STAT 3014/3914

Semester 2	Applied Statistics	2015

Computer Exercise 12 for Sample Survey

Practice Problems

1. Open the data set `fuel.csv`. Set `Fuel` to be the y variable and `Weight` to be the x variable. Calculate the population means of x and y and their correlation coefficient ρ. Comment on the use of x as the auxiliary variable to improve the estimation of the mean of y.

```
data=read.csv("http://www.maths.usyd.edu.au/u/UG/IM/STAT2012/r/fuel.csv")
attach(data)
data
```

2. Draw a simple random sample (SRS) of (x, y) of size $n = 12$. To draw the order pair (x, y), one can first draw the labels i from $1, \ldots, n$ and then identify the pair (x_i, y_i). Use `set.seed(12345)` *once* to set `12345` as the seed for drawing subsequent samples.

```
set.seed(12345)
isam=sample(1:N,n,replace=FALSE,prob=NULL) #draw SRS of labels
isam
ysam=y[isam]
xsam=x[isam]
rbind(xsam,ysam)
```

   (a) Estimate the population mean and its standard error estimate from the SRS using the *ordinary* estimator.

   (b) Estimate the population mean and its standard error estimate from the SRS using the *ratio* estimator.

   (c) Estimate the population mean and its standard error estimate from the SRS using the *regression* estimator and compare the three estimators. Plot the sample of (x, y) and the fitted line. Compare this estimator with the estimators in (a) and (b) and comment.

   (d) Estimate the population mean using the *Hartley Ross* estimator.

3. (Advanced must; Normal optional) Stratify the population into two strata according to the `Type` of car because `Fuel` and `Weight` are more internally homogenous across cars with similar `Type`. The first subpopulation contains all small cars with `Type=Small` and `Sporty`. From each subpopulation, draw a SRS of (x_i, y_i) of size $n_i = 8$.

(a) Estimate the population mean and its standard error estimate using the *separate* ratio estimator. Report the separate ratios r_l and the vectors of $z_{sr,li} = y_{li} - r_l x_{li},\ i = 1, 2$. Compare this estimator with the ratio estimator in 2(b) based on just one SRS and comment.

(b) Estimate the population mean and its standard error estimate using the *combined* ratio estimator. Report the combined ratio r_c and the vectors of $z_{cr,li} = y_{li} - r_c x_{li},\ i = 1, 2$. Compare this estimator with the separate ratio estimator in 3(a) and comment. Which estimator, the separate or combined ratio estimator, do you prefer? Why?