Homework assignments will be drawn from problems from the text, as well as from other sources. Some of these will involve data files. All files will be available on the Web site www.stern.nyu.edu/~gsimon/statdata.

1. Consider this set of eight values:

\[ 250 \ 206 \ 235 \ 211 \ 261 \ 208 \ 174 \ 214 \]

Find the mean, median, and standard deviation for this set.

SOLUTION: The total of the values is 1,759, so the average (mean) is \[ \bar{x} = \frac{1,759}{8} = 219.875 \]

The values, sorted into order, look like this:

\[ 174 \ 206 \ 208 \ 211 \ 214 \ 235 \ 250 \ 261 \]

The median value occupies positions 5 and 6, so we report \[ \bar{x} = x_{\text{median}} = \frac{211 + 214}{2} = 212.5. \]

There are several strategies to compute the standard deviation \[ s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2}{n} - \frac{\left( \sum x_i \right)^2}{n}}. \]

The tough part is the sum \[ \sum_{i=1}^{8} (x_i - \bar{x})^2. \] You can get directly the sum \[ (x_1 - \bar{x})^2 + \ldots + (x_8 - \bar{x})^2, \] but since \[ \bar{x} \] is given to six significant figures, this procedure looks very awkward. Moreover, it is error-prone. The best strategy uses the next result.

You can use \[ \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{\left( \sum x_i \right)^2}{n}. \] The hardest item here is \[ \sum x_i^2, \] but it’s still not bad. Find

\[ \sum x_i^2 = 250^2 + 206^2 + 235^2 + 211^2 + 261^2 + 208^2 + 174^2 + 214^2 = 392,139 \]

Then \[ \sum (x_i - \bar{x})^2 = 392,139 - \frac{(1,759)^2}{8} = 131,826 - \frac{3,094,081}{8} = \frac{1,252,625}{8} = 156,578.125. \]
We are able then to assemble \[ s = \sqrt{\frac{5,378.875}{8-1}} \approx \sqrt{768.410714} \approx 27.72. \]

You could bypass all this work and just enter the eight values into a data column in Minitab. Your sheet would look this this:

![Worksheet](image)

The command `Stat ➔ Basic Statistics ➔ Display Descriptive Statistics` would then give

**Descriptive Statistics: C1**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>8</td>
<td>0</td>
<td>219.88</td>
<td>9.80</td>
<td>27.72</td>
<td>174.00</td>
<td>206.50</td>
<td>212.50</td>
<td>246.25</td>
<td>261.00</td>
</tr>
</tbody>
</table>

2. Consider these values:  
11 17 18 10 22 23 15 17 14 13 10 12 18 18 11 14

Find the mean, median, and mode for these.

**SOLUTION:** There are 16 values, and these 16 values have a total of 243, so the mean (or average) is 243 ÷ 16 = 15.1875. If you sort the values into ascending order, you get

10 10 11 11 12 13 14 14 15 17 17 18 18 18 20 23

The median of the 16 values occurs between the 8th and 9th largest, at the position indicated by the arrow. We give 14.5 as the median.

The value 18 occurs three times, more often than any other value, so we report 18 as the mode.
3. A set of \( n = 80 \) values has average 14,880.16. After all the work is completed, you discover that a value originally recorded as 12,148 should have been 11,248. If you replace the value 12,148 by 11,248, what will be the corrected average?

SOLUTION: The original total was apparently \( 80 \times 14,880.16 = 1,190,412.8 \). If you remove the incorrect 12,148 and replace it by the correct 11,248, the total will change by \( 11,248 - 12,148 = -900 \). The corrected total is then \( 1,190,412.8 - 900 = 1,189,512.8 \), and the corrected average is \( \frac{1,189,512.8}{80} = 14,868.91 \).

There are other ways to think about this. For example, you could note that decreasing the total by 900 translates to decreasing the average by \( \frac{900}{80} = 11.25 \). The new average must be 14,880.16 - 11.25 = 14,868.91.

