MATH2033 Mathematical Statistics
Solution 3

1. The population mean is $\mu = \frac{1}{5}(-1+2+2+4+8) = 3$ with a variance

$$
\begin{aligned}
\sigma^2 &= \frac{1}{5}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&= \frac{1}{5}\left((-4)^2+(-1)^2+(-1)^2+(1)^2+(5)^2\right) \\
&= 8.8
\end{aligned}
$$

All possible samples of size two (with their respective means) are:

$$
\begin{aligned}
&(-1,2)=0.5 &&(-1,2)=0.5 \\
&(-1,4)=1.5 &&(-1,8)=3.5 \\
&(2,2)=2 &&(2,4)=3 \\
&(2,8)=5 &&(2,4)=3 \\
&(2,8)=5 &&(4,8)=6
\end{aligned}
$$

The mean of the sampling distribution is $\frac{1}{10}(0.5+0.5+1.5+3.5+2+3+5+3+5+6) = 3$ and the variance of the sampling distribution is

$$
\begin{aligned}
\sigma^2_{\overline{\mathrm{X}}} &= \frac{1}{10}\sum_{i=1}^{n}\left(\bar{x}_i-\mu_{\overline{\mathrm{X}}}\right)^2 \\
&= \frac{1}{10}\left((-2.5)^2+(-2.5)^2+(-1.5)^2+(0.5)^2+(-1)^2+(0)^2+(2)^2+(0)^2+(2)^2+(3)^2\right) \\
&= 3.3
\end{aligned}
$$

The formula from class said that we should have $\sigma^2_{\overline{\mathbf{X}}} = \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{8.8}{2}\times\frac{3}{4} = 3.3$, as desired.

2. Given 1,2,2,4 and 8, The population proportion p is the number of data values greater than 3 divided by the total number of values:

$$
p = \frac{2}{5} = 0.4
$$

The sample proportion is the number of values greater than 3 divided by the total number of values, which is illustrated in the next table for this sampling.

First sample value	Second sample value	Sample proportion
-1	2	0
-1	2	0
-1	4	0.5
-1	8	0.5
2	-1	0
2	2	0
2	4	0.5
2	8	0.5
2	-1	0
2	2	0
2	4	0.5
2	8	0.5
4	-1	0.5
4	2	0.5
4	2	0.5
4	8	1
8	-1	0.5
8	2	0.5
8	2	0.5
8	4	1

Thus,

$$P[p=0]=\frac{3}{10},\quad P[p=0.5]=\frac{3}{5},\quad P[p=1]=\frac{1}{10}.$$

The mean is

$$\mu_{\hat{p}}=\frac{0+0+0.5+0+\ldots+0.5+0.5+0.5+1}{20}=\frac{8}{20}=0.4$$

The population variance is

$$\sigma^2_{\hat{p}}=\frac{(0-0.4)^2+\ldots.+(1-0.4)^2}{20}\approx 0.09$$

3. (a) All possible samples of size 2 are (repetition of values is not possible);

$$(x_1,x_2),(x_1,x_3),(x_1,x_4),(x_2,x_3),(x_2,x_4),(x_3,x_4)$$

Thus there are 6 possibles samples of sample size two.

(b) $(x_1,x_2),(x_1,x_4),(x_2,x_3),(x_3,x_4)$
Each of these four outcomes will be equally likely:

$$P(\text{ sample })=\frac{1}{4}$$

The expected value of the mean is then the product of the sample means and the probability that the sample occurs. The sample mean is the sum of all values divided by the number of values:

$$E(\bar{X})=\frac{x_1+x_2}{2}\frac{1}{4}+\frac{x_1+x_4}{2}\frac{1}{4}+\frac{x_2+x_3}{2}\frac{1}{4}+\frac{x_3+x_4}{2}\frac{1}{4}=\frac{2x_1+2x_2+2x_3+2x_4}{8}$$
$$=\frac{x_1+x_2+x_3+x_4}{4}=\mu$$

We note that the sample mean is also unbiased in this case, because the expected value of the sample mean is the population mean.

4. (a) Random variables U_i are defined as

$$U_i = \begin{cases} 1, & i\text{-th member of the population is in the sample} \\ 0, & \text{otherwise} \end{cases},$$

for any $i \in \{1, 2, \dots, N\}$. Let S be the set of all indexes i, such that $U_i = 1$. Since the sample is of size n, then S contains exactly n elements; let's name them $i_1, \dots, i_n$. So, $U_i = 1$ for any $i \in S$, and $U_i = 0$, for any $i \notin S$.
Therefore,

$$\begin{aligned} \frac{1}{n} \cdot \sum_{i=1}^{N} U_i x_i &= \frac{1}{n} \cdot \sum_{i \in S} \underbrace{U_i}_{=1} \cdot x_i + \frac{1}{n} \cdot \sum_{i \notin S} \underbrace{U_i}_{=0} \cdot x_i = \frac{1}{n} \cdot \sum_{i \in S} x_i \\ &= \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \bar{X} \end{aligned}$$

(b) We know that

$$P(U_i = 1) = \frac{n}{N}.$$

and note that the probability is the same for all i since each unit in the population is equally likely to be the i-th one .
Therefore,

$$E(U_i) = 1 \cdot P(U_1 = 1) + 0 \cdot P(U_i = 0) = 1 \cdot \frac{n}{N} = \frac{n}{N}.$$

