Questions from class notes HT 9-2-15

For the example from class notes, and scores $\mathbf{y} = (-1, 3, 2, 1)$

1. Verify the table of $T_{ij}$ give above, correct as necessary.

   Since $T_{ij} = P(\text{in both in the sample})$ and $T_{ii} = T_{i}$

   So, we have: $T_{11} = T_{1} = P(1 \text{ in the sample $A$}) = 0.1 + 0.4 = 0.5$

   $T_{12} = P(1 \text{ in Sample $B$}) = 0.4$

   $T_{13} = P(1 \text{ in Sample $B$}) = 0.1$

   $T_{14} = P(1 \text{ in Sample $B$}) = 0.4$

   $T_{21} = P(2 \text{ in Sample $B$}) = 0.4$

   $T_{22} = T_{2} = P(2 \text{ in Sample $B$}) = 0.2 + 0.4 + 0.1 = 0.9$

   $T_{23} = P(2 \text{ in Sample $B$}) = 0.1$

   $T_{24} = P(2 \text{ in Sample $B$}) = 0.9$

   $T_{31} = P(3 \text{ in Sample $B$}) = 0.1$

   $T_{32} = P(3 \text{ in Sample $B$}) = 0.4$

   $T_{33} = T_{3} = P(3 \text{ in Sample $B$}) = 0.1 + 0.2 = 0.3$

   $T_{34} = P(3 \text{ in Sample $B$}) = 0.2$

   $T_{41} = P(4 \text{ in Sample $B$}) = 0.4$

   $T_{42} = P(4 \text{ in Sample $B$}) = 0.9$

   $T_{43} = P(4 \text{ in Sample $B$}) = 0.1$

   $T_{44} = P(4 \text{ in Sample $B$}) = 0.9$

   \[
   \begin{align*}
   T_{11} + T_{12} + T_{13} + T_{14} &= 1 \\
   T_{21} + T_{22} + T_{23} + T_{24} &= 1 \\
   T_{31} + T_{32} + T_{33} + T_{34} &= 1 \\
   T_{41} + T_{42} + T_{43} + T_{44} &= 1
   \end{align*}
   \]

   \[ \Rightarrow \text{what we got above is the same as the table of $T_{ij}$} \]

   So, the table of $T_{ij}$ is verified.
1. Calculate $T$ and all four $T_{ij}$, one for each possible sample $\mathbf{\alpha}$.

For $\mathbf{\alpha} = \{1, 3\}$

$$ T_{13} = \sum_{i \in \{1, 3\}} \frac{y_i}{p_i} = \frac{y_1}{p_1} + \frac{y_3}{p_3} = \frac{-1}{0.5} + \frac{2}{0.5} = \frac{4}{3} $$

For $\mathbf{\alpha} = \{2, 3, 4\}$

$$ T_{234} = \sum_{i \in \{2, 3, 4\}} \frac{y_i}{p_i} = \frac{y_2}{p_2} + \frac{y_3}{p_3} + \frac{y_4}{p_4} = \frac{3}{0.9} + \frac{2}{0.3} + \frac{1}{0.9} = \frac{100}{9} $$

For $\mathbf{\alpha} = \{2, 4\}$

$$ T_{24} = \sum_{i \in \{2, 4\}} \frac{y_i}{p_i} = \frac{y_2}{p_2} + \frac{y_4}{p_4} = \frac{3}{0.9} + \frac{1}{0.9} = \frac{40}{9} $$

For $\mathbf{\alpha} = \{1, 2, 4\}$

$$ T_{124} = \sum_{i \in \{1, 2, 4\}} \frac{y_i}{p_i} = \frac{y_1}{p_1} + \frac{y_2}{p_2} + \frac{y_4}{p_4} = \frac{-1}{0.5} + \frac{3}{0.9} + \frac{1}{0.9} = \frac{22}{9} $$

So, $T = 5$. $T_{ij} = \frac{14}{3}, \frac{100}{9}, \frac{40}{9}, \frac{22}{9}$ for each sample $\mathbf{\alpha}$.