4. The file SWA\97EMPLOY.MTP contains five columns. (This file is available on the Stern Web site.) Data are numbers of employees for various US airlines in 1997; these are approximately the averages of the 12 monthly employee counts.

Variable names:

- AIRLINE
- FULLTIME
- PARTTIME
- TOTAL  Sum of FULLTIME + PARTTIME
- FTequiv  FULLTIME + 0.5 * PARTTIME
- TYPE  1=MAJOR, 2=NATIONAL, 3=LARGE REGIONAL, 4=MEDIUM REGIONAL

a. The values of FTequiv vary substantially according to TYPE. Produce a display in which there are four side-by-side boxplots, so that the size differences among the TYPES are displayed graphically. HINT: Use either **Graph ⇒ Boxplot** or **Stat ⇒ EDA ⇒ Boxplot**. Then ask for **With Groups**; in the **Graph variables** box, name FTequiv (or C5) and in the **Categorical Variables** box name Type (or C6).
b. The display created in part a will leave you with only one clear impression. We can see more if we utilize logarithms of FTequiv to create a new variable which is $\log_e(FTequiv)$. Use \texttt{Calc $\Rightarrow$ Calculator} to make $C7 = \log_e(FTequiv)$; the Minitab code for base-$e$ logarithms is LOGE. Now produce four side-by-side boxplots for the values of TYPE. Name this new variable as LogFTequiv.

NOTE: In most work we prefer base-$e$ logarithms, although some people find it easier to deal with base-10 logarithms. The notation $\ln$ (for “natural logarithm”) is often used, but most of the time we’ll just use log or perhaps $\log_e$.

You will find an unusual airline in the group TYPE=2. What is this airline?
HINT: Try \texttt{Editor $\Rightarrow$ Brush} on the graph. Please note that this calls for \texttt{Editor}, not \texttt{Edit}.

c. Produce a graph showing PARTTIME on the vertical axis and FULLTIME on the horizontal axis. Identify any airline which seems to have an unusual mix of part-time and full-time employees.

d. Part c might come out differently if you plot log(PARTTIME) versus log(FULLTIME). Continue to use base-$e$ logarithms. Do you still find the same unusual airline(s)?

NOTE: A number of the airlines have no part-time employees. Use the transformation LOGE(PARTTIME+0.5).

SOLUTION: For part a we do \texttt{Stat $\Rightarrow$ EDA $\Rightarrow$ Boxplot} and put C6 (or TYPE) in the X column. The result will be this:

![Boxplot of FTequiv vs TYPE](image)

This certainly succeeds in showing us that the major airlines have many more employees. Unfortunately, there is not enough resolution in the picture to make judgments about the other three groups.
Observe that the four TYPEs are cleanly separated. We see that the values in the TYPE=4 group (medium regional) spread out around $e^4 \approx 55$, the values in the TYPE=3 group (large regional) are spread around $e^6 \approx 400$, and the values in the TYPE=2 group (national) spread out around $e^{7.2} \approx 1,350$.

The low outlier in group TYPE=2 is point 36, Rich Airlines. Has anyone heard of this? Has anyone actually flown it?

For part c, we do **Graph ⇒ Scatterplot ⇒** and name PARTTIME as Y, FULLTIME as X. The resulting picture will be this:
This suggests one very unusual point. Using **Editor ⇒ Brush** on the graph, this can be identified as point 7, Federal Express. This is interesting. We see that Federal Express has a very large number of part-time people. This is, of course, *not* a passenger airline.

d. Using **Calc ⇒ Calculator** ⇒ create the logarithms of these two variables. Now the plot will be this:

![Scatterplot of logPARTTIME vs logFULLTIME](image)

Here Federal Express is still the point at the upper right, however it takes a very subtle eye to decide that it is unusual. Another point noted as possibly unusual is point 73, Eastwind Airlines. This airline has twice as many part-time as full-time people!

