Lab Session 12: Multiple Regression Analyses:

Please use datasheets provided on our website: http://courses.eller.arizona.edu/mgmt/delaney/d15s_sbs200_labs.htm

This lab will provide you with an opportunity to complete two multiple regression analyses. Remember, multiple regression analysis allows us to use correlational analyses to evaluate many variables at once. As you will see, while multiple regression is a very powerful tool that can be complicated and involved, simpler questions can be fairly straightforward to actually run and can also provide powerfully useful information from the results.

Today you will want to
- interpret each of the simple correlations (positive vs negative and weak vs strong relationships)
- generate scatterplots for all of the simple correlations
- identify the intention of the regression
  - what are we trying to predict? – we predict the single dependent variable
  - what are we using to predict it? – and several predictor variables (independent variable)
- find the regression equation and interpret the slopes

Problem 1: We are interested in predicting heating cost for the month of January from these three variables:
- Average January temperatures (measured in Fahrenheit degrees)
- Thickness of insulation in the attic (measured in inches)
- Age of the furnace (measured in years)

Part 1. Predict the simple correlation for each independent variable with the dependent variable by completing these three scatterplots, and predicting a value for r. (Remember “Heating Cost” is the predicted variable and will go on the Y axis.)

<table>
<thead>
<tr>
<th>What you predict the correlation between Average Temperatures and Heating Cost to be?</th>
<th>What you predict the correlation between Thickness of Insulation and Heating Cost to be?</th>
<th>What you predict the correlation between the Age of the Furnace and Heating Cost to be?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What do you predict this correlation to be? 
$r =$
According to the matrix below, what did this correlation actually turn out to be? 
$r =$

What do you predict this correlation to be? 
$r =$
According to the matrix below, what did this correlation actually turn out to be? 
$r =$

What do you predict this correlation to be? 
$r =$
According to the matrix below, what did this correlation actually turn out to be? 
$r =$

What do you predict this correlation to be? 
$r =$
According to the matrix below, what did this correlation actually turn out to be? 
$r =$
Part 2. Open the Excel database found on our website:
http://courses.eller.arizona.edu/mgmt/delaney/d15s_sbs200_labs.htm
It should look like this:

Part 3: Create a Correlation Matrix describing the three correlations you predicted in Part 1.

On toolbar on top
- click “Data” and choose
  “Data Analysis”
- Then choose “Correlation”

For the “Input range” we want to choose all four columns (including labels).

Part 4. Let’s complete a Multiple Regression
- On toolbar on top
- click “Data” and choose
  “Data Analysis”
- Then choose “Regression”
  You’ll see a box like this:
  - For the “Input Y range” we want to choose the 20 scores and the title for Heating $ Remember to choose just that one column
  - For the “Input X range” we want to choose the 20 scores and the title for the other three variables (Temp, Insulation, Age) Remember to choose all values for all three columns
  - Click okay and you should get an output that has a lot of information including your regression coefficients:

Part 5. Let’s interpret our results

What was your regression coefficient for “Intercept” ______________

What was your regression coefficient for “Temp” ______________

What was your regression coefficient for “Insulation” ______________

What was your regression coefficient for “Age of Furnace” ______________

What is your regression equation ___________________________

\[ Y' = a + b_1X_1 + b_2X_2 + b_3X_3 \]

Interpreting slopes:
We can predict heating cost by knowing the outside temperature:
  - As the outside temperature increases, the cost to heat the home decreases. If we increase temperature by 1 degree and hold the other two independent variables constant, we can estimate a decrease of $________ in monthly heating cost.

We can predict heating cost by knowing the amount of insulation:
  - For each additional inch of insulation, we expect the cost to heat the home to decline $________ per month, regardless of the outside temperature or the age of the furnace.

