TUTORIAL 2 SOLUTIONS

#7.7.11 Consider a population of size four, the members of which have values $x_1, x_2, x_3, x_4$.

**a.** If simple random sampling were used, how many samples of size two are there?

**b.** Suppose that rather than simple random sampling, the following sampling scheme is used. The possible samples of size two are

$$\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_1, x_4\}$$

and the sampling is done in such a way that each of these four possible samples is equally likely. Is the sample mean unbiased?

**Solution**

**a.** The number of simple random samples of size 2 is $\binom{4}{2} = 6$. 
b. Let $\mu$ denote the population mean. I.e.

$$\mu = \frac{x_1 + x_2 + x_3 + x_4}{4}.$$

We observe that

$$E(\bar{X}) = \frac{1}{4}\left[\frac{x_1 + x_2}{2} + \frac{x_2 + x_3}{2} + \frac{x_3 + x_4}{2} + \frac{x_1 + x_4}{2}\right]$$

$$= \frac{x_1 + x_2 + x_3 + x_4}{4}$$

$$= \mu.$$

This implies that the sample mean is an unbiased estimate of the population mean (even though the sample is not a s.r.s.).
#7.7.12 Consider simple random sampling with replacement.

a. Show that

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

is an unbiased estimate of \( \sigma^2 \).

b. Is \( s \) an unbiased estimate of \( \sigma \)?

c. Show that \( n^{-1} s^2 \) is an unbiased estimate of \( \sigma^2 \bar{X} \).

d. Show that \( n^{-1} N^2 s^2 \) is an unbiased estimate of \( \sigma^2_T \).

e. Show that \( \hat{p}(1 - \hat{p})/(n - 1) \) is an unbiased estimate of \( \sigma^2_{\hat{p}} \).

**Solution** First we observe that the sample \( X_1, \ldots, X_n \) is an i.i.d. sequence of random variables.
a. Let $\mu = E(X_1)$ and $Y_i = X_i - \mu$. Then

$$E(Y_i) = 0,$$

$$\text{Var}(Y_i) = \text{Var}(X_i) = \sigma^2,$$

$$E(s^2) = \frac{1}{n-1} \sum_{i=1}^{n} E[(X_i - \mu - (\bar{X} - \mu))^2]$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} E(Y_i - \bar{Y})^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} E(Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2)$$

$$= \frac{n\sigma^2}{n-1} - \frac{2n}{n-1} E(\bar{Y}^2) + \frac{n}{n-1} E(\bar{Y}^2)$$

$$= \frac{n\sigma^2}{n-1} - \frac{n}{n-1} E(\bar{Y}^2).$$
Since \( E(\bar{Y}^2) = \text{Var}(\bar{Y}) = \sigma^2/n \),

\[
E(s^2) = \frac{n\sigma^2}{n-1} - \frac{\sigma^2}{n-1} = \sigma^2.
\]

This implies that \( s^2 \) is an unbiased estimate of \( \sigma^2 \).

**b.** No, \( s \) is not an unbiased estimate of \( \sigma \) since in general

\[
E(s) = E(\sqrt{s^2}) < \sqrt{E(s^2)} = \sqrt{\sigma^2} = \sigma.
\]

This follows from Jensen’s inequality and

\( E(s) = \sqrt{E(s^2)} \) only if \( s \) is a constant.
c. From a. we have

\[ E(n^{-1}s^2) = \frac{1}{n}E(s^2) \]

\[ = \frac{\sigma^2}{n} \]

\[ = \sigma^2 \bar{X}. \]

That is \( n^{-1}s^2 \) is an unbiased estimate of the variance of \( \bar{X} \).

d. Recall that \( T = N\bar{X} \) and \( \text{Var}(T) \) is given by

\[ \sigma_T^2 = \text{Var}(N\bar{X}) \]

\[ = N^2\text{Var}(\bar{X}) \]

\[ = N^2\frac{\sigma^2}{n}. \]

Consequently,

\[ E(n^{-1}N^2s^2) = n^{-1}N^2E(s^2) \]

\[ = n^{-1}N^2\sigma^2 \]

\[ = \sigma_T^2. \]
This shows that $n^{-1}N^2s^2$ is an unbiased estimate of $\sigma_T^2$.

e. Recall that $\hat{p} = \bar{X}$ when the $X_i$’s only take values 0 or 1. Observe that

$$\sigma_{\hat{p}}^2 = \sigma_X^2$$

$$= \frac{\sigma^2}{n}.$$

Consequently,

$$E\frac{\hat{p}(1 - \hat{p})}{n - 1} = \frac{1}{n - 1}E\left[\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2\right]$$

$$= \frac{1}{n(n - 1)} \sum_{i=1}^{n} E(X_i - \bar{X})^2$$

$$= \frac{1}{n}E(s^2) = \frac{\sigma^2}{n}.$$  

This shows that $(1 - \hat{p})/(n - 1)$ is an unbiased estimate of $\sigma_{\hat{p}}^2$. 

