STAT3003 Problem Sheet 1

ZHENG Weijia (William, 1155124322)

Spring, 2021

1 Q1

We want to prove $E[\hat{\sigma}^2] = \frac{N}{N-1}\sigma^2$.
Note that $\sigma^2 = \frac{1}{N}\sum_{j=1}^{N}(u_j - \mu) = \frac{1}{N^2}[(N-1)\sum_{j=1}^{N} u_j^2 - \sum_{j=1}^{N}\sum_{k\neq j} u_j u_k]$
Hence R.H.S.=$\frac{N}{N-1}\sigma^2 = \frac{1}{N(N-1)}[\sum_{j=1}^{N}\sum_{k\neq j} u_j^2 - \sum_{j=1}^{N}\sum_{k\neq j} u_j u_k]$.
Also note that $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{1}{n(n-1)}[(n-1)\sum_{i=1}^{n} Y_i^2 - \sum_{j=1}^{n}\sum_{k\neq j} Y_j Y_k]$
$= \frac{1}{n(n-1)}[(n-1)\sum_{i=1}^{n} u_i^2 Z_i^2 - \sum_{j=1}^{n}\sum_{k\neq j} u_j u_k Z_j Z_k]$.
Hence L.H.S.=$E[\hat{\sigma}^2] = \frac{1}{n}\sum_{i=1}^{n} u_i^2 E[Z_i^2] - \frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{k\neq j} u_j u_k E[Z_j Z_k]$
$= \frac{1}{n}\sum_{i=1}^{n} u_i^2 E[Z_i^2] - \frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{k\neq j} u_j u_k E[Z_j Z_k]$
$= \frac{1}{n}\sum_{i=1}^{n} u_i^2 \frac{n}{N} - \frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{k\neq j} u_j u_k \frac{n(n-1)}{N(N-1)}$
$= \frac{1}{N}\sum_{i=1}^{n} u_i^2 - \sum_{j=1}^{n}\sum_{k\neq j} u_j u_k \frac{1}{N(N-1)} = \frac{1}{N(N-1)}[\sum_{j=1}^{N}\sum_{k\neq j} u_j^2 - \sum_{j=1}^{N}\sum_{k\neq j} u_j u_k]$=R.H.S..
Q.E.D.

2 Q2

Population: all worms in the field.

Sampling units: do partition the field into say, n, disjoint parts with equal size. We take $k \leq n$ of them. And all worms in each chosen part is a sampling unit.

Frame: the process stated above is to construct a frame, i.e., the frame is consisted of those k sampling units.

Perform SRS: we can label each partition with $i, 1 \leq i \leq n$. We can use random numbers in Excel to select k from the n.

No, the size does not matter. Because the probability of elements in a larger area or a smaller area have the same probability of being selected.

We need to consider the cost, because investigation into a larger sample cost more naturally.

3 Q3

Note that $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{430}{1000} = 0.43$. $n = 1000, N = 9900$
Also, $1 - \alpha = 95\%$, hence $\alpha = 0.05$. $t_{999,0.975} = 1.96$.
Note that both $n\hat{p}, n(1-\hat{p}) > 5$.

To calculate the 95% confidence interval for p. It should be

$$\hat{p} \pm \sqrt{(1-\frac{n}{N})\frac{1}{n-1}\hat{p}(1-\hat{p})}t_{999,0.975} = [0.401, 0.459].$$

4 Q4

By the formulae in the lecture note, we need

$$n = \frac{1}{\frac{1}{N} + \frac{d^2}{p(1-p)z^2_{1-\frac{\alpha}{2}}}(1-\frac{1}{N})}$$

With $N = 9900, d = 0.02$ and $\alpha = 0.05, z_{1-\frac{\alpha}{2}} = 1.96$. For the p, we utilize the result from last question, plug p with 0.43.

Hence

$$n = \frac{1}{\frac{1}{9900} + \frac{0.02^2}{0.43(1-0.43)1.96^2}(1-\frac{1}{9900})} = 1901.9.$$

Hence 1902 can be a proper sample size.

5 Q5

Denote the estimated total as $\hat{\tau}$. Hence the 95% C.I. for τ is

$$(\hat{\tau} \pm \sqrt{N^2(1-\frac{n}{N})\frac{s^2}{n}}t_{n-1,1-\frac{\alpha}{2}}).$$

With $N = 1500, n = 100, s^2 = 136, \bar{y} = 22.5$, we can first derive that

$$\hat{\tau} = N\bar{y} = 1500 \cdot 22.5 = 33750.$$

Hence the interval should be

$$(33750 \pm \sqrt{1500^2(1-\frac{100}{1500})\frac{136}{100}}\ 1.9842\ ) = (33750 \pm 3353.2393).$$

6 Q6

Note that we have a pre-sample estimate for σ, which is $\sqrt{s^2} = 11.66$.
Hence the proper sample size can be

$$n = \frac{1}{\frac{1}{1500} + \frac{1500^2}{1500^2 s^2 t^2_{n-1,1-\frac{\alpha}{2}}}}.$$

Consult the table we found that when $n = 565$, R.H.S.=565.8022 and when $n = 566$, R.H.S.=565.7979, hence we choose n=566.

7 Q7

Note that the sample mean $\bar{y} = 2$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{20}{9}$, hence $s = 1.49$. Note that $\alpha = 1 - 0.95 = 0.05$. And the the interval should be

$$\bar{y} \pm t_{9,0.975}\sqrt{\frac{N-n}{N}}\frac{s}{\sqrt{n}} = 2 \pm 2.2622 \cdot \sqrt{\frac{100-10}{100}}\frac{1.49}{\sqrt{10}} = 2 \pm 1.0112 = [1, 3].$$