Tables & Graphs

Given a set of observations \{ x_1, x_2, ..., x_n \}, one can summarize using:

- time series plot
- stem and leaf \(5\) to \(20\) stems
- frequency table \(5\) to \(20\) classes, \(\approx \sqrt{(#\text{\ observations})}\)
- bar chart \(5\) to \(20\) classes
- histogram \(5\) to \(20\) classes
- pie chart \(3\) to \(8\) sectors
- pictogram (or other methods) – is the scale by length, area or volume? [poor choice!]

Example 2.01 Course text (Devore, seventh edition), ex. 1.2 p. 23 q. 24, modified

[This data set is available from the course web site, at www.engr.mun.ca/~ggeorge/3423/Minitab/s01DescStat/index.html]

The data set below consists of observations on shear strengths \(x\) (in pounds) of ultrasonic spot welds made on a certain type of alclad sheet.

5434 4948 4521 4570 4990 5241 5112 5015 4659 4806 4637 5670 4381 4820 5043 4886 4599 5288 5299 4848 5378 5260 5055 5828 5218 4859 4780 5027 5008 4609 4772 5133 5095 4618 4848 5089 5518 5333 5164 5342 5069 4755 4925 5001 4803 4951 5679 5256 5207 5621 4918 5138 4786 4500 5461 5049 4974 4592 4173 5296 4965 5170 4740 5173 4568 5653 5078 4900 4968 5248 5245 4723 5275 5419 5205 4452 5227 5555 5388 5498 4681 5076 4774 4931 4493 5309 5582 4308 4823 4417 5364 5640 5069 5188 5764 5273 5042 5189 4986

(a) Produce a stem and leaf display of the data.
(b) Construct a bar chart of the data, using ten class intervals of equal width, with the first interval having lower limit 4000 (inclusive) and upper limit 4200 (exclusive). [Such a bar chart will be consistent with one that appeared in the paper “Comparison of Properties of Joints Prepared by Ultrasonic Welding and Other Means”, J. Aircraft, 1983, pp. 552-556.]
By itself, this table is not very helpful as we try to grasp the overall picture of shear strengths. One way to improve visibility is simply to rearrange these data into ascending order:

4173 4308 4381 4417 4452 4493 4500 4521 4568 4570 4592 4599
4609 4618 4637 4659 4681 4723 4740 4755 4772 4774 4780 4786
4803 4806 4820 4823 4848 4848 4859 4886 4900 4918 4925 4931
4948 4951 4965 4968 4974 4986 4990 5001 5008 5015 5027 5042
5043 5049 5055 5069 5069 5076 5078 5089 5095 5112 5133 5138
5164 5170 5173 5188 5189 5205 5207 5218 5227 5241 5245 5248
5256 5260 5273 5275 5286 5296 5299 5309 5333 5342 5364 5378
5388 5419 5434 5461 5498 5518 5555 5582 5621 5640 5653 5670
5679 5702 5764 5828

An additional improvement to the visual appearance is the **stem and leaf** display. Part of the output from a **MINITAB** session, using default values, is reproduced on the left hand side below. The left-most column is a cumulative frequency count from the nearer end. Note how **MINITAB** returns only the thousands and hundreds digits in the stem and the tens digit in the leaf. The units digit is truncated (lost altogether). On the right hand side (to the right of the comment markers ###) is shown a manual version that retains both digits of the leaf.