We can predict heating cost by knowing the age of the home:
  - for each additional year older the furnace is, we expect the cost to increase $________ per month.
Applications for regression equation:

- MaryEllen wants to estimate her heating cost. The mean temperature outside is 30 degrees, there are 5 inches of insulation in the attic and the furnace is 10 years old. What will she predict for her heating cost? _______________

- MaryEllen wants to estimate her heating cost again the very next year. Everything is the same (temperature still equals 30 degrees and insulation still equals 5 inches), but now her furnace is one year older. How much more will her bill be? There are two ways to solve this problem, what are they?
  o __________________________________________________________________________
  o __________________________________________________________________________

Problem 2: You have just opened up a school designed to teach high school graduates to become Paralegal professionals. As an instructor you want to predict which students applying to get into your program will do the best in the paralegal courses. You are interested in predicting GPA in the Paralegal courses from these three variables:

- High School GPA (HS-GPA)
- Scores on Verbal portion of the S.A.T (SAT–V)
- Scores on the Mathematical portion of the S.A.T (SAT–M)

Part 1. Predict the simple correlation for each independent variable with the dependent variable by completing these three scatterplots, and predicting a value for r. (Remember “Paralegal GPA” is the predicted variable and will go on the Y axis.)

<table>
<thead>
<tr>
<th>What you predict the correlation between Paralegal GPA and high school GPA to be?</th>
<th>What you predict the correlation between Paralegal GPA and SAT-Verbal scores to be?</th>
<th>What you predict the correlation between Paralegal GPA and SAT-Math scores to be?</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

What do you predict this correlation to be?

r =

According to the matrix below, what did this correlation actually turn out to be?

r =

What do you predict this correlation to be?

r =

According to the matrix below, what did this correlation actually turn out to be?

r =

What do you predict this correlation to be?

r =

According to the matrix below, what did this correlation actually turn out to be?

r =
Part 2: Open the Excel database found on our website, it should look like this:

<table>
<thead>
<tr>
<th></th>
<th>Student</th>
<th>Paralegal (Y)</th>
<th>HS GPA (X₁)</th>
<th>SAT-V (X₂)</th>
<th>SAT-M (X₃)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3.21</td>
<td>3.25</td>
<td>480</td>
<td>410</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.68</td>
<td>1.80</td>
<td>290</td>
<td>270</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3.58</td>
<td>2.89</td>
<td>420</td>
<td>410</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3.92</td>
<td>3.13</td>
<td>500</td>
<td>600</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>3.00</td>
<td>2.81</td>
<td>430</td>
<td>460</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>2.82</td>
<td>3.21</td>
<td>320</td>
<td>490</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>1.65</td>
<td>2.20</td>
<td>530</td>
<td>460</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>2.30</td>
<td>2.14</td>
<td>469</td>
<td>440</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>2.33</td>
<td>2.63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Part 3: Create a Correlation Matrix describing the three correlations you predicted in Part 1. “Refer to the instructions included in the previous problem if you are not sure how to do this.”

Part 4: Let’s complete a Multiple Regression
- On toolbar on top click “Data” and choose “Data Analysis” Then choose “Regression”
- For the “Input Y range” we want to choose the 9 scores and the title for Paralegal (Y) Remember to choose just that one column
- For the “Input X range” we want to choose the 9 scores and the title for the other three variables (HS GPA, SAT-V, and SAT-M) Remember to choose all values for all three columns
- Your output should look like this:

Part 5: Let’s interpret our results

What was your regression coefficient for “Intercept” ___________; Is the p < 0.05? ___________

What was your regression coefficient for “HS GPA” ___________; Is the p < 0.05? ___________

What was your regression coefficient for “SAT-V” ___________; Is the p < 0.05? ___________

What was your regression coefficient for “SAT-M” ___________; Is the p < 0.05? ___________

According to the p values, which variable is most significant? ________________________________

What is your regression equation _______________________________________________________

\[ Y' = a + b_1X_1 + b_2X_2 + b_3X_3 \]

Interpreting slopes:
- As the High School GPA increases, the predicted GPA for Paralegal studies will increase. If we increase High School GPA 1 full point (from 2.0 to 3.0) and hold the other two independent variables constant, we can estimate an increase of ________ in paralegal GPA.

