Exercise 7.38

Burgers. Fast food is often considered unhealthy because much of it is high in both fat and calories. But are the two related? Here are the fat contents and calories of several brands of burgers. Analyze the association between fat content and calories.

<table>
<thead>
<tr>
<th>Fat (g)</th>
<th>19</th>
<th>31</th>
<th>34</th>
<th>35</th>
<th>39</th>
<th>39</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calories</td>
<td>410</td>
<td>580</td>
<td>580</td>
<td>570</td>
<td>640</td>
<td>680</td>
<td>660</td>
</tr>
</tbody>
</table>

Let us first plot the data.

Y-axis?
X-axis?

So, there seems to be a correlation?

The formula given in class notes follows:

\[ r = \frac{\sum z_x z_y}{n-1} \]

Alternatively (formula at the bottom of page 172 in textbook):

\[ r = \frac{\sum (x - \overline{x})(y - \overline{y})}{(n-1)s_x s_y} \]
Before calculating the correlation coefficient we need to find the mean and standard deviation of \( x \) (fat) and \( y \) (calories).

From what we learn in chapter 4 we have:

\[
\bar{x} =
\]

\[
\sigma_x =
\]

and

\[
\bar{y} =
\]

\[
\sigma_y =
\]

<table>
<thead>
<tr>
<th>( x )</th>
<th>( y )</th>
<th>( x - \bar{x} )</th>
<th>( y - \bar{y} )</th>
<th>((x - \bar{x})(y - \bar{y}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>410</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>580</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>590</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>570</td>
<td>0.7</td>
<td>-20.0</td>
<td>-14.3</td>
</tr>
<tr>
<td>39</td>
<td>640</td>
<td>4.7</td>
<td>50.0</td>
<td>235.7</td>
</tr>
<tr>
<td>39</td>
<td>680</td>
<td>4.7</td>
<td>90.0</td>
<td>424.3</td>
</tr>
<tr>
<td>43</td>
<td>660</td>
<td>8.7</td>
<td>70.0</td>
<td>610.0</td>
</tr>
</tbody>
</table>

\[ \sum (x - \bar{x})(y - \bar{y}) = 4040.0 \]

Thus, the correlation coefficient between “fat” and “calories” is given by

\[
r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1) \sigma_x \sigma_y}
\]

\[ = \]

\[ = \]

A correlation coefficient of 0.96 indicates a

Since the coefficient is positive, when fat increases, the calories intake
Exercise 7.31
The scatterplot of the housing cost index versus the median family income for 10 states is shown below. The correlation is 0.625.

![Scatterplot](image)

Describe the relationship between median family income and housing cost index.

If we standardized both variables what would the correlation coefficient between the standardized variables be?

If we had measured the family income in thousand of dollars instead of dollars, how would the correlation coefficient change?

Another state has a housing cost index of 400 and a median family income of $35,000. If this state was added to the data set, how would the correlation coefficient change?

Does the data provide proof that by raising the median income in a state, housing cost index will rise as a result?
Exercise 7.44
The number of flights by U.S. Airlines has grown rapidly. Here are the number of flights flown in each year from 1995 to 2005.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Flights</td>
<td>5.33</td>
<td>5.35</td>
<td>5.41</td>
<td>5.38</td>
<td>5.53</td>
<td>5.68</td>
<td>5.97</td>
<td>5.27</td>
<td>6.49</td>
<td>7.13</td>
<td>7.14</td>
</tr>
</tbody>
</table>

Find the correlation coefficient for this data. (Hint: \( \sum (x - \bar{x})(y - \bar{y}) = 19.63, \ s_x = 3.32, \ s_y = 0.72 \)

\[ r = \]

= 

Make a scatterplot for this data and describe the trend.

Note two reasons that the correlation we found above is not a suitable summary of the strength of the association. Can you explain any possible reasons for any outliers?