2. Use (1), and the four $P(i)$ values, to verify that indeed $E[T_{ij}] = T$.

By using (1) and $P(i)$ values, we have:

$$ E[T_{ij}] = \frac{14}{3} \cdot 0.1 + \frac{100}{9} \cdot 0.2 + \frac{40}{9} \cdot 0.3 + \frac{22}{9} \cdot 0.4 $$

$$ = \frac{14}{3} + \frac{40.8}{9} $$

$$ = \frac{145}{9} $$

$$ = 5 $$

$$ = T $$

So, $E[T_{ij}] = T$. \( \Box \)
3. Use (1), and the four P(s) values, to calculate \( \text{Var} \hat{\theta}_N \).

By using (1), and the P(s) values, we have:

\[
\text{Var} \hat{\theta}_N = 0.1 \times \left( \frac{16}{5} - 5 \right)^2 + 0.2 \times \left( \frac{100}{8} - 5 \right)^2 + 0.13 \times \left( \frac{80}{7} - 5 \right)^2 + 0.14 \times \left( \frac{22}{9} - 5 \right)^2
\]

\[
= \frac{275}{27}
\]

So, \( \text{Var} \hat{\theta}_N = \frac{275}{27} \).

Also, we can use:

\[
\text{Var} \hat{\theta}_N = \frac{1}{n} \left( \text{PM}_1 - \text{PM}_2 \right) \frac{y_i}{n} \frac{y_j}{n} \quad \text{to calculate with all the situations = (0.4 - 0.3)(0.01) + \ldots + (0.9 - 0.9)(0.9) + \ldots} \]

\[
= \frac{275}{27}
\]

4. Use (1), and sample \( x = \{2, 4 \} \), to calculate \( \text{Var} \hat{\theta}_N \).

Since we have:

\[
\text{Var} \hat{\theta}_N = \Sigma_{ij} a \frac{\text{PM}_i - \text{PM}_j}{\text{PM}_i \times \text{PM}_j} \frac{y_i}{n} \frac{y_j}{n}
\]

So, when \( x = \{2, 4 \} \), we have four possible combinations: (2,2), (4,4), (2,4), (4,2).

\[
\text{Var} \hat{\theta}_N = \Sigma_{ij \in \{2,4\}} \frac{\text{PM}_i - \text{PM}_j}{\text{PM}_i \times \text{PM}_j} \frac{y_i}{n} \frac{y_j}{n}
\]

\[
= \frac{\text{PM}_2 - \text{PM}_2}{\text{PM}_2 \times \text{PM}_2} y_2^2 + \frac{\text{PM}_4 - \text{PM}_4}{\text{PM}_4 \times \text{PM}_4} y_4^2 + \frac{\text{PM}_2 - \text{PM}_4}{\text{PM}_2 \times \text{PM}_4} y_2 \cdot y_4
\]

\[
= \frac{10}{9} + \frac{10}{9} + \frac{10}{9} + \frac{10}{21}
\]

\[
= \frac{160}{81}.
\]
5. If CLT applied (NOT so), what would be the 95% CI for $T$ based on Sample (2.14) ?

Since CLT applied, then the 95% CI for $T$ should be:

$$
\hat{T}_{HT} \pm Z \sqrt{\text{Var}\hat{T}_{HT}}
$$

where $Z = 1.96$, $\text{Var}\hat{T}_{HT} = \frac{160}{81}$, $\hat{T}_{HT} = \frac{40}{9}$ for $Z = 1.96$.

$$
\Rightarrow \hat{T}_{HT} \pm 1.96\sqrt{\frac{160}{81}}
$$

So the 95% CI for $L.H.$ = (1.69, 17.20).

6. From Sample (2.14) calculate $\hat{N}_{HT}$ and the ratio estimator $\hat{\theta}_{HT} = \hat{T}_{HT} / \hat{N}_{HT}$ of $\theta = T/N$.

We have $\hat{N}_{HT} = \frac{N}{\sum_{i=1}^{N}(2i+1)} \frac{1}{\Pi_i}$

for sample (2.14):

$$
\hat{N}_{HT} = \frac{N}{\sum_{i=1}^{9}(2i+1)} \frac{1}{\Pi_i} = \frac{1}{\Pi_3} + \frac{1}{\Pi_4} = \frac{20}{9}
$$

Ratio estimator $\hat{\theta}_{HT} = \frac{\hat{T}_{HT}}{\hat{N}_{HT}} = \frac{\frac{40}{9}}{\frac{20}{9}} = \frac{9}{50}$

Therefore, $\hat{\theta}_{HT} = 2$. 
7. Comment on the potential importance of the fact that each of the HT estimators considered above may be calculated after the collection of the sample and only for our particular sample.

