HW 4 due 9-11-15

**Exercises**

<table>
<thead>
<tr>
<th>D</th>
<th>P(x)</th>
<th>(y_i ), (i \in \Delta)</th>
<th>(y_i/y_i)</th>
<th>(y_i/y_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31,33</td>
<td>1</td>
<td>1</td>
<td>.5</td>
</tr>
<tr>
<td>2</td>
<td>32,3,43</td>
<td>2</td>
<td>2</td>
<td>.4</td>
</tr>
<tr>
<td>3</td>
<td>32,43</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>31,2,43</td>
<td>4</td>
<td>4</td>
<td>.4</td>
</tr>
</tbody>
</table>

\[
y = (-1,3,2,1)\]
\[
\theta = (1,2,3,4)
\]

\[
\Pi_{i,j} = 1, 2, 3, 4
\]

all values agree with the probabilities of each \( \Delta \)

\[
\begin{align*}
\text{D1: } \hat{\pi}_{HT} & = \sum_{i=1}^{\Delta} \left( \pi_i \right) \frac{y_i}{y_i} = (1) -\frac{1}{5} + (0) \frac{3}{9} + (1) \frac{2}{3} + (0) \frac{1}{9} \\
& = 4.67 \\
\text{D2: } \hat{\pi}_{HT} & = (0) -\frac{1}{5} + (1) \frac{3}{9} + (1) \frac{2}{3} + (1) \frac{1}{9} \approx 11.11 \\
\text{D3: } \hat{\pi}_{HT} & = (0) -\frac{1}{5} + (1) \frac{3}{9} + (0) \frac{2}{3} + (1) \frac{1}{9} \approx 4.44 \\
\text{D4: } \hat{\pi}_{HT} & = (1) -\frac{1}{5} + (1) \frac{3}{9} + (0) \frac{2}{3} + (1) \frac{1}{9} \approx 2.44
\end{align*}
\]

\[
\text{E(} \hat{\pi}_{HT} \text{)} = 1 (4.67) + 2 (11.11) + 3 (4.44) + 4 (2.44)
\]

\[\pi \leq \hat{\pi} \leq \sqrt{5} \checkmark\]

\[
\text{Var} \hat{\pi}_{HT} = \sum_{i,j} \left( \pi_{ij} \right) \frac{y_i}{y_i} \cdot (\bar{y}_i/y_i - \bar{y}_i)^2
\]

\[
= \frac{1}{5} (4.67 - 5)^2 + 2 (11.11 - 5)^2 + 3 (4.44 - 5)^2 + 4 (2.44 - 5)^2 \approx 10.19
\]

\[
\hat{\gamma} = \sum_{i,j} \left( \pi_{ij} \right) \frac{y_i}{y_i} = (1) \left( \frac{9-\frac{1}{9}}{9} \right) \cdot 3^2 + (1) \left( \frac{9-\frac{1}{9}}{9} \right) \cdot 2^2 + (\sqrt{2}) \left( \frac{9-\frac{1}{9}}{9} \right) \cdot 1^2 + (\sqrt{3}) \left( \frac{9-\frac{1}{9}}{9} \right) \cdot 0 \approx 6.04
\]
\[ S - (1.96)\text{(SD)}; S + (1.96)\text{(SD)} \]
\[ = [S - (1.96)(5.6), S + (1.96)(5.6)] \]
\[ = [2.52, 7.48] \]

\[ \hat{\theta}_HT = \frac{\sum_1^n y_i x_i}{\sum_1^n y_i} = \frac{7.9}{9.9} \approx 0.8 \]

The fact that the HT estimators can be calculated after the collection of a sample is significant because it means that data researchers can account for missing data without knowing ahead of time which data is missing.

Some cases may not need to take data

Going outside the data frame (we would not know about it)
Ch. 2
Exercises

1. a. \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{10}{0 + 1 + 1 + 1 + 2 + 1 + 0 + 1 + 0} = 1 \)
   \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} (0 + 1 + 1 + 1 + 2 + 1 + 0 + 1 + 0) \)
   \( = 1 \)
   \( \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i = \frac{100}{0 + 1 + 1 + 1 + 2 + 1 + 0 + 1 + 0} = 100 \)
   \( \text{Vár (x)} = n (n - n) \frac{s^2}{n} = \frac{100}{100 - 10} (1.56) \approx 1.40 \)
   \( s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{9} [(-1)^2 + (-2)^2 + (0)^2 + (-1)^2]
   \approx 1.44 \)
   \( \text{Vár (y)} = \frac{100}{100 - 10} (1.51) \approx 13.59 \)

b. \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} (0 + 1 + 1 + 1 + 2 + 1 + 0 + 1 + 0) = 1.2 \)
   \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} (0 + 1 + 1 + 1 + 2 + 1 + 0 + 1 + 0) = 1.2 \)
   \( \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i = \frac{100}{0 + 1 + 1 + 1 + 2 + 1 + 0 + 1 + 0} = 100 \)
   \( \text{Vár (x)} = n (n - n) \frac{s^2}{n} = \frac{100}{100 - 10} (1.56) \approx 1.40 \)
   \( s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{9} [(-1)^2 + (-2)^2 + (0)^2 + (-1)^2)
   \approx 1.51 \)
   \( \text{Vár (y)} = \frac{100}{100 - 10} (1.51) \approx 13.59 \)