Besides the graphical methods, you might simply look at the obvious ratio \( \text{PARTTIME} \div \text{FULLTIME} \). Based on this, we would find

- There are 14 airlines with no part-time people at all.
- Beyond these, there are an additional 18 airlines with this ratio in the interval \((0, 0.05)\).
- There are three airlines for which this ratio exceeds 0.50. These are Executive (0.55), Federal Express (0.65), and Eastwind (1.97).
5. Indicate (without computation) which sample in each set has the higher standard deviation.

Set 1, Sample A: 86, 86, 86, 86, 86
Set 1, Sample B: 85, 86, 86, 86, 86

Set 2, Sample A: 20, 25, 25, 25, 30
Set 2, Sample B: 15, 25, 25, 25, 35

Set 3, Sample A: 20, 20, 30, 40, 40
Set 3, Sample B: 20, 25, 30, 35, 40

SOLUTION: The standard deviation is defined as 
\[ s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \]. Since all
these lists have \( n = 5 \), the comparisons can be based on 
\[ \sum_{i=1}^{5} (x_i - \bar{x})^2 \].

In Set 1, Sample A has \( \bar{x} = 86 \) and then
\[ \sum_{i=1}^{5} (x_i - \bar{x})^2 = 0. \] For Sample B,
\[ \sum_{i=1}^{5} (x_i - \bar{x})^2 > 0, \] so sample B has the larger standard deviation.

In Set 2, Sample A has \( \bar{x} = 25 \) and
\[ \sum_{i=1}^{5} (x_i - \bar{x})^2 = (-5)^2 + (5)^2 = 50. \] For Sample B,
\( \bar{x} = 25 \) and
\[ \sum_{i=1}^{5} (x_i - \bar{x})^2 = (-10)^2 + (10)^2 = 200. \] Thus Sample B has the larger standard
deviation.

In Set 3, Sample A has \( \bar{x} = 30 \) and
\[ \sum_{i=1}^{5} (x_i - \bar{x})^2 = (-10)^2 + (-10)^2 + (10)^2 + (10)^2 \]
\[ = 400. \] Sample B has \( \bar{x} = 30 \) and
\[ \sum_{i=1}^{5} (x_i - \bar{x})^2 = (-10)^2 + (-5)^2 + (5)^2 + (10)^2 = 250. \]
Thus Sample A has the larger standard deviation.
6. MBS11, problem 2.66, page 69. The Minitab file presents the data to you in two formats.

Separate columns format:
- Column C1 gives the scores for the 33 children in the Honey group.
- Column C2 gives the scores for the 35 children in the DM group.
- C3 gives the scores for 37 children in the Control group.

Single column plus identifier format:
- Column C5 gives the scores for all 33 + 35 + 37 = 105 children.
- Column C6 gives the labels corresponding to the three groups.

You can do this problem through either of the two formats provided. The single column plus identifier format generalizes to other situations and is vastly more useful.

SOLUTION: For parts a to c, you can use the separate columns format by asking for Stat ⇒ Basic Statistics ⇒ Display Descriptive Statistics for columns C1, C2, C3. The panel will look like this:

![Display Descriptive Statistics Panel](image)

The output is the following:

**Descriptive Statistics: Honey, DM, Control**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honey</td>
<td>35</td>
<td>0</td>
<td>10.714</td>
<td>0.483</td>
<td>2.855</td>
<td>4.000</td>
<td>9.000</td>
<td>11.000</td>
<td>12.000</td>
<td>16.000</td>
</tr>
<tr>
<td>DM</td>
<td>33</td>
<td>0</td>
<td>8.333</td>
<td>0.567</td>
<td>3.256</td>
<td>3.000</td>
<td>6.000</td>
<td>9.000</td>
<td>11.500</td>
<td>15.000</td>
</tr>
<tr>
<td>Control</td>
<td>37</td>
<td>0</td>
<td>6.514</td>
<td>0.483</td>
<td>2.940</td>
<td>0.000</td>
<td>5.000</td>
<td>7.000</td>
<td>8.000</td>
<td>12.000</td>
</tr>
</tbody>
</table>

