).

Treat the data in the file as the population of all baseball players in 2004.

1. Use 'R' to take a random sample of 50 ordered pairs (*G*, *salary*). ie. 50 players' ordered pairs of games played and salary. Copy and paste your sample data.

   Under your sample data, clearly indicate your answers, including output where necessary:

   a) Estimate the ratio of the mean number of games played to mean salary.

   b) Calculate the standard error of your estimate from part a).

   c) Create a 95% CI for the population ratio of interest.

   d) Find the population ratio. Does your CI include the parameter?

   e) Comment on the validity of using Ratio Estimation in this case by including appropriate plots, statistics, etc. Would an SRS be better?

2. Use 'R' to take a one-stage cluster sample of *salary:* Use *POS* (primary position played) as clusters and take an SRS of 5 clusters. Copy and paste your sample data.

   Under your sample data, clearly indicate your answers, including output where necessary:

   a) Estimate the population mean salary using your sample data.

   b) Calculate the standard error of your estimate in part a).

   c) Comment on the appropriateness of using a cluster sample here. Would it be better to use Stratified Sampling or Cluster Sampling in this case? Include the relevant statistics, calculations, ANOVA tables, etc. to justify your answer.