STAT 5201 HW6

6-1. Edition 2: Chapter6-Problem 4 (Edition 1: Chapter6-Problem 6)

Note: The handout "Studying the Horvitz-Thompson Estimator" will help you do the following problem.

6-2. Consider a population where x is a random sample from some distribution and y is generated from x using the following relationship:

$$y_i = \beta x_i + e_i$$

where the e_i's are independent $N(0, v(x_i)^2)$, $v(\cdot)$ is some function. If $v(x_i) = \lambda x_i$ for some constant λ then the ratio estimator will be a good estimator of

$$Y = \sum_{i=1}^{N} y_i,$$

the population total.

For pps sampling based on x we consider the "approximate" HT estimator:

$$\delta_{HT}(y_{smp}) = \sum_{i \in smp} w_i y_i, \text{where } \; w_i = \frac{\sum_{j=1}^{N} x_j}{n x_i}$$

with its estimate of variance

$$\frac{1}{n} \sum_{i \in smp} (n w_i y_i - \delta_{HT}(y_{smp}))^2/(n-1)$$

In the above mentioned handout the function

```
compar3estlp(popy,popx,design,n,R)
```

with

```
design<-popx
```

will take R samples of size n(using pps sampling based on popx) and find the usual estimate of the population total $N \times \bar{y}_{smp}$, the ratio estimator and the approximate" HT estimator. For each estimator it returns its average value, average absolute error, the average lower bound and length of the nominal 95% confidence interval and the proportions of intervals which contained the true population total of y. If you set

```
design<-rep(1,length(popy))
```

you will just be doing simple random sampling without replacement.

(a) For the example in the handout with

```
design<-popx
```

We see that both the ratio estimator and HT estimator do much better than the usual estimator. Explain.

(b) For the example in the handout with

```
design<-rep(1,length(popy))
```

We see that the usual estimator and the HT estimator behave about the same. Explain.

(c) For the example in the handout and other choices of the design, I found situations where the average absolute error of the HT estimator was 4 times as large as the average absolute error of the ratio estimator. Find a design for which this happens.

(d) The correlation between *popy* and *popx* for the example in the handout is 0.92. Construct two new populations, one with correlation around 0.60 and the other with correlation around 0.30, where the ratio estimator and the HT estimator behave similarly when the design is *popx* but the ratio estimator outperforms the HT estimator for another design.

In the handout *popx* was generated as a random sample from a log-normal distribution. The random variable U is said to have a log-normal if $\log U$ is normal distributed. That is U is of the form $exp(V)$ where V is normally distributed. So the R command

$$rlnorm(500, \mu, \sigma)$$

gives a random sample of size 500 from a distribution where μ and σ are the mean and standard deviation of the logarithm. The mean and variance are:

$$E(U) = exp(\mu + \sigma^2/2);\ Var(U) = exp(2\mu + \sigma^2)(exp(\sigma^2) - 1)$$