

## MTH 1210

### Lab 7 Scatterplots, Correlation and Linear Regression

#### Part A: Scatterplots

The file **diamonds.mpx** contains data on  $n = 308$  cut diamonds that appeared in a jeweler's advertisement in the Feb. 18, 2000 edition of Singapore's *Business Times*. The data set includes the following variables:

**Carats:** Weight of the diamond in carats (1 carat = 0.2 grams)

**Color:** The color purity class of the diamond (D = highest level of purity, then E, F, G, H and I = lowest level of purity)

**Clarity:** The clarity class of the diamond (IF = Internally flawless (highest level of clarity), VVS1 = Very very slightly imperfect 1, VVS2 = Very very slightly imperfect 2, VS1 = Very slightly imperfect 1, or VS2 = Very slightly imperfect 2 (lowest level of clarity))

**CertBody:** The certification body associated with the diamond (GIA = Gemmological Institute of America, IGI= International Gemmological Institute, and HRD = Hoge Raad Voor Diamant (in Antwerp))

**Price:** The price of the diamond in U.S. dollars.

1. Open the MINITAB *worksheet* **diamonds.mpx**. (If you saved it, then either double click its icon or start MINITAB and use FILE > OPEN).
2. Make a scatterplot of **Price** (response variable,  $y$ -axis) versus **Carats** (explanatory variable,  $x$ -axis):

GRAPH > SCATTERPLOT

This will take you to a selection of types of scatterplots.

Choose:       SIMPLE

< OK >

In the dialog box:     Select the variables: Y-variable, X-variable

< OK >

This will produce a nice scatterplot in a graphics window.

You can double click on the  $x$ -axis or the  $y$ -axis to modify the labeling.

You can double click inside the graph for coloring and other attributes.

You can double click on any of the dots to modify the shape, size and color of the dots.

3. **Copy and paste the scatterplot into the Answer Sheet.**
4. **Please answer Questions 1 and 2 on the Answer Sheet.**

## Part B: Adding a Qualitative Variable to a Scatterplot

We can include the qualitative variable **Color** in the scatterplot by using different plot symbols to represent different colors of diamond.

1. Make another scatterplot with **Carats** on the  $x$ -axis and **Price** on the  $y$ -axis, but with different diamond **Colors** represented as different plot symbols:

GRAPH > SCATTERPLOT

This will take you to a selection of types of scatterplots.

Choose: WITH GROUPS

< OK >

In the dialog box: Select the variables: Y-variable, X-variable  
Select the qualitative variable for grouping.

< OK >

This will produce a nice scatter plot in a graphics window, with different categories of the qualitative variable represented by different plot symbols.

2. **Copy and paste the scatterplot into the Answer Sheet.**
3. **Please answer Question 1 on the Answer Sheet.**

## Part C: Correlation

The file **golfscores.mpx** has data representing golf scores for 11 members of a women's golf team in two rounds of a tournament.

1. Open the data file **golfscores.mpx**.
2. Compute the correlation  $r$  between **Round1** and **Round2** scores:

STAT > BASIC STATISTICS > CORRELATION

In the dialog box: Select the variables, X and Y.

Clear the option "Display P-Value" (click on the check mark next to it)

Also make sure that the other box is cleared.

<OK>

This will give the correlation, under the title "Correlations: x, y." It is referred to as 'Pearson correlation'.

3. **Please answer Questions 1 and 2 on the Answer Sheet.**

## Part D: Linear Regression

We can fit a regression line to the **Round1** and **Round 2** golf scores and use it to make predictions of second round scores based on first round scores.

1. Make a scatterplot of the data *with the regression line*. Put **Round1** on the  $x$  axis and **Round2** on the  $y$  axis:

GRAPH > SCATTER PLOT

This will take you to a selection of types of scatter plots.

Choose: WITH REGRESSION

<OK>

In the dialog box: Select the variables: Y-variable, X-variable

< OK >

2. **Copy and paste the scatterplot into the Answer Sheet.**
3. Now find the equation of the regression line:

STAT > REGRESSION > REGRESSION

In the dialog box: Response: Double click on the Y-variable

Predictors: Double click on the X-variable

If you need the  $\hat{y}$  values (i.e. the predicted, or fitted values), and the errors (i.e. the residuals):

Click on < RESULTS >

Then choose the 4<sup>th</sup> option.

This will give you the equation of the regression line in the Session window, under the title: "Regression Analysis: y versus x."

Note: The regression output comes with a lot of other information besides the equation of the line.

4. **Please answer Questions 1, 2, 3 and 4 on the Answer Sheet.**