---

Stem-and-leaf of Shear st  N = 100
Leaf Unit = 10

### Manual version, retaining both digits of the leaf:

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41 7</td>
</tr>
<tr>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>3</td>
<td>43 08</td>
</tr>
<tr>
<td>6</td>
<td>44 159</td>
</tr>
<tr>
<td>12</td>
<td>45 026799</td>
</tr>
<tr>
<td>17</td>
<td>46 01358</td>
</tr>
<tr>
<td>24</td>
<td>47 2457788</td>
</tr>
<tr>
<td>32</td>
<td>48 00224458</td>
</tr>
<tr>
<td>43</td>
<td>49 01234566789</td>
</tr>
<tr>
<td>(14)</td>
<td>50 00124445667789</td>
</tr>
<tr>
<td>43</td>
<td>51 13367788</td>
</tr>
<tr>
<td>35</td>
<td>52 00124445677899</td>
</tr>
<tr>
<td>21</td>
<td>53 034678</td>
</tr>
<tr>
<td>15</td>
<td>54 1369</td>
</tr>
<tr>
<td>11</td>
<td>55 158</td>
</tr>
<tr>
<td>8</td>
<td>56 24577</td>
</tr>
<tr>
<td>3</td>
<td>57 06</td>
</tr>
<tr>
<td>1</td>
<td>58 2</td>
</tr>
</tbody>
</table>

The "(14)" means that that stem has 14 leaves, *including the median.*
With the "Increment" option of "Graph > Stem-and-leaf" set to 200 instead of the default value of 100, the number of stems is reduced to the appropriate number, namely \(\sqrt{100} = 10\). MINITAB's output is then

```
Stem-and-leaf of Shear st  N = 100
Leaf Unit = 100

1  4 1
3  4 33
12 4 4445555
24 4 6666777777
43 4 8888889999999999
(22) 5 000000000000111111
35 5 222222222222333333
15 5 444555
8  5 666677
1  5 8
```

Notice how the leaf unit is now 100, not 10, so that the last two digits of each value are now lost. The median class is \([5000, 5200)\) and contains 22 values, as indicated by the "(22)" in the first column.

From this stem and leaf display it is easy to generate a frequency table for shear stress manually, in the form required by part (b) of the question.

<table>
<thead>
<tr>
<th>Interval for (x)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4000 \leq x &lt; 4200)</td>
<td>1</td>
</tr>
<tr>
<td>(4200 \leq x &lt; 4400)</td>
<td>2</td>
</tr>
<tr>
<td>(4400 \leq x &lt; 4600)</td>
<td>9</td>
</tr>
<tr>
<td>(4600 \leq x &lt; 4800)</td>
<td>12</td>
</tr>
<tr>
<td>(4800 \leq x &lt; 5000)</td>
<td>19</td>
</tr>
<tr>
<td>(5000 \leq x &lt; 5200)</td>
<td>22</td>
</tr>
<tr>
<td>(5200 \leq x &lt; 5400)</td>
<td>20</td>
</tr>
<tr>
<td>(5400 \leq x &lt; 5600)</td>
<td>7</td>
</tr>
<tr>
<td>(5600 \leq x &lt; 5800)</td>
<td>7</td>
</tr>
<tr>
<td>(5800 \leq x &lt; 6000)</td>
<td>1</td>
</tr>
</tbody>
</table>

Total: 100
The bar chart generated by MINITAB (which it calls an "histogram") also provides the frequency table:

---

### Devore Exercise 1-24,

with 10 classes

---

There is a subtle difference between a “bar chart” and an “histogram”. A **bar chart** is used for **discrete** (countable) data (such as “number of defective items found in one run of a process”) or nonnumeric data (such as “engineering discipline chosen by students”). The bars are drawn with arbitrary (often equal) width. No two bars should touch each other. The height of each bar is proportional to the frequency.

An **histogram** is used for **continuous** data (such as “shear stress” or “weight” or “time”, where between any two possible values another possible value can always be found). [An histogram can also be used for discrete data.] Each bar covers a continuous interval of values and just touches its neighbouring bars without overlapping. Every possible value lies in exactly one interval. Unlike a bar chart, it is the **area** of each bar that is proportional to the frequency in that interval. Only if all intervals are of equal width will the histogram have the same shape as the bar chart.
The relative frequency in an interval is the proportion of the total number of observations that fall inside that interval. A relative frequency histogram can then be generated, with bar height given by

\[
\text{Bar height} = \frac{\text{Relative Frequency}}{\text{Class Width}} = \frac{\text{Frequency}}{(\text{Total Freq.})(\text{Class Width})}
\]

The total area of all bars in a relative frequency histogram is always 1. In chapter 8 we will see that the relative frequency histogram is related to the graph of a probability density function, the total area under which is also 1.