For example, the figure on the right side shows our plan, we can have many ways to walk to gather the sample, and we may encounter some incidence during walking, for instance the shading part, but we will find our way for each sample. For HT estimators, if we gather the samples and we have all the \( p(i) \) and \( p(i, j) \) (we're obligated to make sure \( p(i, j) > 0 \)), we can calculate the HT estimators anyway, and HT estimators are more flexible, only if we have the sample and \( p(i) \) and \( p(i, j) \), the problem solves itself without any difficult mathematic calculations, and that's the potential importance of this fact.

Exercises from chapter 2.

1. (a) A random sample without repl of \( n = 10 \) was selected from \( N = 120 \).

   1. List the sample data.
   2. Use sample to estimate the number of objects in figure.
   3. Estimate variance of estimator.

   \( \bar{Y} = \frac{1}{10} (0 + 1 + 1 + 1 + 0 + 2 + 4 + 1 + 0 + 1) = 1.1 \)
   \( S^2 = \frac{1}{10-1} \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = \frac{1}{9} (0.21 + 0.01 + 0.01 + 0.51 + 0.81 + 0.04 + 0.01 + 0.21 + 0.01) = \frac{43}{30} \)
   \( \text{Var}(\bar{Y}) = \frac{(N-n)}{N} \cdot \frac{S^2}{n} = \frac{90}{100} \cdot \frac{43}{30} \cdot \frac{1}{10} = 0.129 \)
   \( \Rightarrow \hat{N} = N \bar{Y} = 120 \times 1.1 = 110 \)

   \( \text{Var}(\hat{N}) = N^2 \text{Var}(\bar{Y}) = N(N-n) \cdot \frac{S^2}{n^2} = 1290 \)
(b) Repeat (a). Selecting another sample of size 10 by SRS and making new estimates. Indicate the positions of the units of samples on sketch.

$$\bar{Y} = \frac{1}{10} (1 + 4 + 2 + 2 + 1 + 0 + 2 + 0 + 0 + 1) = \frac{13}{10} = 1.3$$

$$S^2 = \frac{1}{10-1} \left( \sum_{i=1}^{10} (X_i - \bar{Y})^2 \right)$$

$$= \frac{1}{9} \left( 0.09 + 2.7^2 + 0.49 + 0.49 + 0.09 + 1.3^2 + 0.49 + 1.3^2 + 1.3^2 + 0.09 \right)$$

$$= \frac{14.5}{9}$$

$$\text{Var}(\bar{Y}) = \frac{(N-n)}{N} \cdot \frac{S^2}{n}$$

$$= \frac{90}{100} \times \frac{14.5}{9} \times \frac{1}{10}$$

$$= 0.1145$$

$$\text{Var} (\hat{Y}) = N \cdot \text{Var}(\bar{Y}) = 100 \times 1.3 = 130$$

$$\text{Var} (\hat{Y}) = N^2 \cdot \text{Var}(\bar{Y}) = 100^2 \times 0.145 = 1450$$

Units selected are indicated by shading in the figure, on the left.
(C) Give inclusion probability for the unit in the upper left-hand corner.

7. How many possible samples are there? What is the prob of selecting the sample you obtained in (a)?

\[ P = \frac{1}{\binom{N}{n}} \]

\[ = \frac{\binom{100}{10}}{10! \cdot 90!} \]

2. SRS of 10 is selected from 100. The data: (2, 5, 1, 4, 4, 3, 2, 5, 2, 3)

(a) Estimate the total number of people in the pop estimate the variance of estimate

\[ N = 100 \]
\[ n = 10 \]

\[ \bar{Y} = \frac{1}{10} (2 + 5 + 1 + 4 + 4 + 3 + 2 + 5 + 2 + 3) = 3.1 \]

\[ S^2 = \frac{1}{10-1} \left( Y_i - \bar{Y} \right)^2 \]

\[ = \frac{9}{9} \left( 1.1^2 + 0.9^2 + 2.1^2 + 0.9^2 + 0.1^2 + 1.1^2 + 0.9^2 + 1.1^2 + 0.1^2 + 0.1^2 \right) \]

