

Business Analytics (108-1)

Assignment 1

Due: 9:00 am, Tue 01-Oct-2019

1.

The exam results of Statistics are listed as follows.

94	81	79	66	96	73	70	68	94	70
64	84	66	64	80	76	87	65	65	91
60	68	62	90	95	62	69	66	80	61
77	60	66	52	82	63	81	62	73	92
81	67	84	70	78	48	86	87	81	79

Please make a numerical and graphical summary of the data, commenting on any features that you find interesting.

2.

A manufacturer produces custom metal blanks that are used by its customers for computer-aided machining. The customer sends a design via computer (a 3-D blueprint), and the manufacturer comes up with an estimated cost per unit, which is then used to determine a price for the customer. This analysis considers the factors that affect the cost to manufacture these blanks. The data for analysis, “production_cost_1.txt”, were sampled from the accounting records of 195 previous orders that were filled during the last 3 months.

- Create a scatterplot for the average cost per item on the material cost per item. Do you find a linear pattern?
- Estimate the linear equation using least squares. Interpret the fitted intercept and slope. Be sure to include their units.
- Interpret R^2 and $\hat{\sigma}$.
- What is the estimated increase in the total cost per finished item if the material cost per unit goes up by \$3?
- One can argue that the slope in this regression should be 1, but it is not. Explain the difference.
- The total cost of an order in these data was \$61.16 per unit with material cost of \$4.18 per unit. Is this a relatively expensive order given the material cost?
- Plot the residuals from this regression. What does the standard deviation of the residuals tell about the fit of this equation?

3.

Utility companies in many older communities still rely on “meter readers” who visit homes to read meters that measure consumption of electricity and gas. Unless someone is home to let the meter reader inside, the utility company has to estimate the amount of energy used. The utility company in this example sells natural gas to homes in the Philadelphia area. Many of these are older homes that have the gas meter in the basement. We can estimate the use of gas in these homes with a simple linear model.

The explanatory variable is the average number of degrees below 65 during the billing period, and the response is the number of hundred cubic feet of natural gas (CCF) consumed during the billing period (about a month). The explanatory variable is set to 0 if the average temperature is above 65 (assuming a homeowner won’t need heating in this case).

- (a) Use R to fit a simple linear model with the data “gas_consumption.txt”. Do the analysis, make the plot, and summarize the results.
- (b) Modify the script provided for this lecture to create the “Confidence and Prediction Intervals” plot shown on lecture note p. 2-32 and re-printed below.