For the example above, the total frequency is 100 and the class width is 200, so the height of each bar in the relative frequency histogram is given by

\[
\text{Bar height} = \frac{\text{Frequency}}{100 \times 200} = \frac{\text{Frequency}}{20000}
\]

The cumulative frequency is the sum of all frequencies up to and including the current class.

Extending the previous table:

<table>
<thead>
<tr>
<th>Interval for (x)</th>
<th>Frequency (f)</th>
<th>Relative Frequency (r)</th>
<th>Height of histogram bar</th>
<th>Cumulative Frequency (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4000 \leq x &lt; 4200)</td>
<td>1</td>
<td>.01</td>
<td>.00005</td>
<td>1</td>
</tr>
<tr>
<td>(4200 \leq x &lt; 4400)</td>
<td>2</td>
<td>.02</td>
<td>.00010</td>
<td>3</td>
</tr>
<tr>
<td>(4400 \leq x &lt; 4600)</td>
<td>9</td>
<td>.09</td>
<td>.00045</td>
<td>12</td>
</tr>
<tr>
<td>(4600 \leq x &lt; 4800)</td>
<td>12</td>
<td>.12</td>
<td>.00060</td>
<td>24</td>
</tr>
<tr>
<td>(4800 \leq x &lt; 5000)</td>
<td>19</td>
<td>.19</td>
<td>.00095</td>
<td>43</td>
</tr>
<tr>
<td>(5000 \leq x &lt; 5200)</td>
<td>22</td>
<td>.22</td>
<td>.00110</td>
<td>65</td>
</tr>
<tr>
<td>(5200 \leq x &lt; 5400)</td>
<td>20</td>
<td>.20</td>
<td>.00100</td>
<td>85</td>
</tr>
<tr>
<td>(5400 \leq x &lt; 5600)</td>
<td>7</td>
<td>.07</td>
<td>.00035</td>
<td>92</td>
</tr>
<tr>
<td>(5600 \leq x &lt; 5800)</td>
<td>7</td>
<td>.07</td>
<td>.00035</td>
<td>99</td>
</tr>
<tr>
<td>(5800 \leq x &lt; 6000)</td>
<td>1</td>
<td>.01</td>
<td>.00005</td>
<td>100</td>
</tr>
</tbody>
</table>

Total: 100 1.00

The relative frequency histogram is on the next page.
Relative frequency histogram for the set of 100 observations of shear strengths (in pounds) of ultrasonic spot welds made on a certain type of alclad sheet.

Devore Exercise 1-24, with 10 classes

From this diagram, the relative frequency of any class can be recovered by calculating the area of the bar. For example, the relative frequency of the class $4800 \leq x < 5000$ is given by

$$\text{relative frequency} = \text{area of bar} = 200 \times 0.00095 = 0.19$$

**Therefore 19\% of the 100 data values are in the interval [4800, 5000).**

If you are absent from the first Minitab tutorial, then view the web page [www.engr.mun.ca/~ggeorge/3423/Minitab/s01DescStat/index.html](http://www.engr.mun.ca/~ggeorge/3423/Minitab/s01DescStat/index.html) carefully.
**Measures of Location**

The **mode** is the most common value.

In example 2.01 the mode is $\text{4848 and 5069 (each occurs twice)}$

From the frequency table, the modal class is $\text{5000} < x < \text{5200 (occurs 22 times)}$

A disadvantage of the mode as a measure of location is that it is not necessarily unique. 
For ungrouped continuous data it is not even well defined.

The **sample median** $\bar{x}$ (or the population median $\mu$) is the “halfway value” in an ordered set.