7
#7.7.23

**a.** Show that the standard error of an estimated proportion is largest when $p = 1/2$.

**b.** Use this result and Corollary B of Section 7.3.2 to conclude that the quantity
\[
\frac{1}{2} \sqrt{\frac{N - n}{N(n - 1)}}
\]

is a conservative estimate of the standard error of $\hat{p}$ no matter what the value of $p$ may be.

**c.** Use the central limit theorem to conclude that the interval
\[
\hat{p} \pm \sqrt{\frac{N - n}{N(n - 1)}}
\]

contains $p$ with probability at least 0.95.
Solution

a. Recall that the sample is s.r.s. From page 214 of the text, the standard error of \( \hat{p} \) is given by

\[
\sigma_{\hat{p}} = \sqrt{\frac{N - n}{N - 1} \frac{p(1 - p)}{n}}.
\]

Treating \( \sigma_{\hat{p}} \) as a function of \( p \), we note that \( f(p) = p(1 - p), 0 \leq p \leq 1 \), is maximized when \( p = 1/2 \).

b. Using a. and Corollary B of Chapter 7.3.2, we have

\[
s_{\hat{p}}^2 = \frac{\hat{p}(1 - \hat{p})}{n - 1} (1 - \frac{n}{N})
\leq \frac{1}{2} \left( 1 - \frac{1}{n} \right) \left( \frac{1}{n - 1} \right) (1 - \frac{n}{N})
\]

\[= \frac{N - n}{4N(n - 1)}.
\]
Taking square root, we have

\[ s\hat{p} \leq \frac{1}{2} \sqrt{\frac{N - n}{N(n - 1)}}. \]

The r.h.s. is a conservative estimate of the standard error of \( \hat{p} \).

**c.** Using the CLT from s.r.s., a 95% CI for \( p \) is

\[ \hat{p} \pm z_{0.025} s\hat{p}. \]

Since \( z_{0.025} = 1.96 \approx 2 \), it follows from **b.** that the interval

\[ \hat{p} \pm \sqrt{\frac{N - n}{N(n - 1)}}. \]

contains \( p \) with probability at least 0.95.
#7.7.24 For a random sample of size $n$ from a population of size $N$, consider the following as an estimate of $\mu$:

$$\bar{X}_c = \sum_{i=1}^{n} c_i X_i$$

where the $c_i$ are fixed numbers and $X_1, \ldots, X_n$ is the sample.

**a.** Find a condition of the $c_i$ such that the estimate is unbiased.

**b.** Show that the choice of $c_i$ that minimizes the variance of the estimate subject to this condition is $c_i = 1/n$, where $i = 1, \ldots, n$. 
**Solution** Note that in this problem, a random sample is meant an i.i.d. sample.

a. \( \bar{X}_c \) is an unbiased estimate of the population mean \( \mu \) iff

\[
\mu = E(\bar{X}_c) = \sum_{i=1}^{n} c_i E(X_i) \\
= (\sum_{i=1}^{n} c_i) \mu.
\]

This implies that the condition for unbiasedness is

\[
\sum_{i=1}^{n} c_i = 1.
\]
b. Let $\text{Var}(X_1) = \sigma^2$. Then by the independence of $X_1, \ldots, X_n$, we have

$$\text{Var}(\bar{X}_c) = \text{Var}(\sum_{i=1}^{n} c_i X_i)$$

$$= \sum_{i=1}^{n} c_i^2 \text{Var}(X_i)$$

$$= \sigma^2 \sum_{i=1}^{n} c_i^2.$$

We want to minimize $\sum_{i=1}^{n} c_i^2$ subject to $\sum_{i=1}^{n} c_i = 1$.

To do that we minimize the following Lagrangian function:

$$f(c_1, \ldots, c_n, \lambda) = \sum_{i=1}^{n} c_i^2 + \lambda (1 - \sum_{i=1}^{n} c_i).$$

Here $\lambda$ is the Lagrangian multiplier.
We partial differentiate $f$ wrt $c_1, \ldots, c_n, \lambda$,

$$\frac{\partial f}{\partial c_i} = 2c_i - \lambda, \quad i = 1, \ldots, n,$$
$$\frac{\partial f}{\partial \lambda} = 1 - \sum_{i=1}^{n} c_i.$$

Equating these partial derivatives to 0, we have

$$c_i = \frac{\lambda}{2}, \quad i = 1, \ldots, n,$$
$$1 = \sum_{i=1}^{n} c_i = \frac{\lambda n}{2}.$$

Solving for $c_i$ and $\lambda$, we have

$$\lambda = \frac{2}{n},$$
$$c_i = \frac{1}{n}, \quad i = 1, \ldots, n.$$
By taking 2nd partial derivatives of $f$, it can be shown that this gives the minimum (and not maximum) of $f$. 