\[ = 0.169 \]

\[ \text{Var}(\bar{Y}) = \frac{(N-n)}{N} \cdot \frac{S^2}{n} = \frac{90}{100} \times \frac{0.169}{10} = 0.169 \]

\[ \Rightarrow \begin{align*}
\hat{\mu} &= \frac{N}{n} \bar{Y} = 100 \times 3.1 = 310 \\
\text{Var}(\hat{\mu}) &= N^2 \text{Var}(\bar{Y}) = 100 \times 100 \times 0.169 = 1690
\end{align*} \]
(b) estimate the mean number of people per household and variance of that estimator.

\[
\bar{y} = \frac{1}{n} (y_1 + \ldots + y_n) \\
\bar{y} = \frac{1}{n} (2+5+4+4+3+2+5+2+2) \\
\bar{y} = 3.1 \\
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
S^2 = \frac{1}{9} (1.1^2 + 3.61 + 4.41 + 0.16 + 0.01 + 1.21 + 3.61 + 1.0 + 0.01) \\
S^2 = \frac{16.9}{9} \\
\hat{\sigma}^2(\bar{y}) = \frac{(N-n)}{N} S^2 \frac{1}{n} = \frac{90}{100} \times \frac{16.9}{9} = 0.169
\]

3. A small pop of \( N = 5 \), labeled 1, 2, 3, 4, 5 with \( y \)-values 3.1, 0.15.

Consider ASRS with \( n = 3 \).

(a) Give the values of the pop parameters \( \mu, \sigma, \sigma^2 \).

(b) List every possible sample of size \( n = 3 \).

(c) For each sample, what is the prob that it is selected?

\( \text{I.} \quad T = 3+1+0+1+5 = 10 \)

\( \mu = \frac{\sum y}{N} = \frac{15.5}{5} = 3.1 \)

\( \sigma^2 = \frac{1}{N-1} \sum (y_i - \bar{y})^2 \\
\sigma^2 = \frac{1}{4} (1.1^2 + 3.61 + 4.41 + 0.16 + 0.01 + 1.21 + 3.61 + 1.0 + 0.01) \\
\sigma^2 = 4.0 \)

\( \text{II. for } n = 3: \quad (1, 2, 3), (1, 2, 4), (1, 2, 5), (1, 3, 4), (1, 3, 5) \\
(2, 3, 4), (2, 3, 5), (3, 4, 5), (1, 4, 5), (2, 4, 5) \)

\( \text{III. } P(a) = \frac{1}{10} \text{ for each sample.} \)
(b) For each sample, compute sample mean \( \bar{y} \) and sample median \( m \).

1. Demonstrate sample mean is unbiased for pop mean and determine whether sample median is unbiased for pop median.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample mean ( \bar{y} )</th>
<th>Sample median ( m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3,1,0)</td>
<td>( \frac{4}{3} = 1.33 )</td>
<td>1</td>
</tr>
<tr>
<td>(3,1,1)</td>
<td>( \frac{5}{3} = 1.67 )</td>
<td>1</td>
</tr>
<tr>
<td>(3,1,5)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>(3,0,1)</td>
<td>( \frac{4}{3} = 1.33 )</td>
<td>1</td>
</tr>
<tr>
<td>(3,0,5)</td>
<td>( \frac{8}{3} = 2.67 )</td>
<td>3</td>
</tr>
<tr>
<td>(1,0,1)</td>
<td>( \frac{2}{3} = 0.67 )</td>
<td>1</td>
</tr>
<tr>
<td>(1,0,5)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(0,1,5)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(3,1,5)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>(1,1,5)</td>
<td>( \frac{7}{3} = 2.33 )</td>
<td>1</td>
</tr>
</tbody>
</table>

2. \[ E(\bar{y}) = \left( \frac{1.33 + 1.67 + 3 + 1.33 + 2.67 + 0.67 + 2 + 2 + 3 + 2.33}{10} \right) \]

\[ = \frac{2}{3} \quad \text{(from (a))} \]

So, sample mean is unbiased for pop mean.