The requested standard deviations are 2.855 (Honey), 3.256 (DM), and 2.940 (Control). We would regard these as very close. You could then supply the answer to d, but the wisest response might be “too close to worry about the differences.”
If you used the single column plus identifier format, the panel would be this:

![Display Descriptive Statistics Panel]

The output would be nearly identical to the previous, except for the labels:

### Descriptive Statistics: TotalScore

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>TotalScore</td>
<td>C</td>
<td>37</td>
<td>0</td>
<td>6.514</td>
<td>0.483</td>
<td>2.940</td>
<td>0.000</td>
<td>5.000</td>
<td>7.000</td>
<td>8.000</td>
<td>12.000</td>
</tr>
<tr>
<td></td>
<td>DM</td>
<td>33</td>
<td>0</td>
<td>8.333</td>
<td>0.567</td>
<td>3.256</td>
<td>3.000</td>
<td>6.000</td>
<td>9.000</td>
<td>11.500</td>
<td>15.000</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>35</td>
<td>0</td>
<td>10.714</td>
<td>0.483</td>
<td>2.855</td>
<td>4.000</td>
<td>9.000</td>
<td>11.000</td>
<td>12.000</td>
<td>16.000</td>
</tr>
</tbody>
</table>

7. MBS11, problem 2.86, page 76. Notice the phrase “likely to contain.” This statement begs for a quantification of “likely.” For your solution, make this mean that the chance is about two-thirds.

**SOLUTION:**

a. There is about a two-thirds chance (probability) that the person will be within one standard deviation of the average. For the math exam, this interval is $19 \pm 65$, or $(-46, 84)$.

b. For the verbal exam, the two-thirds chance is with interval $7 \pm 49$, or $(-42, 56)$.

c. For the math exam, the 140 increase represents $\frac{140 - 19}{65} \approx 1.86$ estimated standard deviations. For the verbal exam, this is $\frac{140 - 7}{49} \approx 2.71$ estimated standard deviations. This is *much* more likely for the math exam.
8. There two groups of salespeople at MetroMega Motors. Group $A$ consists of salespeople with no previous experience in sales, and group $B$ consists of those who do have sales experience.

a. There are 14 people in group $A$. During January, these group $A$ people had average sales of $170,000. There are 20 people in group $B$. During January, the group $B$ people had average sales of $240,000. Find the average sales during January for all 34 salespersons.

b. The 14 people in group $A$ have median age 28.4 years. The 20 people in group $B$ have median age 36.2 years. What can you say about the median of the combined group of 34?

HINT: Consider the list of values $x_1, x_2, \ldots, x_n$ and suppose that this list has median $M$. Suppose that $n$ is even. Then at least $\frac{n}{2}$ values are in the interval $(-\infty, M]$ and also at least $\frac{n}{2}$ values are in the interval $[M, \infty)$.

If $n$ is odd, the statements are that at least $\frac{n+1}{2}$ values are in $(-\infty, M]$ and at least $\frac{n+1}{2}$ values are in $[M, \infty)$.

SOLUTION: Let’s consider a. The $A$ people had total sales of $14 \times 170,000 = 2,380,000$. The $B$ people had total sales of $20 \times 240,000 = 4,800,000$. The total sales is therefore $7,180,000$. The average for all 34 people is then $\frac{7,180,000}{34} \approx \$211,176.47$. We might reasonably round this answer to $\$211,000$.

Part b. is tougher.

