Section 8.3: Variance and Standard Deviation

Example: Find the mean and median of these data sets.

A) 40, 40, 40, 100, 100, 100

\[ \text{Mean} = 70 \]
\[ \text{Median} = 70 \]

B) 60, 65, 65, 70, 75, 75, 80

Definition: The **standard deviation** of a data set is the measure of how the data is spread about its mean.

The **variance** of a data set is the average of the square of the distance from the data value to the mean.

\[ \sqrt{\text{Var}} = \text{st. dev.} \]

Notation:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_x$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>st. deviation</td>
<td>$\mu$</td>
</tr>
</tbody>
</table>

\[ \text{Variance} = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_n - \mu)^2}{n} \]

Sample Variance formula:

\[ \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n - 1} \]
Example: Compute the standard deviation and the variance.

<table>
<thead>
<tr>
<th>data</th>
<th>freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
</tr>
</tbody>
</table>

\[ \sigma = 3.142337 \]

\[ \text{Var} = (3.142337)^2 = 9.8737 \]

Example: Compute the standard deviation and the variance. Let \( X \) = the number of Dr. Peppers drank during a semester.

<table>
<thead>
<tr>
<th>X</th>
<th>freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>43</td>
<td>49</td>
</tr>
<tr>
<td>26</td>
<td>73</td>
</tr>
<tr>
<td>250</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ S_X = 19.55819 \]

\[ \text{Var} = (S_X)^2 = 382.522796 \]

Example: Compute the standard deviation.

<table>
<thead>
<tr>
<th>data</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
</tr>
<tr>
<td>8</td>
<td>0.4</td>
</tr>
<tr>
<td>10</td>
<td>0.1</td>
</tr>
</tbody>
</table>

\[ \sigma = 3.41906 \]
Example: Let X be a discrete random variable such that $0 \leq X \leq 99$. The random variable has an expected value of $15.3$ and standard deviation of $3.5036$.

A) What values of $X$ are within $1.25$ standard deviations of the mean?

$$
\mu - 1.25\sigma = 15.3 - 1.25(3.5036) = 10.6455
$$

$$
\mu + 1.25\sigma = 15.3 + 1.25(3.5036) = 19.75
$$

Answer: 11, 12, 13, ..., 19

B) What values of $X$ are up to $2$ standard deviations above the mean?

$$
\mu + 2\sigma = 22.47
$$

Answer: 0, 1, 2, ..., 22

---

Chebyshev’s Inequality: Let $X$ be a random variable with mean $\mu$ and standard deviation $\sigma$. Then the probability that $X$ will be within $k$ standard deviations of the mean is

$$
P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}
$$

Example: The random variable $X$ has a mean of 50 and a standard deviation of 14. Estimate

$$
P(30 \leq X \leq 70).
$$

$$
P(30 \leq X \leq 70) \geq 1 - \frac{1}{(\frac{14}{5})^2} = 0.51
$$

$$
50 + 14k = 70
$$

$$
14k = 20
$$

$$
k = \frac{20}{14} = \frac{10}{7}
$$

at least 0.51
Example: The expected lifetime of a product is 2 years with a standard deviation of 3.5 months. 
For a shipment of 5000 items, estimate the number of item that will last between 17.7 months and 30.3 months.

\[
P(17.7 \leq x \leq 30.3) \geq 1 - \frac{1}{1.8^2} = 0.6914
\]

\[
30.3 = 24 + 3.5k
\]
\[
6.3 = 3.5k
\]
\[
k = \frac{6.3}{3.5} = 1.8
\]

\[
5000(0.6914) = 3457
\]

at least 3457 items