3. Since pop median is 1,

\[ E(m) = \frac{1 + 1 + 3 + 1 + 3 + 1 + 1 + 1 + 3 + 1}{10} \]

\[ = 1.6 \]

\( \neq \) pop median

So, sample median is not unbiased.
4. Show that \( E(S^2) = \sigma^2 \) in SRS.

\[
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2
\]

\[
= \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \mu + \mu - \bar{y})^2
\]

\[
= \frac{1}{n-1} \left( \sum_{i=1}^{n} (y_i - \mu)^2 + n(\bar{y} - \mu)^2 - 2 \sum_{i=1}^{n} (y_i - \mu)(\bar{y} - \mu) \right)
\]

\[
= \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \mu)^2 + n(\bar{y} - \mu)^2 - 2n(\bar{y} - \mu)^2
\]

\[
= \frac{n}{n-1} (y_i - \mu)^2 - n(\bar{y} - \mu)^2
\]

\[
\Rightarrow E(S^2) = E\left( \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right)
\]

\[
= \frac{1}{n-1} E\left( \sum_{i=1}^{n} (y_i - \bar{y})^2 \right)
\]

\[
= \frac{1}{n-1} E\left( \sum_{i=1}^{n} (y_i - \mu)^2 - n(\bar{y} - \mu)^2 \right)
\]

\[
= \frac{1}{n-1} E\left( \sum_{i=1}^{n} (y_i - \mu)^2 \right) - \frac{n}{n-1} E(\bar{y} - \mu)^2
\]

\[
= \frac{1}{n-1} E\left( \sum_{i=1}^{n} (y_i - \mu)^2 \right) - 0
\]

\[
= \sigma^2
\]
5. (a) What practical problems arise in establishing a frame?

Take the tree estimating as one example, here we’re interested in estimating the number of trees on MSU campus by dividing the campus into plots. The practical problem that arise here is that we have a map of MSU campus and divide it into N plots, randomly, but some plots may have too many trees and some may do not have trees.

(b) How is the sample selection actually carried out?

We have N plots, plot1, 2, ... N. We want to select n plots from N.

We select n numbers from 1, 2, ... N randomly, by using R we’ll write the code like:

```R
set.seed(1)
N = 3
n = 2
sample(1:N, n)
```

(c) What special problems arise in observing the units selected?

When observing the plots, we may find that some plots contain too much trees, but some plots may not contain any tree because we divide the map randomly with no actual information about the locations of the trees in campus.

(d) Estimate the pop mean and total

We use sample mean to estimate population mean: \( \bar{y} = \frac{1}{n}(y_1 + y_2 + ... + y_n) \)
and unbiased estimator of population total is \( \widehat{T} = N \bar{y} \).

(e) Estimate the variance of the estimators in (d)

The variance of \( \hat{y} \) is:
\[
\text{Var}(\hat{y}) = \frac{(N-n)}{N} \frac{s^2}{n}
\]

the variance of \( \hat{T} \) is:
\[
\text{Var}(\hat{T}) = N^2 \text{Var}(\hat{y}) = N(N-n) \frac{s^2}{n}
\]

(f) How would you improve the survey procedure if you were to do it again?

I will do a prior investigation of the distribution of tree’s location in MSU campus, to get a rough information in order to avoid the bias that happens when divide the campus into N plots, that some plots contain too many.
true but some plots contain nothing, because this will influence our estimation.
6. Repeat the simulation exercise with the tree data