- If we increase SAT-Verbal score by 1 point and hold the other two independent variables constant, we can estimate an increase of ________ in paralegal GPA.

- If we increase SAT-Math score by 1 point and hold the other two independent variables constant, we can estimate an decrease of ________ in paralegal GPA.
Applications for regression equation:

- Here comes Alice. Her HS GPA was 2.80, her SAT-V score was 430 and her SAT-M score was 460. What will you predict for her Paralegal GPA? _______________

- Here comes Alberta. Her HS GPA was 3.80, her SAT-V score was 430 and her SAT-M score was 460. What will you predict for her Paralegal GPA? _______________

- The only thing that changed between Alice and Alberta was the HS GPA. Alberta’s was exactly one point higher. Her predicted GPA for paralegal school was 1.201 higher than Alice. Why is this interesting, why was it 1.2, could we have gotten that from our regression equation? 

Please propose a multiple regression that would be relevant to issues in your own major or interests
(let’s keep all of our variables numeric)

What would you want to predict? 
This is your dependent variable and will go on your “Y” axis

What variables would predict this well?
Each of these will be your independent variables and will go on your “X” axis

1) 
2) 
3) 

<table>
<thead>
<tr>
<th>What would you predict the correlation to be between your predicted (DV) and your first predictor variable (IV)</th>
<th>What would you predict the correlation to be between your predicted (DV) and your second predictor variable (IV)</th>
<th>What would you predict the correlation to be between your predicted (DV) and your second predictor variable (IV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Draw a scatterplot and predict what you think this correlation would be? 
$r =$

Draw a scatterplot and predict what you think this correlation would be? 
$r =$

Draw a scatterplot and predict what you think this correlation would be? 
$r =$
Problem 1 Correlation Matrix:

<table>
<thead>
<tr>
<th></th>
<th>Heating $(Y)$</th>
<th>Temp (X1)</th>
<th>Insulation (X2)</th>
<th>Age Furnace (X3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heating $(Y)$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temp (X1)</td>
<td>-0.811508835</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulation (X2)</td>
<td>-0.257101335</td>
<td>-0.103016129</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Age Furnace (X3)</td>
<td>0.536727562</td>
<td>-0.485987697</td>
<td>0.063617402</td>
<td>1</td>
</tr>
</tbody>
</table>

Problem 1 Regression Analysis Output:

<table>
<thead>
<tr>
<th>SUMMARY OUTPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression 0.890755299</td>
</tr>
<tr>
<td>R Square 0.804170066</td>
</tr>
<tr>
<td>Adjusted R Square 0.767451554</td>
</tr>
<tr>
<td>Standard Error 5.104855358</td>
</tr>
<tr>
<td>Observations 20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Regression 3</td>
</tr>
<tr>
<td>Residual 16</td>
</tr>
<tr>
<td>Total 19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95.0%</th>
<th>Upper 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 427.1938033</td>
<td>59.60142931</td>
<td>7.167509374</td>
<td>2.23764E-06</td>
<td>300.8444183</td>
<td>553.5431883</td>
<td>300.8444183</td>
<td>553.5431883</td>
</tr>
<tr>
<td>Temp -4.562663626</td>
<td>0.772818958</td>
<td>-5.983638915</td>
<td>2.10035E-05</td>
<td>-6.219900505</td>
<td>-2.945418747</td>
<td>-6.219900505</td>
<td>-2.945418747</td>
</tr>
<tr>
<td>Age Furnace 6.101032061</td>
<td>4.011210664</td>
<td>1.520050381</td>
<td>0.147862484</td>
<td>-2.404282684</td>
<td>14.60634681</td>
<td>-2.404282684</td>
<td>14.60634681</td>
</tr>
</tbody>
</table>

Problem 1 Regression Equation: $Y' = a + b_1X_1 + b_2X_2 + b_3X_3$

Heating Cost = 427.19 + (-4.58)(Temperature) + (-14.83)(Insulation) + (6.10)(Age)