For $n$ data, the median is the $(n+1)/2$ th value if $n$ is odd.
The median is the semi-sum of the two central values if $n$ is even, 
(that is median = $\left(\frac{n}{2} \text{ th value} + ((n/2 + 1)\text{th value})}{2}\right)$.

For the example above, $n = \text{100 (even)} \Rightarrow n/2 = 50$

sample median $\bar{x} = \frac{x_{50} + x_{51}}{2} = \frac{5049 + 5055}{2} = \text{5052}$

In the table of grouped values, the 50th and 51st values fall in the same class.
The median class is therefore $\text{5000} < x < \text{5200 (same as the modal class)}$.

The **sample arithmetic mean** $\bar{x}$ (or the population mean $\mu$) is the ratio of the sum of the observations to the number of observations.

From individual observations, 
$$\bar{x} = \frac{\sum x}{n} \quad \text{(sample)}; \quad \mu = \frac{\sum x}{N} \quad \text{(pop'ln)}$$

and from a frequency table, 
$$\bar{x} = \frac{\sum f \cdot x}{\sum f}$$

For the example above, from the 100 raw data (not from the frequency table), 
$$\bar{x} = \frac{504916}{100} = \text{5049.16}$$
The relative advantages of the mean and the median can be seen from a pair of smaller samples.

**Example 2.02**

Let $A = \{1, 2, 3, 4, 5\}$ and $B = \{1, 2, 3, 23654, 5\}$.

_sorted into order:_ $B = \{1, 2, 3, 5, 23654\}$.

Then

$$\bar{x} = 3 \quad \text{for set } A \quad \text{and} \quad \bar{x} = 3 \quad \text{for set } B,$$

while

$$\bar{x} = \frac{1+2+3+4+5}{5} = 3 \quad \text{for set } A \quad \text{and} \quad \bar{x} = \frac{1+2+3+23654+5}{5} = 4733 \quad \text{for set } B.$$

Note that the mode is not well defined for either set.

A disadvantage of the mean as a measure of location is that it is very sensitive to outliers (extreme values).

Advantages of the mean over the median include

- the median uses only the central value(s) while the mean uses all values.
- calculus methods work much better with the mean.

For a symmetric population, the mean $\mu$ and the median $\tilde{\mu}$ will be equal. If the mode is unique, then it will also be equal to the mean and median of a symmetric population.
Measures of Variation

The simplest measure of variation is the **range** = (largest value − smallest value).

A disadvantage of the sample range is

> **it often increases as \( n \)** increases.

A disadvantage of the population range is

> **it may be infinite.**

The effect of outliers can be eliminated by using the distance between the **quartiles** of the data as a measure of spread instead of the full range.

The **lower quartile** \( q_L \) is the \( \{ (n+1) / 4 \} \) th smallest value.
The **upper quartile** \( q_U \) is the \( \{ 3(n+1) / 4 \} \) th smallest value.

[Close relatives of the quartiles are the **fourths.**
The lower fourth is the median of the lower half of the data, (including the median if and only if the number \( n \) of data is odd).
The upper fourth is the median of the upper half of the data, (including the median if and only if the number \( n \) of data is odd).
In practice there is often little or no difference between the value of a quartile and the value of the corresponding fourth.]

The **interquartile range** is \( IQR = q_U - q_L \) and
the **semi-interquartile range** is \( SIQR = (q_U - q_L) / 2 \)

**Example 2.01:**

\( n = 100 \) \( \Rightarrow \) \( (n+1) / 4 = 25.25 \) \( \Rightarrow \) \( q_L = \) value 1/4 of the way from \( x_{25} \) to \( x_{26} \)

\[
q_L = \frac{3x_{25} + x_{26}}{4} = \frac{3 \times 4803 + 4806}{4} = 4803.75
\]

and \( 3 (n+1) / 4 = 75.75 \) \( \Rightarrow \) \( q_U = \) value 3/4 of the way from \( x_{75} \) to \( x_{76} \)

\[
q_U = \frac{x_{75} + 3x_{76}}{4} = \frac{5273 + 3 \times 5275}{4} = 5274.50
\]

The semi-interquartile range is then \( \frac{5274.50 - 4803.75}{2} = 235.375 \).
The **boxplot** illustrates the median, quartiles, outliers and skewness in a compact visual form. The boxplot for example 2.01, as generated by an older version of MINITAB, is shown below. [See the tutorial session for a more modern version of this output.]

```
MTB > BoxPlot C1.
```

Unequal whisker lengths reveal skewness. The whiskers extend as far as the last observation before the inner fence. The fences are *not* plotted by MINITAB.

