

**Homework #8 – Due Monday, Dec. 4**  
COR1-GB.1305 – Statistics and Data Analysis

**Problem 1**

The file `DiamondPrices.CSV` contains data on retail prices (in Dollars) for 617 round shaped diamonds. As the predictor variable, we will focus on Carats, a measure of weight. (One Carat = 200 mg.)

- (a) Make a scatterplot of Price versus Carats, and comment on the reasonableness of fitting a linear regression model to this data.
- (b) Run the regression of Price on Carats, using *Stat*  $\Rightarrow$  *Regression*  $\Rightarrow$  *Regression*  $\Rightarrow$  *Fit Regression Model*, set *Responses: Price*, and *Continuous Predictors: Carats*. Copy and paste the Minitab regression output for **Model Summary**, **Coefficients**, and **Regression Equation**
- (c) What is the equation of the fitted line? Use this equation to predict the price of a diamond ring which weighs 1.25 carats.
- (d) Is there evidence of a significant linear relationship between the price and the weight of the diamond? Justify your answer.
- (e) Interpret the estimated slope of the fitted model, and construct a 95% confidence interval for the true slope coefficient. What is the practical meaning of the true slope coefficient?
- (f) Discuss and give a practical interpretation of the coefficient of determination,  $R^2$ .
- (g) Does the negative estimated intercept of the fitted model bother you? What is the interpretation of the true intercept?
- (h) What is the estimate of the typical fluctuation of data points from the true regression line, measured in the vertical direction?
- (i) Using Minitab, construct a 95% confidence interval for the expected price of a ring which weighs 1.25 Carats. (To do this , after running the regression click on *Stat*  $\Rightarrow$  *Regression*  $\Rightarrow$  *Regression*  $\Rightarrow$  *Predict*. Type in 1.25 in the first line under Carats.)

.....

## Problem 2

Consider the data in `MARKET.CSV`.

- (a) Construct the fitted line plot for `IBMRet` versus `MarketReturn`. Does this suggest a linear relationship between the two variables?
- (b) Identify the outlier in the lower left-hand corner of the plot by resting the cursor over the point and then going to the spreadsheet to find the corresponding case.
- (c) Run the regression of `IBMRet` versus `MarketReturn`. Write the equation for the fitted model. (In finance, this is called the market model.) What is the slope of the fitted line? (In finance, they call this the “beta” for IBM, but actually it’s just an estimate of the true slope,  $\beta_1$ .)
- (d) Is there strong evidence of a linear relationship between `MarketReturn` and `IBMRet`?
- (e) Find a 95% confidence interval for the true slope. Does this interval contain the value 1?
- (f) In finance, the performance of an investment compared to the market is often measured by the “alpha”, which is equal to the estimated intercept,  $\hat{\beta}_0$ . What was the value of  $\hat{\beta}_0$  for IBM? What is the interpretation of this value?
- (g) Is there evidence that the true  $\beta_0$  for IBM is nonzero? What is the relevant  $p$ -value? Interpret this  $p$ -value.

.....