Jan 30th

bill@utstat.toronto.ca

NOTES ABOUT QUIZ 1

Simple Random Sampling

Given population N X from population

$X_1, X_2, \cdots, X_n$

$\bar{X} = \frac{\sum X_i}{n}$

What are we intended in?

mean $E[X] = \mu$

Variance $E[(X-\mu)^2]$

$E[\bar{X}] = \mu$

Let $\hat{\sigma}^2 = \sum (X - \bar{X})^2 / n$ Note $\bar{X}$ is random

$E[\hat{\sigma}] = \sigma^2 \left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right) \leq \sigma^2$

$E\left[\left(\frac{n}{n-1}\right)\left(\frac{N-1}{N}\right) \cdot \hat{\sigma}^2\right] = \sigma^2$

$Var(\bar{X}) = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) \longrightarrow 0$ as $n \to N$

standard error $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ not $Var(\bar{X})$

unbiased estimator of $Var(\bar{X})$ is $\frac{s^2}{n}\left(1 - \frac{n}{N}\right) = S_{\bar{x}}^2$

Confidence intervals

If n large, $\frac{n}{N}$ small

95% CI of μ given $X_1, X_2, \cdots, X_n$

$\bar{X} \pm 1.96 S_{\bar{x}}$

Q1: A population has 6 units with respo-nse values {1,2,3,4,5,6}, a sample of size 3 is taken from the population using simple random sampling without replacement. Calculate

a) the population mean and variance

b). the mean and variance of the distribution of sample mean (Consider sample mean as a random variable).

#1. {1, 2, 3, 4, 5, 6}

1) population mean, variance

2). mean, variance of $\bar{X} = \frac{X_1+X_2+X_3}{3}$

1). $\frac{\sum X_i}{6} = 3.5$

$\frac{1}{6}\sum (X-3.5)^2 = 2.92$

2). $E[\bar{X}] = 3.5$

$E\left[\frac{X_1+X_2+X_3}{3}\right] = \frac{1}{3}\sum E[X_i] = 3.5$

$Var(\bar{X}) = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right) \doteq 0.582$

#2 a). population $N=100$, sample size 10. $\bar{X}=3$, $s^2=5=\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1}$

$E[X]=\bar{X}=3$

$Var(X)=\sigma^2=5\cdot\frac{99}{100}=4.95$

$E[s^2\frac{N-1}{N}]=\sigma^2$

b). 95% CI for μ

$Var(\bar{X})=\frac{s^2}{n}(1-\frac{n}{N})=0.45$

$3\pm1.96\cdot\sqrt{0.45}$

Q2: A sample of size 10 is taken from a normally distributed population of size 100. the mean of this sample is 3, and the variance of this sample is 5. Calculate

a). the best estimation of population mean & variance

b). the 95% confidence interval of the population mean. For a standard normal distribution Z.

$P(Z<1.645)=0.95$,

$P(Z<1.96)=0.975$.

#3. Population size 1000 mean=10 Var=20

Find minimum n s.t. $Var(\bar{X})\leq 0.8$

$Var(\bar{X})=\frac{20}{n}(1-\frac{n-1}{N-1})\leq 0.8$

$20(\frac{N-n}{N-1})\leq 0.8n$

$\Rightarrow n\geq 24.4$

$n\geq 25$

Q3: A population of size 1000 has mean 10 and variance 20. What is the minimum sample size required such that the variance of sample mean (consider sample mean as a random variable) is no bigger than 0.8

#4 Population $N=10000$

Choose k from $\{1,2,\cdots,500\}$ Take $k+500, k+500\times2, \cdots, k+500\times19$

a). P(unit included) k,

a) $\bigcup_{k=1}^{500}\{k+500,\cdots,k+500\times19\}=(1,2,\cdots,10000)$

$P(\{k,k+500,\cdots,k+500\cdot19\})=\frac{1}{500}$

$P(\text{unit } i)=\frac{1}{500}$

b). P(unit i and unit j are included)

$P(\text{unit}_j \mid \text{unit}_i \text{ in sample})=\begin{cases}1 & \text{if } i-j=n\cdot500 \text{ for some } n\in\mathbb{Z}\\ 0 & \text{o.w.}\end{cases}$

$P(i,j)=P(j\mid i)P(i)=\begin{cases}\frac{1}{500} & \text{if } \cdots\\ 0 & \text{o.w.}\end{cases}$

$P(A\mid B)=\frac{P(A\cap B)}{P(B)}$

Q4: given a population size $N=10000$, each unit is uniquely labeled with a number 1 to 10000. The sampling plan is as following: first, randomly choose one unit from the first 500 units, say this is labeled as k, then we choose unit $k+500, k+500\times2, \cdots, k+500\times19$

Calculate the probability that

a). a unit is labeled as i is included in the sample.

b) two units i and j are both included in the sample