1. According to the U.S. Census Bureau, the distribution of family sizes in the United States are as shown on the accompanying table.

<table>
<thead>
<tr>
<th>Number of persons</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of families</td>
<td>.25</td>
<td>.32</td>
<td>.17</td>
<td>.16</td>
<td>.07</td>
<td>.02</td>
<td>.01</td>
</tr>
</tbody>
</table>

That is, 32% of the families in the United States have 2 members, while only 2% have 6 members. (Families with more than 7 children are very rare.)

(a) Find the mean family size, approximately, from this distribution of family sizes. Will this approximation be too large or too small? Explain.

(b) Find the approximate standard deviation of family sizes.

(c) Suppose Nielsen randomly selects 400 families from this population. Describe, as closely as you can, the shape, center, and spread of the 400 data values that might occur in the sample.

(d) Nielsen is actually interested in the mean number of persons per family in samples of 400 families. Describe, as closely as you can, the shape, center, and spread of the distribution of possible values of the sample mean in random samples of 400 families.

Solution

(a). \( \mu = Ex = \sum xp(x) = 1 \times 0.25 + 2 \times 0.32 + \cdots + 7 \times 0.01 = 2.58 \). This approximation is regular because only 3% have 6 members or more, and 90% have members between 1 and 4.

(b).

\[
\sum x^2p(x) = 1^2 \times 0.25 + 2^2 \times 0.32 + \cdots + 7^2 \times 0.01 = 8.58 \\
Var(x) = \sum x^2p(x) - \mu^2 = 8.58 - 6.6564 = 1.9236 \\
\sigma = \sqrt{Var(x)} = 1.39
\]

(c). Shape is close to the distribution of the population.

Center is approximately 2.58.

Spread: standard deviation is approximately 1.39.

(d). \( \bar{x} \) has approximately normal distribution with mean 2.58 and standard deviation \( \sigma/\sqrt{n} = 1.39/\sqrt{20} = 0.0695 \) (noting \( (N - n)/(N - 1) \approx 1.)

2. List all possible simple random samples of size \( n = 2 \) that can be selected from the population \{0, 1, 2, 3, 4\}. Calculate \( \sigma^2 \) for the population and \( V(\bar{y}) \) for the sample.

Solution

possible samples: \{0, 1\} \{0, 2\} \{0, 3\} \{0, 4\} \{1, 2\} \{1, 3\}
sample mean $\bar{y}$: 0.5, 1, 1.5, 2, 1.5, 2
possible samples: \{1, 4\} \{2, 3\} \{2, 4\} \{3, 4\}
sample mean $\bar{y}$: 2.5, 2.5, 3, 3.5
The probability distribution of sample mean $\bar{y}$ is
$\bar{y}$: 0.5, 1, 1.5, 2, 2.5, 3, 3.5
\[p(\bar{y}): 0.1, 0.1, 0.2, 0.2, 0.2, 0.1, 0.1\]

\[E \bar{y} = \sum \bar{y}p(\bar{y}) = 0.5 \times 0.1 + 1 \times 0.1 + \cdots + 3.5 \times 0.1 = 2\]
\[E(\bar{y})^2 = \sum (\bar{y})^2p(\bar{y}) = 0.5^2 \times 0.1 + 1^2 \times 0.1 + \cdots + 3.5^2 \times 0.1 = 4.75\]
\[Var(\bar{y}) = E(\bar{y})^2 - (E\bar{y})^2 = 4.75 - 4 = 0.75\]

Method 2 The probability distribution of $y$ is
$y$: 0, 1, 2, 3, 4
\[p(y): 0.2, 0.2, 0.2, 0.2, 0.2\]
\[Ey = \sum yp(y) = 0 \times 0.2 + 1 \times 0.2 + \cdots + 4 \times 0.2 = 2\]
\[Ey^2 = \sum y^2p(y) = 0^2 \times 0.2 + 1^2 \times 0.2 + \cdots + 4^2 \times 0.2 = 6\]
\[\sigma^2 = Ey^2 - (Ey)^2 = 6 - 4 = 2\]
\[Var(\bar{y}) = \frac{N-n}{N-1} \sigma^2 = \frac{5-2}{5-1} \times 2 = 3/4 = 0.75\]

3. For the simple random samples generated in the above one, calculate $s^2$ for each sample. Show numerically that $Es^2 = \frac{N}{N-1} \sigma^2$. How to prove $Es^2 = \frac{N}{N-1} \sigma^2$?

Solution
possible samples: {0, 1} {0, 2} {0, 3} {0, 4} {1, 2} {1, 3}
sample variance $s^2$: 0.5, 2.0, 4.5, 8, 0.5, 2.0
possible samples: {1, 4} {2, 3} {2, 4} {3, 4}
sample variance $s^2$: 4.5, 0.5, 2.0, 0.5
where $s^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$.
The probability distribution of sample variance $s^2$ is
\[s^2: 0.5, 2.0, 4.5, 8.0\]
\[p(s^2): 0.4, 0.3, 0.2, 0.1\]
\[Es^2 = \sum s^2p(s^2) = 0.5 \times 0.4 + \cdots + 8.0 \times 0.1 = 2.5\]
But \[Es^2 = \frac{N}{N-1} \sigma^2 = \frac{5}{4} \times 2 = 2.5\]

4. State park officials were interested in the proportion of campers who consider the campsite spacing adequate in a particular campground. They decided to take a simple random sample of $n = 30$ from the first $N = 300$ camping parties that visit the campground. Let $y_i = 0$ if the head of the $i$th party sampled does not think the campsite spacing is adequate and $y_i = 1$ if he does ($i = 1, \cdots, 30$). Now suppose we have $\sum_{i=1}^{30} y_i = 25$. Please estimate $p$, the proportion of campers who consider the campsite spacing adequate. Place a bound on the error of estimation. ($\alpha = 5\%$)

Solution
\[\hat{p} = \frac{\sum y_i}{n} = \frac{25}{30} = \frac{5}{6}\]
Bound on the error of estimation: \( z_{\alpha/2} \sqrt{\text{Var}(\bar{y})} \), \( \alpha = 5\% \), \( z_{0.05} = 1.96 \approx 2 \).

\[
B = 2 \sqrt{\text{Var}(\bar{y})} = 2 \sqrt{\frac{s^2 (N - n)}{n - 1}} \\
= 2 \sqrt{\frac{\hat{p}\hat{q} (N - n)}{n - 1}} \\
= 2 \sqrt{\frac{\frac{5}{6} \times \frac{1}{6} 300 - 30}{300}} \\
= 0.131
\]

\( \hat{\text{Var}}(\bar{y}) = \frac{s^2 (N - n)}{n} \) is an unbiased estimator of \( \text{Var}(\bar{y}) = \frac{\sigma^2 (N - n)}{n-1} \).

\[
s^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{n}{n - 1} (\bar{y} - \bar{y})^2 = \frac{n}{n - 1} \hat{p}\hat{q}
\]