The inner fences are 1.5 interquartile ranges beyond the nearer quartile, at

\[ x_L - 1.5 \times IQR \text{ (lower)} \quad \text{and} \quad x_U + 1.5 \times IQR \text{ (upper)} \quad \text{[4097.625 and 5980.625 here]} \]

The outer fences are twice as far away from the nearer quartile, at

\[ x_L - 3 \times IQR \text{ (lower)} \quad \text{and} \quad x_U + 3 \times IQR \text{ (upper)} \quad \text{[3391.500 and 6686.750 here]} \]

Any observations between inner & outer fences are **mild outliers**, which would be indicated by an open circle (or, in MINITAB, by an asterisk). There are no outliers in this example.

Any observations beyond outer fences are **extreme outliers**, which would be indicated by a closed circle (or, in MINITAB, by an asterisk or a zero).

If you encounter an extreme outlier, then check if the measurement is incorrect or is from a different population. If the observation is genuine, then it is a rare event (< 0.01% in most populations).

Measures of variability based on quartiles are not easy to manipulate using calculus methods.
The deviation of the \( i \)th observation from the sample mean is \( (x_i - \bar{x}) \). At first sight, one might consider that the sum of all these deviations could serve as a measure of variability. However:

\[
\sum_{i=1}^{n}(x - \bar{x}) = \sum_{i=1}^{n}x - \bar{x}\sum_{i=1}^{n}1 = n\bar{x} - n\bar{x} = 0
\]

An alternative is the **mean absolute deviation from the mean**, defined as

\[
MAD = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|
\]

Unfortunately, the function \( f(x) = |x - \bar{x}| \) is not differentiable at the one point where the derivative is most needed, at \( x = \bar{x} \). Instead, the mean **square** deviation from the mean is used:

The **population variance** \( \sigma^2 \) for a finite population of \( N \) values is given by

\[
\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2
\]

and the **sample variance** \( s^2 \) of a sample of \( n \) values is given by

\[
s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

The square root of a variance is called the **standard deviation** and is **positive** (unless all values are exactly the same, in which case the standard deviation is zero). The reason for the different divisor \( (n - 1) \) in the expression for the sample variance \( s^2 \) will be explained later.

The **MINITAB** output for various summary statistics for example 2.01 is shown here:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>MEAN</th>
<th>MEDIAN</th>
<th>TRMEAN</th>
<th>STDEV</th>
<th>SEMEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>100</td>
<td>5049.2</td>
<td>5052.0</td>
<td>5050.5</td>
<td>351.5</td>
<td>35.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIN</td>
<td>MAX</td>
<td>Q1</td>
<td>Q3</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>4173.0</td>
<td>5828.0</td>
<td>4803.8</td>
<td>5274.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
When calculating a sample variance by hand or on some hand held calculators, one of the following shortcut formulæ may be easier to use:

\[ s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right)^2}{n - 1} \quad \text{or} \quad s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2}{n - 1} \quad \text{or} \]

\[ s^2 = \frac{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2}{n (n - 1)}. \]