For the $A$ people, we know that
at least 7 people are in the interval $(-\infty, 28.4]
at least 7 people are in the interval $[28.4, \infty)$

For the $B$ people, we know that
at least 10 people are in the interval $(-\infty, 36.2]
at least 10 people are in the interval $[36.2, \infty)$

We are now certain that
at least 17 people are in the interval $(-\infty, 36.2]
at least 17 people are in the interval $[28.4, \infty)$

We are left with the unsatisfying statement that the median is known only to be in the interval $[28.4, 36.2]$.
In any set of data, the mode is the value that occurs most often. The figures below are a local bakery's daily flour utilization, in pounds, for twenty consecutive weekdays:

440  677  481  690  707
514  671  488  483  554
611  638  572  514  623
664  631  570  484  612

a. Find the mode of this set of values.
b. Round the values to the nearest ten pounds. After rounding, find the mode.
c. Round the original values to the nearest hundred pounds. After rounding, find the mode.
d. Describe the sensitivity of the mode to rounding.

SOLUTION: The values, sorted into increasing order, are these:

440  481  483  484  488
514  514  554  570  572
611  612  623  631  638
664  671  677  690  707

Sorting makes it easier to identify the mode.

For part a, we note that the mode is 514, the only value which occurs twice.

For part b, we round to the nearest 10, producing these values:

440  480  480  480  490
510  510  550  570  570
610  610  620  630  640
660  670  680  690  710

Now the mode is 480, which occurs three times.

For part c, we round to the nearest 100, giving this:

400  500  500  500  500
500  500  600  600  600
600  600  600  600  600
700  700  700  700  700

The mode is 600.

We have thus produced modes of 514, 480, and 600. These values differed only because we used different rules for rounding. For part d we must certainly conclude that the mode behaves in an unstable way with regard to rounding. Indeed, we must reach the generalization that the mode should not be used for continuous (measured) data.
10. Larcenous Larry has a problem. The data column SALES had the following summary:

**Descriptive Statistics: SALES**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>StDev</th>
<th>Sum</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calls</td>
<td>28</td>
<td>0</td>
<td>4319</td>
<td>2693</td>
<td>120925</td>
<td>1674</td>
<td>1842</td>
<td>3062</td>
<td>6542</td>
<td>8947</td>
</tr>
</tbody>
</table>

Unfortunately, he had promised that SALES would average at least 5,000. To cover this up, he maliciously edited the values to this:

**Descriptive Statistics: SALES**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>StDev</th>
<th>Sum</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calls</td>
<td>28</td>
<td>0</td>
<td>5019</td>
<td>2693</td>
<td>120925</td>
<td>1674</td>
<td>1842</td>
<td>3062</td>
<td>6542</td>
<td>8947</td>
</tr>
</tbody>
</table>

When an auditor looks at the latter display, Larry’s misdeeds will be quickly detected. How?

**SOLUTION:** The auditor will most likely notice the disconnect between the average and the total. If Larry changed the mean to 5,019, he should also have changed the total to $28 \times 5,019 = 140,532$.

11. Devious Danny computed these values:

**Descriptive Statistics: OVERHEAD**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>486</td>
<td>0</td>
<td>24.035</td>
<td>10.867</td>
<td>10.000</td>
<td>17.000</td>
<td>20.000</td>
<td>27.000</td>
<td>60.000</td>
</tr>
</tbody>
</table>

His supervisor had wanted to see an average (mean) of at most 15. Danny decided to subtract 10 from each summary number (except N and N*) and reported the following:

**Descriptive Statistics: OVERHEAD**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>486</td>
<td>0</td>
<td>14.035</td>
<td>0.867</td>
<td>0.000</td>
<td>7.000</td>
<td>10.000</td>
<td>17.000</td>
<td>50.000</td>
</tr>
</tbody>
</table>

Will an auditor detect this mischief? If so, how?

**SOLUTION:** The standard deviation of 0.867 is too low to believe for a set of $n = 486$ values with a range of Maximum – Minimum = 50. Danny was clever to alter all the location measures (mean, minimum, $Q_1$, median, $Q_3$, maximum) by the same amount, but he should not have touched the standard deviation, which is a dispersion measure. He might have gotten away with this if he had left the standard deviation at the original 10.867.
12. If A is an event with probability \( p \), we will write \( P(A) = p \). (This is a multiple use of \( p / P \), but we can handle this level of confusion.) We will say that the odds on event \( A \) are \( o = \frac{p}{1-p} \). This will usually be written as \( \frac{p}{1-p} : 1 \). The inverse relationship is \( p = \frac{o}{o+1} \).