2. a. \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} (2 + 3 + 4 + 4 + 5 + 6 + 7 + 8 + 9) = 3.1 \)
   \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} (2 + 5 + 1 + 4 + 1 + 3 + 2 + 5 + 2 + 3) = 3.1 \)
   \( \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i = \frac{1}{n} (1 + 2 + 3 + 6 + 1 + 4 + 9 + 1 + 8) \)
   \( = \frac{1}{n} (1 + 2 + 3 + 6 + 1 + 4 + 9 + 1 + 8) \)
   \( = 310 \)
   \( s^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{9} [(-1)^2 + (1)^2 + (1)^2 + (2)^2 + (1)^2 + (1)^2 + (3)^2 + (2)^2 + (1)^2]
   \approx 1.87 \)
   \( \text{Vár (x)} = n (n - n) \frac{s^2}{n} = \frac{100}{100 - 10} (1.87) \approx 168.3 \)
   \( \text{Vár (y)} = \frac{100}{100 - 10} (1.87) \approx 168.3 \)
   \( \text{Vár (z)} = \frac{100}{100 - 10} (1.87) \approx 168.3 \)

b. \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = 3.1 \)
   \( \text{Vár (y)} = \frac{100}{100 - 10} (1.87) \approx 168.3 \)

C. \( \frac{10}{100} \times 100\% = 10% \)
   \( \text{Vár (y)} = \frac{100}{100} (100 - 10) = 1.73 \times 10^3 \) different samples
   probability of sample in fact (a): \( 1.73 \times 10^3 \)
\[ E(\bar{Y}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_i - \mu) \]

Population mean: \( \mu = 1.6 \)

So \( \bar{Y} \) is unbiased for the population mean in this problem.

The estimator of \( \sigma^2 \) is:

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

\[ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \]
5a. Problems establishing a frame include selecting the boundaries for the study area and selecting the size of the units because both are in some sense arbitrary decisions and can affect the results of the study.

b. I chose to study how many McDonald's can be found within a 9-square-mile square of land area throughout the metro Detroit area. To do so, I Googled "McDonald's" on Google Maps, took a screen shot with Bloomfield Twp as the NW boundary, Charter Twp of Clinton in the NE, Dearborn Heights in the SW, and the city of Detroit in the SE. I then overlaid a grid of ~9-square-mile blocks. The population area has a width of eight such blocks and a height of six, so that 48 units comprise the population. I sampled 8 of these: (4,2), (13,1), (16,2), (20,2), (30,3), (31,2), (35,3), (43,3)

c. Special problems include: the dots the represent McDonald's locations are sometime shared evenly between two units, requiring a coin toss for which one belongs in, and Detroit is bordered by a body of water to the northeast, so that parts of some eastern units cannot contain McDonald's.

\[
\begin{align*}
\bar{y} &= \frac{1}{8} \sum_{i=1}^{8} y_i = \frac{1}{8} (2+1+2+2+3+2+3+3) = 2.25 \\
S^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
&= \frac{1}{7} [(-2.25)^2 + (-1.25)^2 + (-2.25)^2 + (-2.25)^2 + (-2.25)^2 + (1.75)^2 + (1.75)^2] = 0.5 \\
\hat{\sigma} &= \sqrt{\frac{S}{n}} = \sqrt{\frac{0.5}{48}} = 0.0852 \\
\text{Var}(\hat{\sigma}) &= \frac{n-1}{n} \frac{S^2}{n} = \frac{48}{48} \left( \frac{S^2}{n} \right) = 0.52
\end{align*}
\]
I would improve the survey procedure by using a broader definition of southeast Detroit. Instead of arbitrarily choosing only cities that fit into a rectangle with Detroit in the SE corner, I would expand the boundaries to include all cities that constitute metro Detroit.

6. a. \( \bar{y} \approx 30.17 \)
   \( \text{var}(\bar{y}) = 17.98 \)
   \( SD(\bar{y}) \approx 4.24 \approx 4.28 = \text{HSE} \)

   b. \( \bar{y} \approx 30.15 \)
   \( \text{var}(\bar{y}) = 9.22 \)
   \( SD(\bar{y}) \approx 3.03 \approx 3.05 = \text{HSE} \)
Histogram of ybar

n = 10
Histogram of ybar

\( n = 15 \)