For integer values of \( x \), the last of these three formulæ allows the sample variance to be expressed exactly as a fraction. The formulæ for data taken from a frequency table with \( m \) classes are similar:

\[ s^2 = \frac{1}{n - 1} \sum_{i=1}^{m} f_i \left( x_i - \bar{x} \right)^2 \quad \text{or} \quad s^2 = \frac{\sum_{i=1}^{m} f_i x_i^2 - \frac{1}{n} \left( \sum_{i=1}^{m} f_i x_i \right)^2}{n - 1} \]

or

\[ s^2 = \frac{\sum_{i=1}^{m} f_i x_i^2 - n \bar{x}^2}{n - 1} \quad \text{or} \quad s^2 = \frac{n \sum_{i=1}^{m} f_i x_i^2 - \left( \sum_{i=1}^{m} f_i x_i \right)^2}{n (n - 1)} \]

where, in each case, \( n = \sum_{i=1}^{m} f_i \) and \( \bar{x} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i} \).

However, all of the shortcut formulæ are more sensitive to round-off errors than the definition is.

**Example 2.03:**

Find the sample variance for the set \{ 100.01, 100.02, 100.03 \} by the definition and by one of the shortcut formulæ, in each case rounding every number that you encounter during your computations to six or seven significant figures, (so that 100.01 \( ^2 \) = 10002.00 to 7 s.f.). The correct value for \( s^2 \) in this case is .0001, but rounding errors will cause all three shortcut formulæ to return an incorrect value of zero. (Try it!).

\[ \Sigma x = 300.06 \Rightarrow (\Sigma x)^2 = 90036.00; \]

\[ \Sigma (x^2) = 10002.00 + 10004.00 + 10006.00 = 30012.00 \]

\[ \Rightarrow n \Sigma (x^2) - (\Sigma x)^2 = 90036.00 - 90036.00 = 0.00 ! \]
Example 2.04:

Find the sample mean and the sample standard deviation for \( x \) = the number of service calls during a warranty period, from the frequency table below.

<table>
<thead>
<tr>
<th>( x_i )</th>
<th>( f_i )</th>
<th>( f_i \cdot x_i )</th>
<th>( f_i \cdot x_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>65</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Sum:</td>
<td>100</td>
<td><strong>42</strong></td>
<td><strong>60</strong></td>
</tr>
</tbody>
</table>

[Note that the mode and median of \( x \) are both 0.]

\[
\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{42}{100} = 0.42
\]

\[
s^2 = \frac{n \sum f_i x_i^2 - \left( \sum f_i x_i \right)^2}{n(n-1)} = \frac{100 \times 60 - 42 \times 42}{100 \times 99} = \frac{4236}{9900} = 0.42787878\ldots
\]

or

\[
s^2 = \frac{1}{n-1} \sum f_i (x_i - \bar{x})^2 = \frac{65 \times (0 - 0.42)^2 + \cdots + 2 \times (3 - 0.42)^2}{99} = \cdots = 0.427878\ldots
\]

- tedious, but
- less sensitive to round-off errors
For any data set:

\[ \geq \frac{3}{4} \text{ of all data lie within two standard deviations of the mean.} \]
\[ \geq \frac{8}{9} \text{ of all data lie within three standard deviations of the mean.} \]
\[ \geq (1 - \frac{1}{k^2}) \text{ of all data lie within } k \text{ standard deviations of the mean (Chebyshev’s inequality).} \]

For a bell-shaped distribution (for which population mean = population median = population mode):

- \(~ 68\%\) of all data lie within one standard deviation of the mean.
- \(~ 95\%\) of all data lie within two standard deviations of the mean.
- \(> 99\%\) of all data lie within three standard deviations of the mean.

[Note that the points on the normal probability curve where \( x = \mu \pm 1\sigma \) are the curve’s points of inflection, where the concavity changes sign.]
Misleading Statistics - Example 2.05

Both graphs below are based on the same information, yet they seem to lead to different conclusions.

“Our profits rose enormously in the last quarter.” vs. “Our profits rose by only 10% in the last quarter.”

Visual displays can be very misleading. Questions to ask when viewing visual summaries of data include,

for graphs:

- Where is the zero?
- Are the scales appropriate?

for bar charts / pictograms:

- Is the frequency proportional to height, area or volume?

[End of the chapter “Descriptive Statistics”]