Probability & Statistics

The Mean

Population vs. Sample

• A **population** is a set of all elements about which information is desired.
• A **sample** is a subset of a population that is analyzed in an attempt to estimate certain properties of the entire population.

Example

• A clothing manufacturer wants to know what style of jeans teens between 13 and 16 will buy. To help answer this question, 200 teens between 13 and 16 were surveyed.
• The **population** is all teens between 13 and 16.
• The **sample** is the 200 teens between 13 and 16 surveyed.

Statistic vs. Parameter

• A numerical descriptive measurement made on a sample is called a **statistic**.
• Such a measurement made on a population is called a **parameter** of the population.
• Since we cannot usually have access to entire populations, we rely on our experimental results to obtain statistics, and we attempt to use the statistics to estimate the parameters of the population.

Mean (Average) of a Sample

Let an experiment have as outcomes the numbers \( x_1, x_2, \ldots, x_r \) with frequencies \( f_1, f_2, \ldots, f_r \) respectively, so that \( f_1 + f_2 + \ldots + f_r = n \). Then the **sample mean** equals

\[
\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_r f_r}{n},
\]

or

\[
\bar{x} = \frac{x_1 f_1}{n} + \frac{x_2 f_2}{n} + \ldots + \frac{x_r f_r}{n}.
\]

Mean (Average) of a Population

If the population has \( x_1, x_2, \ldots, x_r \) with frequencies \( f_1, f_2, \ldots, f_r \) respectively. Then the **population mean** equals

\[
\mu = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_r f_r}{N},
\]

or

\[
\mu = \frac{x_1}{N} f_1 + \frac{x_2}{N} f_2 + \ldots + \frac{x_r}{N} f_r.
\]

Note: Greek letters are used for parameters.
An ecologist observes the life expectancy of a certain species of deer held in captivity. The table shows the data observed on a population of 1000 deer.

<table>
<thead>
<tr>
<th>Age at death (years)</th>
<th>Number observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>180</td>
</tr>
<tr>
<td>4</td>
<td>250</td>
</tr>
<tr>
<td>5</td>
<td>200</td>
</tr>
<tr>
<td>6</td>
<td>120</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>120</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
</tr>
</tbody>
</table>

What is the mean life expectancy of this population?

The expected value of the random variable $X$ which can take on the values $x_1, x_2, \ldots, x_N$ with $\Pr(X = x_i) = p_i, \Pr(X = x_j) = p_j, \ldots, \Pr(X = x_N) = p_N$ is

$$E(X) = x_1p_1 + x_2p_2 + \ldots + x_Np_N.$$
Example

• An Olympic gymnast received the following scores from six judges: 9.8, 9.8, 9.4, 9.2, 9.2, 9.0
  a.) Find the average score.
  b.) Create a relative frequency table.
  c.) Find the mean of the relative frequency distribution.