(a) \( n = 10 \)

```r
> y <- trees$Volume
> N <- 31
> n <- 10
> s <- sample(1:N, n)
> s
[1] 18  7 17 29 12 14 26 11 23  5
> y[s]
[1] 27.4 15.6 33.8 51.5 21.0 21.3 55.4 24.2 36.3 18.8
> mean(y[s])
[1] 30.53
> s <- sample(1:N, n)
> s
[1]  6 24 14  8 17 25 28 29  5 26
> mean(y[s])
[1] 35.79
> mu <- mean(y)
> mu
[1] 30.17097
```

```r
[4]
> b <- 6
> ybar <- numeric(b)
> for (k in 1:b){
+ s <- sample(1:N, n)
+ ybar[k] <- mean(y[s])
+ }
> ybar
[1] 25.52 29.95 24.23 27.02 26.35 19.82
> b <- 10000
> for (k in 1:b){
+ s <- sample(1:N, n)
+ ybar[k] <- mean(y[s])
+ }
> ybar
[12] 20.89 23.85 32.84 30.20 27.26 32.39 27.59 27.77 27.76 27.45 27.44
[34] 24.98 36.61 35.74 29.25 32.77 33.73 27.04 26.36 21.71 28.10 30.49
[45] 23.89 29.79 28.11 35.65 35.47 33.54 31.37 23.23 27.32 30.29 28.05
[56] 26.53 24.51 31.84 31.12 30.96 24.61 35.61 38.88 29.91 30.76 32.41
[67] 31.31 29.90 31.36 28.80 32.36 26.69 31.89 28.95 27.12 28.15 31.82
[78] 31.45 20.16 32.56 30.40 31.78 26.77 29.17 32.64 30.27 34.91 27.49
[89] 37.15 21.23 27.70 36.71 31.50 36.58 30.45 33.63 31.12 30.66 35.10
[100] 25.15 26.92 36.08 24.70 28.52 25.25 32.77 22.08 30.33 29.42 36.93
```
```r
[9912] 22.89 35.89 32.02 29.67 37.59 32.88 31.69 31.54 27.36 34.00 21.82
[9923] 36.30 25.39 26.69 34.50 33.75 26.69 35.57 31.42 34.01 34.44 26.25
[9934] 33.51 31.44 32.16 26.28 27.75 30.02 24.51 21.38 34.36 28.16 36.04
[9945] 31.32 27.86 30.15 25.39 29.05 24.78 37.94 21.22 32.63 30.92 29.60
[9956] 30.36 28.52 29.39 29.71 22.98 38.31 36.69 29.60 30.36 37.90 37.66
[9967] 31.48 27.13 25.86 35.07 36.07 33.21 24.06 23.52 34.67 23.71 29.34
[9978] 30.92 35.60 28.84 34.64 22.65 32.40 25.26 32.75 27.77 25.61 30.89
[9989] 19.61 34.96 25.89 37.31 25.17 31.67 32.96 22.39 32.73 28.95 30.01
[10000] 25.02
>
> hist(ybar)
> mean(ybar)
[1] 30.15357
> var(ybar)
[1] 18.26799
> (1-n/N)*var(3)/n
[1] 18.30406
> sd(ybar)
[1] 4.27646
> sqrt((1-n/N)*var(y)/n)
[1] 4.276324
> mean(ybar - mu)^2
[1] 10.25646
> }

almost the same as those in the textbook.
```
From above we can see that my results are almost the same as those in the text when $n=10$. 
(b) n=15

```
> y <- trees$Volume
> N <- 31
> n <- 15
> s <- sample(1:N, n)
> y[s]
[1] 23 19 15 5 29 10 18 1 17 12 3 30 6 25 20
> mean(y[s])
[1] 27.44
> s <- sample(1:N, n)
> y[s]
> mean(y[s])
[1] 30.48
> m2 <- mean(y)
> m2
[1] 30.17097
> b <- 6
> ybar <- numeric(6)
> for (k in 1:b){
+ s <- sample(1:N,n)
+ ybar[k] <- mean(y[s])
+ }
> ybar
```

```
[1] 27.61333 32.66667 32.22222 27.05556 33.03333 27.78000
```

```
> b <- 10000
> for (k in 1:b){
+ s <- sample(1:N,n)
+ ybar[k] <- mean(y[s])
+ }
> ybar
```

```
[57] 29.25333 29.42667 32.41333 31.66667 33.79333 34.54667 29.27333 28.86667  
[81] 28.32667 37.76000 34.02000 29.45333 36.89333 32.08000 30.53333 25.31333  
```
The variance we get here is much more smaller than the variance we get when n=10. Their mean is almost the same.
Compare this result to those of part (a), we can see that the histogram becomes more precise and narrows the interval of ybar down, also by comparing the numerical results mean and variance of the estimator we can see that mean is almost the same but the variance becomes smaller very obviously which is a great improvement when n becomes 15.

\[
\begin{array}{llll}
  n = 10 & n = 15 \\
  \text{mean} & 30.15 & 30.23 & \rightarrow \text{do not change a lot} \\
  \text{variance} & 18.28 & 9.07 & \rightarrow \text{becomes much more smaller}
\end{array}
\]