(c) Remember that the variance of the random variable U_i can be found as

$$\text{Var}(U_i) = E\left(U_i^2\right) - \left[E(U_i)\right]^2.$$

Using the definition of the expected value of a (function of a) discrete random variable, we have that

$$E\left(U_i^2\right) = 1^2 \cdot P(U_i = 1) + 0^2 \cdot P(U_i = 0) = \frac{n}{N}.$$

Therefore,

$$\text{Var}(U_i) = \frac{n}{N} - \frac{n^2}{N^2} = \frac{n}{N} \cdot \left(1 - \frac{n}{N}\right).$$

(d) Using the definition of the expected value of a (function of a) discrete random variable, we have that for $i \neq j$, the following holds:

$$\begin{aligned} E(U_i \cdot U_j) &= \sum_{k=0}^{1} \sum_{l=0}^{1} k \cdot l \cdot P(U_i = k, U_j = l) = P(U_i = 1, U_j = 1) \\ &= P(U_i = 1 \mid U_j = 1) \cdot P(U_j = 1) = \frac{n-1}{N-1} \cdot \frac{n}{N}. \end{aligned}$$

(e) Remember that we can find the covariance of two random variables as

$$\text{Cov}(U_i, U_j) = E(U_i \cdot U_j) - E(U_i) \cdot E(U_j).$$

Therefore, using the results of the b) and d) part of this exercise, we have that the covariance of U_i and U_j, when $i \neq j$, is

$$\operatorname{Cov}(U_i, U_j) = \frac{n \cdot (n-1)}{N \cdot (N-1)} - \frac{n^2}{N^2} = \frac{n \cdot (n-N)}{N^2 \cdot (N-1)}.$$

(f) Let σ^2 be the population variance Remember that

$$\begin{aligned}
\sigma^2 &= \frac{1}{N} \cdot \sum_{i=1}^{N} x_i^2 - \bar{X}^2 \\
&= \frac{1}{N} \cdot \sum_{i=1}^{N} x_i^2 - \frac{1}{N^2} \cdot \left(\sum_{i=1}^{N} x_i \right)^2 \\
&= \frac{1}{N} \cdot \sum_{i=1}^{N} x_i^2 - \frac{1}{N^2} \cdot \left(\sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} x_i \cdot x_j \right) \\
&= \frac{N-1}{N^2} \cdot \sum_{i=1}^{N} x_i^2 - \frac{1}{N^2} \cdot \sum_{i=1}^{N} \sum_{\substack{N=1 \\ j \neq i}}^{N} x_i \cdot x_j.
\end{aligned}$$

Also, remember one property of the covariance that we will use in the calculations:

$$\operatorname{Cov}(aX, bY) = ab \cdot \operatorname{Cov}(X, Y), \quad \forall a, b \in \mathbb{R}.$$

Now it follows that the variance of $\bar{X}$ (using the representation from the a) part) can be calculated as

$$\begin{aligned}
&\operatorname{Var}(\bar{X}) = \operatorname{Var}\left(\frac{1}{n} \cdot \sum_{i=1}^{N} U_i x_i \right) \\
&= \frac{1}{n^2} \cdot \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \operatorname{Cov}(U_i x_i, U_j x_j) + \frac{1}{n^2} \cdot \sum_{i=1}^{N} \operatorname{Var}(U_i x_i) \\
&= \frac{1}{n^2} \cdot \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} x_i \cdot x_j \cdot \operatorname{Cov}(U_i, U_j) + \frac{1}{n^2} \cdot \sum_{i=1}^{N} x_i^2 \cdot \operatorname{Var}(U_i) \\
&= \frac{n-N}{n \cdot N^2 \cdot (N-1)} \cdot \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} x_i \cdot x_j + \frac{N-n}{n \cdot N^2} \cdot \sum_{i=1}^{N} x_i^2 \\
&= \frac{N-n}{n \cdot (N-1)} \cdot \underbrace{\left(\frac{N-1}{N^2} \cdot \sum_{i=1}^{N} x_i^2 - \frac{1}{N^2} \cdot \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} x_i \cdot x_j \right)}_{=\sigma^2} \\
&= \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}.
\end{aligned}$$

5. (a)

$$\bar{X} = \frac{1}{25} \cdot \sum_{i=1}^{25} x_i = \frac{104 + 109 + 111 + \cdots + 92 + 97}{25} = \frac{2451}{25} = 98.04.$$

(b)

$$\hat{\sigma}^2 = \frac{N-1}{N} \cdot S^2 = \frac{N-1}{N \cdot (n-1)} \cdot \sum_{i=1}^{n} \left(x_i - \bar{X}\right)^2.$$

Substituting all the values into the formula for S^2 yleids

$$S^2 = \frac{1}{25-1} \cdot \sum_{i=1}^{25} (x_i - 98.04)^2 = \frac{3208.96}{24} = 133.71.$$

Therefore,

$$\sigma^2 = \frac{1999}{2000} \cdot 133.71 = 133.64.$$

Next, remember that an unbiased estimate of the variance of the sample mean is given as

$$S_{\bar{X}}^2 = \frac{S^2}{n} \cdot \frac{N-n}{N}.$$

So, substituting all the values into that formula yields that

$$S_{\bar{X}}^2 = \frac{133.71}{25} \cdot \frac{1975}{2000} = 5.281545$$

(c) Remember that an approximate $(1-\alpha)$ confidence interval for the population mean μ is of the form

$$\left[\bar{X} - z(\alpha/2) \cdot s_X, \bar{X} + z(\alpha/2) \cdot s_X\right],$$

Taking values in it, we obtain

$$[98.04 - 1.96 \cdot 2.298, 98.04 + 1.96 - 2.298] = [93.536, 102.544]$$