Probability & Statistics

Frequency & Probability Distributions

Our Goals

- Describe a given set of data in terms that allow for interpretation and comparison.
- Interpret data collected in a meaningful way and make decisions based on those interpretations.

The mathematical tools that allow us to do this belong to the branch of mathematics called statistics.

Statistics

- Statistics is the branch of mathematics that deals with data: their collection, description, analysis, and use in prediction.
- Data can be presented in raw form or organized and displayed in tables or charts.

Frequency Table

A table like the one below is called a frequency table since it presents the frequency with which each response occurs.

<table>
<thead>
<tr>
<th>Highest Degree Planned</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>68,600</td>
<td>23.7</td>
</tr>
<tr>
<td>Master's</td>
<td>118,675</td>
<td>41.0</td>
</tr>
<tr>
<td>Doctorate</td>
<td>50,365</td>
<td>17.4</td>
</tr>
<tr>
<td>Medical</td>
<td>26,919</td>
<td>9.3</td>
</tr>
<tr>
<td>Law</td>
<td>14,183</td>
<td>4.9</td>
</tr>
<tr>
<td>Other</td>
<td>10,710</td>
<td>3.7</td>
</tr>
<tr>
<td>Total</td>
<td>289,452</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Example: Frequency Table & Histogram

The grades for the first quiz in a class of 25 students are

8 7 6 10 5 10 7 1 8 0 7 0 8 10 5 9 3 8 6 10 4 9 10 9 5

(a) Organize the data into a frequency table.
(b) Create a histogram for the data.

Histogram

- When the data is numeric data, then it can be represented by a histogram which is similar to a bar chart but there is no space between the bars.
- Vertical axis: represents frequency, proportion, or probability
- Horizontal axis: represents numeric data
**Frequency Distribution & Relative Frequency Distribution**

- A table that includes every possible value of a statistical variable with its number of occurrences is called a *frequency distribution*.
- If instead of recording the number of occurrences, the proportion of occurrences are recorded, the table is called a *relative frequency distribution*.

**Example**

- The number of cars waiting to be served at a gas station was counted at the beginning of every minute during the morning rush hour and the frequency tabulated in the table at right.
- Determine the relative frequency distribution associated with this data and draw the histogram.

<table>
<thead>
<tr>
<th>Number of Cars Waiting</th>
<th>Number of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

**Probability Distributions**

- Generally, frequency distributions and relative frequency distributions are created from actual experiments where collections of sample data were obtained by actually recording the outcomes of experiments.
- Probability distributions are theoretical models of experiments.

**Histogram of a Probability Distribution**

- The histogram for a probability distribution is constructed in the same way as the histogram for a relative frequency distribution. Each outcome is represented on the number line, and above each outcome we erect a rectangle of width 1 and of height equal to the probability corresponding to that outcome.
Probability of an Event in Histogram

In a histogram of a probability distribution, the probability of an event \( E \) is the sum of the areas of the rectangles corresponding to the outcomes in \( E \).

Example

- Construct the probability distribution for the experiment in which a coin is tossed five times and the number of occurrences of heads is recorded.
- Construct the histogram for this probability distribution.
- Shade the histogram to represent the probability of the event “at least 3 heads”.

Random Variable

Consider a theoretical experiment with numerical outcomes. Denote the outcome of the experiment by the letter \( X \).

Since the values of \( X \) are determined by the unpredictable random outcomes of the experiment, \( X \) is called a random variable.

The probability distribution of \( X \) is a table listing the various values of \( X \) and their associated probabilities \( p_i \) with \( p_1 + p_2 + \ldots + p_r = 1 \).

Example

- Construct the probability distribution for the experiment in which a coin is tossed five times and the number of occurrences of heads is recorded.
- Construct the histogram for this probability distribution.
- Shade the histogram to represent the probability of the event “at least 3 heads”.

Example

- Consider a child’s toy box that contains 5 red cars and 2 blue cars. A sample of three cars is chosen at random from the toy box. Let \( X \) denote the number of blue cars in the sample. Find the probability distribution of \( X \).