ISM/UNEX 270, Winter 2008
Homework 1

Prof. Kevin Ross, kross@soe.ucsc.edu
T.A. Geoff Ryder, gryder@gmail.com

January 9, 2008

Goals: Homework 1 starts out with a review of essential concepts from statistics in Problem 1—we will build on these to do more complicated operations later on. Then in P2 we introduce spreadsheet programming using Excel. Finally, in P3 we solve a managerial decision problem using data envelopment analysis (DEA) with the Excel solver.

Due Date: At the beginning of class on Thursday, January 24 (two weeks).

Assessment: 10% of course grade. Weighting: P1, 3%; P2, 3%; P3, 4%.

1 Statistics Review

We will pass out reference notes in class to help you solve the following problems. Your textbooks from prior courses and online math references may also help. You will need to refresh your memory regarding statistical concepts to be able to use SAS well.

Part A — combinatorial probability. A certain shelf has 5 operations research books and 4 computer science books. If they are continually being used and then put back in random order, find the probability that when they are all back on the shelf, 3 specific operations research books will happen to be together.

Part B — more probability. An urn contains 6 red, 3 green, and 8 blue balls. Three balls are drawn at random without replacement. Compute the probability that at least one is green.

Part C — The cumulative distribution function (CDF) for continuous random variable X is

$$F(x) = \begin{cases} 1 - e^{-3x} & : \quad x \geq 0 \\ 0 & : \quad x < 0 \end{cases}$$

Find the probability density function (PDF) for this random variable, and the probability that $X > 3$.

Part D — joint PDF. Given the joint PDF below, compute the value of the constant k. Find the marginal distribution function for X, and for Y. Are these two random variables independent?

$$f(x,y) = \begin{cases} k(2x+3y) & : \quad 1 < x < 4, \quad 2 < y < 5 \\ 0 & : \quad \text{else} \end{cases}$$

Table 1: Data for regression—the effect of Adwords spending on revenue

obs. (i)	Adwords budget (X)	actual revenue (Y)	predicted revenue ($\hat{Y}$)
	($1000's)	($1000's)	($1000's)
1	40	222	
2	50	335	
3	50	293	
4	65	377	
5	75	459	
6	95	540	
7	90	508	
8	90	492	
9	105	577	
10	115	606	
11	120	584	
12	120	599	

Part E — conditional PDF. Find the conditional PDF of X given Y, or $f(x|y)$, for the following joint distribution:

$$f(x,y) = \begin{cases} e^{-(x+y)} & : \quad x \geq 0, y \geq 0 \\ 0 & : \quad x < 0, y < 0 \end{cases}$$

Part F — Apply the definition of expected value to prove in 2-4 lines that for continuous random variables X and Y, $E[X+Y] = E[X] + E[Y]$.

Part G — Write in mathematical notation a statement of the central limit theorem (as may be found in a textbook), and then briefly summarize in your own words what it means.

2 Introduction to Spreadsheet Programming: Regression

Sometimes we want to analyze the relationship between a continuous dependent variable Y and one or more independent variables X_k. A *regression analysis* identifies a function that can be used to predict Y, when given specific values for the X_k. In this problem, we will fit a *linear regression* function for a vector of observations (values) of a single variable, $\hat{Y}_i = b_0 + b_1 \cdot X_i$. Consider the data in Table 1.

A company that sells technical books online used Google's Adwords to advertise in twelve different market segments over a year's time. These market segments are considered to be about the same size, and the customers in each segment have similar purchasing patterns. An analyst at the company would like to determine the functional relationship between Adwords spending and annual revenue. (She knows revenue is driven by more than just advertising, and this over-simple regression analysis is just a first step.)

Here we will look at the R^2 statistic, or coefficient of determination, that indicates how much of the variation in the dependent variable Y is accounted for by changes in the independent variable X.

$$\begin{aligned} \text{TSS} &= \sum_{i=1}^{n} (Y_i - \hat{Y})^2 \\ Y_i - \hat{Y}_i &= (Y_i - \hat{Y}_i) + (\hat{Y}_i - \hat{Y}) \end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \hat{Y})^2 &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \hat{Y})^2 \\ \text{...or, TSS} &= \text{ESS} + \text{RSS} \\ R^2 &= \frac{\text{RSS}}{\text{TSS}} \end{aligned} \quad (1)$$

This says that the total sum of squares (TSS) equals the sum of two components. Term ESS, unexplained sum of squares, is the amount of variation in the dependent variable Y that is *unexplained* by the independent variable X. Term RSS, or regression sum of squares, give the amount of variation in Y that is explained by X. The R^2 statistic is then computed as Equation 1.

Open Excel, and enter the data in Table 1 into the cell range "A4:C15". Then from the menu options choose Tools-Macro-Visual Basic Editor. Type in the following script:

```
Sub H1P2_Linest()
    Dim X As Variant
    Dim Y As Variant

    X = Range("b4:b15")
    Y = Range("c4:c15")
    Range("e4:f8").Formula = Application.WorksheetFunction.LinEst(Y, X, True, True)
End Sub
```

Compile it from the Debug menu, and try running it. From Excel's Help on LINEST, you can see what the different parameters mean.

