MAT 120 REGRESSION ASSIGNMENT

Due: January 7, 2013

In real-world applications we often encounter numerical data in the form of a table. The powerful mathematical tool Regression Analysis is a process for finding a function that provides a useful model for a set of data points. One uses computer software to create a graph (a scatter plot) of the set of points and find the function of best fit. The function can then be used to predict values of the dependent variable y for any given value of the independent variable x.

Review §1.3, 2.3, 2.5, of the textbook and try the following example before starting the assignment.

1. EXAMPLE

Prices for round-shaped diamonds taken from an online trader are given in Table 1. Weight is the independent variable x and price is the dependent variable y. We will analyze this data in the computer.

For Windows users, I suggest using Excel (Microsoft Office). For Mac users, I suggest using Numbers (iWork '09). If you do not have this software, it is available in the university computer labs. More information about the available software can be found at

http://its.uncg.edu/Software/.

For Excel and Numbers, enter `chart` into the Help menu to find out how to create charts, and enter `trend` into the Help menu to find detailed instructions on creating and manipulating trendlines for data series.

TABLE 1. Prices for round-shaped diamonds

| Weight (carats)	Price ($)
0.5	2790
0.6	3191
0.7	3694
0.8	4154
0.9	5018
1.0	5898

(1) Enter the data into a spreadsheet (including the header Weight and Price).
(2) Create a scatter plot of the data. Note that the scatter plot for this data appears roughly linear.
(3) Compute the linear regression equation (trendline). The trendline will be $y = ax + b$ for some values of a and b.
(4) Adjust the minimum/maximum values on the axes of the graph to 0.4 and 1.2 for the weight and 2000 and 6500 for the price.

A sample of the output created on Mac Numbers is given below. The regression equation $y = 6137.4x - 478.9$ is on the graph. To predict the price of a diamond weighing 0.85 carats, calculate price by plugging in $x = 0.85$ into the regression equation

$$y = 6137.4(0.85) - 478.9 = \$4737.89.$$

Diamond Prices for round-shaped diamonds

Weight (carats)	Price ($)
0.5	$2,790.00
0.6	$3,191.00
0.7	$3,694.00
0.8	$4,154.00
0.9	$5,018.00
1.0	$5,898.00

Diamond prices

$y = 6137.4x - 478.9$

+ Weight (carats)   — Price
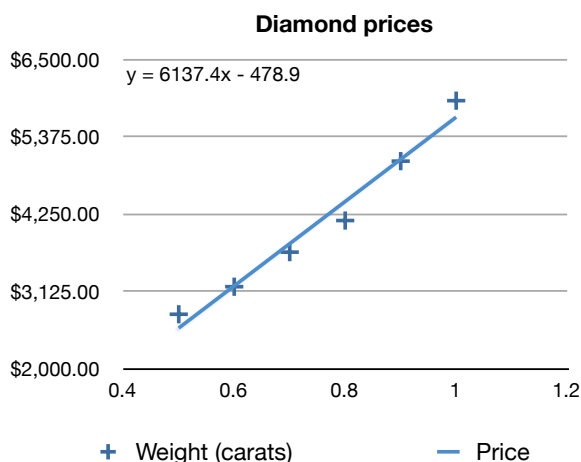
Predicted price for 0.85 carat diamond:	$4,737.89

2. ASSIGNMENT

Complete the following exercises in Excel, Numbers, or similar software. Save your work often. You answers should be entered in the Regression Assignment under the TESTS tab of MyMathLab. Your work can be submitted any time before January 7, 2013.

Exercises.

(1) Table 2 contains price-supply data and price-demand data for soybeans.
- Enter the data into a spreadsheet.
- Create the scatter plots for the price-supply, where x is the supply (in billions of bushels) and y is the price (in dollars). Do the same for price-demand.
- Adjust the minimum and maximum of the axes of each plot to slightly below and slightly above the data values.
- Compute the regression equations for supply and for demand using linear regression on each of the plots. The trendline will be $y = ax + b$ for some values of a and b.
- Find the equilibrium price for soybeans. (Hint: The supply model will be an increasing linear function. The price model will be a decreasing linear function.

Set the two equations equal to each other and solve for the equilibrium value x, and then find the corresponding value for y equilibrium price.)

TABLE 2. Supply and demand for soybeans.

Supply (billion bu)	Price ($/bu)	Demand (billion bu)	Price ($/bu)
1.55	5.15	2.60	4.93
1.86	5.79	2.40	5.48
1.94	5.88	2.18	5.71
2.08	6.07	2.05	6.07
2.15	6.15	1.95	6.40
2.27	6.25	1.85	6.66

(2) Table 3 contains data for fuel consumption (mpg) of an outboard motor at various rpm.
- Enter the data into a spreadsheet so that x represents the rpm in thousands. e.g. enter $x = 1.5$ for 1500, enter $x = 2.0$ for 2000 etc.
- Create the scatter plot for the fuel consumption y (mpg) as a function of engine speed x (rpm).
- Adjust the minimum and maximum of the axes of each plot to slightly below and slightly above the data values.
- Compute the regression equation using quadratic (polynomial order 2) regression. The trendline will be $y = ax^2 + bx + c$ for some values of a, b, and c.
- Use your regression equation to estimate the fuel consumption at 2300 rpm.

TABLE 3. Fuel consumption for outboard motor.

rpm	mpg
1500	8.2
2000	6.9
2500	4.8
3000	4.1
3500	3.7

(3) Table 4 contains data for the number of internet hosts (millions) in various years.
- Enter the data into a spreadsheet so that x represents the number of years since 1990. e.g enter $x = 4$ for 1994, enter $x = 7$ for 1997, etc.
- Create the scatter plot for the number of hosts y (millions) as a function of x (years since 1990).
- Adjust the minimum and maximum of the axes of each plot to slightly below and slightly above the data values.
- Compute the regression equation using exponential regression. The trendline will be $y = ae^{bx}$ for some values of a and b.
- Use your regression equation to estimate the number of hosts in 2018 ($x = 2018 - 1990 = 28$).

TABLE 4. Internet hosts.

Year	Hosts (millions)
1994	2.4
1997	16.1
2000	72.4
2003	171.6
2006	394.0
2009	625.2

(4) Table 5 contains data for the number of dairy cows (thousands) in the U.X. in various years.
   - Enter the data into a spreadsheet so that x represents the number of years since 1940. e.g enter $x = 10$ for 1950, enter $x = 20$ for 1960, etc.
   - Create the scatter plot for the number of cows y (thousands) as a function of x (years since 1940).
   - Adjust the minimum and maximum of the axes of each plot to slightly below and slightly above the data values.
   - Compute the regression equation using logarithmic regression. The trendline will be $y = a \ln(x) + b$ for some values of a and b.
   - Use your regression equation to estimate the number of dairy cows in 2020 ($x = 2020 - 1940 = 80$).

TABLE 5. Dairy cows on farms in the U.S.

Year	Cows (thousands)
1950	23853
1960	19527
1970	12091
1980	10758
1990	10015
2000	9190