The purpose of finding a regression line is description of the relation between the predictors and the response in the population. The regression equation gives the best linear fit to the data (sample) and the coefficient of determination ($R^2$) measures the closeness of that fit to the data, not to the population. The two main issues for “usefulness” are determining whether the data indicate a linear relation in the population (and determining which predictors are useful) and (if so) estimating the rate of change of the response (based on the different predictors) and using the equation to predict means and individual values.

**LEARNING OBJECTIVES**

1. Work as a team, using the team roles
2. Gain experience with linear regression concepts
3. Learn how to determine if a regression equation gives evidence of a linear relation in a population
4. Learn to identify (statistically) significant predictors in a regression situation
5. Learn to use the regression line for prediction

**CRITERIA**

1. Success in completing the exercises.
2. Success in answering questions about the model
3. Success in working as a team

**RESOURCES**

1. The “Multiple Regression” handout from Monday
2. Your Text - especially sections 15.5 and 15.6 and the table of data on p. 688
3. Your Calculator
4. 40 minutes

**PLAN**

1. Select roles, if you have not already done so, and decide how you will carry out steps 2 and 3 (5 minutes)
2. Work through the exercises given here - be sure everyone understands all results (30 minutes)
3. Assess the team’s work and roles performances and prepare the Reflector’s and Recorder’s reports including team grade (5 minutes).
4. Be prepared to discuss your results.

**EXERCISE**

The table given in Case problem 3 (p687-688 in your text) gives information on a sample of 48 national universities, showing graduation rate (percent of entering first-year students who graduate from the school), percent of classes that are under 20 (students), the student-faculty ratio (# students per faculty member) and the alumni/ae giving rate (percent of alumni/ae who donate to the university). A national fundraising organization wishes to see if graduation rate, % of classes under size 20 and student-faculty ratio can be used to effectively predict the alumni/ae giving rate. The Minitab printout of the regression calculation is given below.

1. Using the regression model for predicting % giving based on all three predictors, answer the questions below (very little calculation is required - mostly explanation).
   (a) Give the regression equation (with context, of course) for predicting alumni/ae giving rate based on these three predictors.
   (b) What does the equation give as the predicted giving rate for New York University (graduation rate 72%, 63% of classes under 20, student-faculty ratio 13)? What is the residual?
(c) What proportion of variation in alumni/ae giving rate is “explained” by the relation to the three predictors?

(d) The coefficient of “graduation rate” in the equation is .748. What does this tell us (be specific — interpret that actual number)?

(e) Set up and carry out [no calculation is necessary — but writing is] the test to determine whether there is evidence of a linear relation useful for predicting alumni/ae giving rate based on the (set of) three predictors used here. Does it appear there is a significant linear relationship?

(f) Which predictors are shown to be useful [statistically significant] for predicting the giving rate (give the evidence)?

(g) Use the information in the printout (you will have to do some calculation) to give a 95% confidence estimate for the decrease in alumni/ae giving rate for each unit increase in the student-faculty ratio. [Note: the standard error $s_b$, of each coefficient is given in the printout, in the column SE Coeff]

(h) The printout shows the 95% “95% CI” for schools with graduation rate 72%, with 63% of classes under 20 and with student-faculty ratio 13 as “(14.20, 24.77)”. What does this mean? [It certainly isn’t the same as g, above]

(i) Similarly, the printout shows the “95% PI” for schools with graduation rate 80%, with 45% of classes under 20 and with student-faculty ratio 15 as “(3.26, 35.70)”. What does this mean [why is it different from the preceding result]?

2. Using also the Minitab printout for the regression which uses only graduation rate and student-faculty ratio as predictors:

(a) Give the regression equation.

(b) Compare the usefulness of this regression model to the three-predictor model, using $R^2$(adj). How does this correspond to your result in part f above?

(c) What does this model give as a 95% confidence interval for the decrease in giving rate for each unit increase in faculty-student ratio? Why doesn’t it match the previous result (1g)?

READING ASSIGNMENT (in preparation for next class)
Read Sections 15.7 (use of qualitative predictors), and 15.8(residual analysis)

SKILL EXERCISES:(hand in - individually - at next class meeting): p.645 #24, 25 p.648 #28, 29, 31 (Use Minitab where possible)

Regression Analysis: %Giving versus %Grad, Under 20, S/FRatio

The regression equation is
\[
\%\text{Giving} = -20.7 + 0.748 \%\text{Grad} + 0.029 \text{Under 20} - 1.19 \text{S/FRatio}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-20.72</td>
<td>17.52</td>
<td>-1.18</td>
<td>0.243</td>
</tr>
<tr>
<td>%Grad</td>
<td>0.7482</td>
<td>0.1660</td>
<td>4.51</td>
<td>0.000</td>
</tr>
<tr>
<td>Under 20</td>
<td>0.0290</td>
<td>0.1393</td>
<td>0.21</td>
<td>0.836</td>
</tr>
<tr>
<td>S/FRatio</td>
<td>-1.1920</td>
<td>0.3867</td>
<td>-3.08</td>
<td>0.004</td>
</tr>
</tbody>
</table>

$S = 7.60972$  $R^2\text{-Sq} = 70.0\%$  $R^2\text{-Sq(adj)} = 67.9\%$

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>3</td>
<td>5943.5</td>
<td>1981.2</td>
<td>34.21</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>44</td>
<td>2547.9</td>
<td>57.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>8491.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Predicted Values for New Observations

New
Obs  Fit  SE Fit  95% CI    95% PI
   1  19.48  2.62  (14.20, 24.77)  (3.26, 35.70)

Values of Predictors for New Observations

New
Obs  %Grad  Under 20  S/FRatio
   1  72.0      63.0     13.0

Regression Analysis: %Giving versus %Grad, S/FRatio

Regression Analysis: %Giving versus %Grad, S/FRatio

The regression equation is
%Giving = - 19.1 + 0.756 %Grad - 1.25 S/FRatio

Predictor   Coef    SE Coef   T     P
Constant    -19.11   15.55     -1.23 0.226
%Grad       0.7557   0.1602     4.72 0.000
S/FRatio    -1.2460  0.2843    -4.38 0.000

S = 7.52841  R-Sq = 70.0%  R-Sq(adj) = 68.6%

Analysis of Variance

Source    DF    SS          MS         F     P
Regression 2  5941.0   2970.5  52.41 0.000
Residual Error 45  2550.5   56.7
Total      47  8491.5

Predicted Values for New Observations

New
Obs  Fit  SE Fit  95% CI    95% PI
   1  19.11  1.90  (15.29, 22.93)  (3.47, 34.75)

Values of Predictors for New Observations

New
Obs  %Grad  S/FRatio
   1  72.0     13.0