Part A —Now, work out which parameters are b_0 and b_1 in the equation $\hat{Y}_i = b_0 + b_1 \cdot X_i$. Use them to populate the range "D4:D15" with values of $\hat{Y}_i$ computed using X_i, b_0, and b_1. Finally, use Excel's graphing function to make a scatterplot of Y and $\hat{Y}$ versus X, and paste it in your worksheet.

Part B —What is the value of R^2 for this regression analysis? Are the Adwords expenditures good predictors of revenue?

Part C —Finally, look up the Excel function TREND, and rewrite the script to use TREND instead of LINEST to populate the range "D4:D15". No plot is needed; just submit a list of your working source code.

3 Data Envelopment Analysis

Download the spreadsheet for this problem, H1P3.xls. Here we will solve a data envelopment-based decision problem similar to the one presented in Fitzsimmons.

- **Goals**
  - become familiar with Excel, and the Excel solver
  - solve a linear programming problem
  - conduct a data envelopment analysis for a service business
- **Tools Needed**

- Fitzsimmons, 5th or 6th edition
- access to Microsoft Excel with its basic solver installed

- **Background** You were recently the manager of a Peaberry Coffee store, one of a chain of fine coffee shops which originated in San Francisco. One day you received a surprise visit from the company president. It turns out that the companys South Bay stores have been underperforming, and she decided to promote a diligent and successful store manager (you) to become the new South Bay Regional Director.

  Your first task is to conduct a data envelopment analysis (DEA) of the companys eight South Bay locations. When it is complete all the stores will be placed in their proper quadrants of the DEA strategic matrix. See page 71 of the text for a description of the matrix. The goal is to classify the locations based on their efficiency and profitability: benchmark, potential, problem, or divest.

  San Francisco headquarters will help you by sponsoring a marketing campaign in your new territory. They have developed a tested marketing strategy aimed at consumers in the 20-to-34 age group, which involves viral marketing, ads on internet sites, local cable and radio ads, and so on. Based on past experience, the response to the campaign should be about $28 in increased annual profit per person among those in the age group that live within a store's zip code.

  Peaberry Coffee currently has four stores clustered in Mountain View, zip code 94041, and four stores in North San Jose, zip code 95134. The stores are expected to divide the additional $28 evenly within each zip code, for a gain of about $7 each. Your job is to complete the H1Template worksheet in the H1.xls workbook: **classify each store in its correct place in the DEA strategic matrix.**

- **Problem Setup**
  Open the file H1P3.xls in your version of Excel, and look at the worksheet titled DEAExamplePage68. Click on the yellow square, E16. Then click on the Tools menu and select Solver. (If the Solver is not present, select Add-ins from the Tools menu and load it.) E16 is the objective function for the linear program. The Solver Parameters window should show: Set Target Cell: E16 Equal To: Max By Changing Cells: C15 : E15 Subject to the Constraints: G21 <= I21 and so on If the solver parameter window is blank, you will need to enter all the information as we discussed in class. A snapshot of the window is given on page 68.

  To compute the DEA efficiencies, substitute one by one the labor and material values as positive numbers in the Inputs line of the spreadsheet, D24 and E24. Then click Solve in the Solver Parameters window to compute the answer. Notice that after each computation the efficiency value in E16 was copied and pasted to the Results Table at the bottom. You can read more about this example problem on pages 66 to 70.

- **Efficiency Analysis**
  In H1P3.xls, open the H1Template worksheet. Cells in gray are more or less completed for you, and your task is to fill in the cells shaded yellow. Note that the relationships among cells are about the same as in the previous example worksheet. The efficiency for each location, or service unit, will show up in the cell with the dark border. When your constraints are entered, complete eight solver runs, and fill in the Efficiency column of the Results Table manually after each run.

- **Profitability Analysis**
  Last time, students were instructed to acquire census data using Excel plugins from StrikeIron (http://www.strikeiron.com). This quarter we saved you the trouble and your sample spreadsheet

already contains the relevant population data in the range D51 to D58. These correspond to the stores in zip code 94041 (Mountain View), and for 95134 (San Jose).

Now observe the Analysis of Profitability table below the Results Table in H1P3.xls. The policy for assigning service units to categories is given below the table: high efficiency is defined as 90% or better of the top benchmark units efficiency, and high annual profit is $75K. For each Peaberry Coffee store, check its efficiency value and column (i) value, and assign it the proper strategic category: benchmark, potential, problem, or divest. Enter its category in its corresponding cell within H52:H58.

Repeat the assessment using column (ii) data, and enter its (post-marketing) DEA category in the right location in I52:I58. Note that the entry for store 1 is done for you: it qualifies as a benchmark location both before and after the marketing campaign. **Are any of the strategic conclusions expected to change after the marketing campaign is done?**