For example, if the odds on event \( B \) are 4 : 1, then \( P(B) = \frac{4}{4+1} = 0.80 \).

When \( \frac{p}{1-p} \) is not an integer, we often convert \( \frac{p}{1-p} : 1 \) to integer form. For instance, if \( P(C) = 0.28 \), then the odds on event \( C \) are \( \frac{0.28}{1-0.28} = \frac{0.28}{0.72} = \frac{7}{18} \approx 0.3889 \). We could give the odds as either 0.3889 : 1 or as \( \frac{7}{18} : 1 \) or as 7 : 11.

In horse racing, the odds are usually given for the complementary event. Say that \( D \) is the event “horse \( D \) will win the race.” When we say “the odds on horse \( D \) are 9 : 1, we’re really giving the odds on event \( D' \). That is, the probability that horse \( D \) will lose is \( \frac{9}{9+1} = 0.90 \).

Suppose that teams \( G, H, \) and \( K \) are in a competition in which there will be exactly one winner. The odds that \( G \) will win are 0.25, and the odds the \( H \) will win are 0.40. What are the odds that team \( K \) will win?

SOLUTION: This is an easy problem in terms of probabilities. Find \( P(G) = \frac{0.25}{0.25+1} = \frac{1}{5} = 0.20 \) and then \( P(H) = \frac{0.4}{0.4+1} = \frac{2}{7} \). It follows that

\[
P(K) = 1 - \{P(G) + P(H)\} = 1 - \left\{ \frac{1}{5} + \frac{2}{7} \right\} = 1 - \frac{17}{35} = \frac{18}{35}
\]

In common-denominator terms \( P(G) = \frac{7}{35}, P(H) = \frac{10}{35} \), and \( P(K) = \frac{18}{35} \). These may look better as the decimals 0.2000, 0.2857, 0.5143.
The corresponding odds on $K$ are $o(K) = \frac{P(K)}{1 - P(K)} = \frac{18}{17} \approx 1.06$. The odds might be quoted as “1.06 to 1.”

13. Better Body, Inc., runs two health salons in Dendron, South Carolina. Here are some summary numbers on their employees. The salaries are weekly numbers.

<table>
<thead>
<tr>
<th>Branch</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Average Salary</td>
</tr>
<tr>
<td>Women’s Body Image of Dendron</td>
<td>25</td>
<td>$525</td>
</tr>
<tr>
<td>Men’s Body Works of Dendron</td>
<td>4</td>
<td>$380</td>
</tr>
</tbody>
</table>

These are not magnificent salaries, but health salons in small-town South Carolina were never places to get rich.

a. What is the average salary for all 29 female employees?
b. What is the average salary for all 16 male employees?
c. Does Better Body, Inc., discriminate against men or does it discriminate against women?

SOLUTION:

a. The female average is $\frac{25 \times $525 + 4 \times $380}{29} = \frac{$14,645}{29} = $505.$
b. The male average is $\frac{2 \times $550 + 14 \times $410}{16} = \frac{$6,840}{16} = $427.50$
c. For all 29 + 16 = 45 employees, it seems that the discrimination is against the men. However, men get the higher average salaries at each of the two locations!

If it could be claimed that, within each location, all people do approximately the same work, then the discrimination would be against the women. After all, men get more money at each location.

Salaries are very different at the two locations, and we are not given any information about why this is so.

If this were a genuine legal case, the court would want to know the specific tasks of all the people involved. It may well be the case that the two men at Women’s Body Image of Dendron are high-level managers and that the four women at Men’s Body Works of Dendron are office clerks. Or not….