**HDA Workshop**

Lab #1

**Exercise 1 (CrossTabs) – 15 Mins**

The file *OpinionSurvey.xlsx* contains data from a survey of 399 people regarding a government environmental policy.

1. Create a crosstabs and an associated column chart for Gender versus Opinion. Express the counts as percentages so that for either gender, the percentages add to 100%. Discuss your findings. Specifically, do the two genders tend to differ in their opinions about the environmental policy?
2. Repeat part a) with Age versus Opinion.
3. Then repeat part a) with Recoded Salary column versus Opinion. (Note how the recoding used VLOOKUP formula – if you are unsure of how to use this, please go through the Lookups & Referencing PowerPoint & examples in Excel Basics)

**Exercise 2 (Pivot Tables) – 20 mins**

The file *Marathon.xlsx* lists the results of about 20,000 runners in the 2008 New York Marathon.

1. For all runners who finished in 3.5 hours or less, create a pivot table and corresponding pivot chart of average of Time by Gender. (To get a fairer comparison in the chart, change it so that the vertical axis starts at zero.) For the same runners, and on the same sheet, create another pivot table and pivot chart of counts by Gender. Comment on the results.

Hint: use the filter to select the portion of the dataset which contains the records for the runners who finished in 3.5 hours or less.

1. For all runners who finished in 3.5 hours or less, create a pivot table and corresponding pivot chart of average of Time by Age. Group by Age so that the teens are in one category, those in their twenties are in another category, and so on. For the same runners, and on the same sheet, create another pivot table and pivot chart of counts of these age groups. Comment on the results.

Hint: by clicking on any of the age labels in the pivot table and selecting *Group*, a dialog box will appear which allows you to group the ages according to some criterion.

1. For all runners who finished in 3.5 hours or less, create a single pivot table of average of Time and of counts, broken down by Country. Then filter so that only the 10 countries with the 10 lowest average times appear. Finally, sort on average times so that the fastest countries rise to the top. Guess who the top two are! Comment on the results.

(If you are unsure of how to filter or sort data, please refer the Filtering, Sorting PowerPoint & examples in Excel Basics)

**Exercise 3 (Trend Lines in Scatterplots) – 10 mins**

Each worksheet in the Excel file *LineFit Data.xlsx* contains a set of data that describes a functional relationship between the dependent variable *y* and the independent variable *x*. Construct a scatter chart of each data set, and use the *Add Trendline* tool to determine the best fitting functions to model these data sets.

**Exercise 4 (Trend Lines in Scatterplots) – 5 mins**

The Excel file *Facebook Survey* provides data gathered from a sample of college students.

1. Create a scatter diagram showing the relationship between Hours online/week and Friends.
2. *Add Trendline* tool to determine the best fitting functions to model the data set.
3. Add the R-squared value to the chart and interpret the meaning of the derived value.