Homework # 2                  due Monday, 12 September

(1) Data were collected on beginning computer sciences students at a large university, to determine how well their success in the first years of college could be predicted. (This is the “CSDATA” data set in the Data Appendix of previous editions of the text.)

There were 224 students. The variables in the data set are:

• obs: observation number
• gpa: the student’s GPA for the first 3 semesters of college, on a 0–4 scale (4 = A).
• HSM: the student’s high school grade average in math, on a 0–10 scale (10 = A).
• HSS: the student’s high school grade average in science, on a 0–10 scale (10 = A).
• HSE: the student’s high school grade average in English, on a 0–10 scale (10 = A).
• SATM: the student’s score on the Mathematics part of the SAT, taken prior to starting college.
• SATV: the student’s score on the Verbal part of the SAT.
• Sex: 1 = male, 2 = female.

The data set will be distributed by email and posted on the course web site. The first four observations are:

<table>
<thead>
<tr>
<th>bs</th>
<th>gpa</th>
<th>hsm</th>
<th>hss</th>
<th>hse</th>
<th>satm</th>
<th>satv</th>
<th>sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.32</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>670</td>
<td>600</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2.26</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>700</td>
<td>640</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2.35</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>640</td>
<td>530</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2.08</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>670</td>
<td>600</td>
<td>1</td>
</tr>
</tbody>
</table>

Your assignment is to examine the relationships among GPA, SATM, and sex:

(a) Do you consider any of these variables to be explanatory or response variables? Why, or why not? If so, which? (Pay attention to the temporal order in which these scores are obtained for each student, and to the purpose of the study as given in the first sentence at the start of this question.)

(b) Make a scatterplot of GPA and SATM, with GPA on the vertical axis, and add a smoother. Describe what you see (form, direction and strength of the relationship, outliers).

(c) Make another scatterplot of GPA and SATM, using different symbols for males and females; add a smoother for each sex. Describe how the relationship of GPA and SATM compares between sexes.

(2) Using the CSDATA from question (1), calculate correlations between GPA and SATM for all students and then separately for males and for females. Describe and explain the differences between the correlation results for all students and those for males and females separately.

(3) Using only the females in the CSDATA from question (1), fit a least-squares regression line to the relationship of SATM (explanatory variable) and GPA (response variable). Report the estimated regression equation, and define the coefficients in it.
A student wonders whether tall women tend to date taller men than do short women. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man each woman dates. She obtains a correlation coefficient between the woman’s height and the man’s of 0.56. Here are the data (heights in inches); you don’t need them to answer the questions but they may be helpful.

<table>
<thead>
<tr>
<th>women</th>
<th>66</th>
<th>64</th>
<th>66</th>
<th>65</th>
<th>70</th>
<th>65</th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td>72</td>
<td>68</td>
<td>70</td>
<td>68</td>
<td>71</td>
<td>65</td>
</tr>
</tbody>
</table>

(a) Does this analysis answer her question of whether, for these women, taller women tend to date taller men? If so, what is the answer?

(b) Does this correlation tell us whether the women tend to date men taller than themselves? How would the correlation change if all the mean were 6 inches shorter than the values in the table above?

(c) If heights were measured in centimeters rather than inches, how would the correlation change?

(d) If every woman dated a man exactly 3 inches taller than herself, what would be the correlation between female and male heights?