Math 1040 Project 2 - Probability Computer Project

INTRODUCTION:

This lab uses the idea of computer simulation to do statistical analysis. Rather than using real data, we use data generated by a computer as the basis of the analysis. Because a computer can quickly generate large amounts of data, it can simulate statistical studies for large populations. This project should be completed using Excel. In addition to your Excel worksheet, you will need to submit a copy of this worksheet with your answers.

OBJECTIVE:

       I. To simulate the probability of getting a 7 when two dice are rolled.

      II. To simulate the probability that two or more people from a class of 25 people will have the same birthday.

We will give an outline and a procedure with explicit instructions to accomplish the first objective. We give only the outline to assist you in accomplishing the second objective.

I. Before we simulate the probability of rolling a 7 with two dice, we first calculate the theoretical probability. To do this, study and then complete the following table**.**

Roll total	(die 1, die 2)	Number of combinations	Probability
2	(1,1)	1	1/36
3	(1,2) or (2,1)	2	____
4	(1,3), (2,2), (3,1)	3	____
5	_____	___	____
6	_____	___	____
7	_____	___	____
8	_____	___	____
9	_____	___	____
10	_____	___	____
11	(5,6) or (6,5)	2	____
12	(6,6)	1	____
	Total _____		

The probability P(7) of rolling a 7 is _____

The OUTLINES *below are intended as a summary.* The PROCEDURES below list details of what you must do to accomplish the steps in the outlines.

OUTLINE for objective I.

> A. Run the spreadsheet.
> B. Generate a column of 100 rolls of a single die.
> > (We use random integers ranging from 1 to 6).
> C. Generate a second column of 100 rolls of a single die.
> D. Add the column values to get a third column with the roll total of both dice.
> E. Sort the numbers from smallest to largest.

From this information we can count the number of occurrences of the number 7 compared to the total number of rolls. This ratio is the probability of rolling a 7 for this data set.

PROCEDURE to accomplish steps A through E of objective I using Excel:

A. Run Excel.

B. **Excel 2003:** Select 'Data Analysis' from the 'Tools' menu. (If there is no 'Data Analysis' entry in the 'Tools' menu, select 'Add Ins' and check the 'Analysis ToolPak' box.)

   Excel 2007: Select 'Data Analysis' from the 'Data' menu. (If there is no 'Data Analysis', click on the 'Office Button' in the upper left corner of the screen, then select 'Excel Options' at the bottom of the menu that opens. Next choose 'Add-Ins' and click on 'Analysis ToolPak' and then 'Go'. Another menu will appear for 'Add-Ins'. Check 'Analysis ToolPak' and then 'OK'. Finally, you will need to request that this add-in be installed.)

   After you have selected 'Data Analysis', scroll down this list and select 'Random Number Generator'. Click OK. Enter the data in the menu as follows:

Number of variables:	1
Number of Random Numbers:	100
Distribution:	Uniform
Parameters	Between 1 and 7

   Check ( )   Output Range, [click in adjoining field],
   > then enter A1 (or click cell A1).
   Click OK. The random numbers will appear in column 1.

C. Repeat the process to get 100 different random numbers in column B.

We need to have whole integer values rather than decimal numbers. To accomplish this, we must truncate the decimal part.

Select cell C1 and type: =Trunc(A1)   press Enter.
Select cell C1. Use your mouse to point to the lower right corner of the cell. You should see a black +.   Drag the mouse from this position to cell C100. The column will be blue. The whole numbers will now be in column C.

Similarly get integer values from column B into column D.

D. We now add the values of the dice. To accomplish this:

Select cell E1 and type   =C1+D1   press Enter.
Select cell E1. Drag the mouse from cell E1 to cell E100 (as above).

We need these sums to be independent of columns C and D so we can sort them. To accomplish this:

Copy column E.
Select cell F1
Select 'Paste Special'.
A submenu will appear. Check the box 'values' and then 'OK'.

E. Column F now contains values that can be sorted so the 7's will all be together. To accomplish this:

Copy column F.
Select 'Sort' (under 'Data' in Excel 2003).
Check 'Continue with the current selection'. Click 'Sort'.
Click OK (defaults are current column in ascending order). The data will be sorted
          from smallest to largest.

CONCLUSION for objective I:

Examine the sorted values of the dice in column F. Generally there will be only a few 2's with a few more 3's and still a few more 4's. Similarly there will be only a few 12's with a few more 11s etc. Usually there are more 7's than any other value. Your results may vary slightly.

The simulated probability of rolling a 7 is: (Total '7') / 100 = _____.

The simulated probability or rolling a 4 is _____? Rolling a 12 _____?

Your simulated probabilities should be reasonably close to the theoretical probabilities you calculated above.

II. The second objective is to simulate the probability that two or more students in a class of 25 students will have the same birthday. What do you think this probability is? Circle your answer!

a. less than 20%   b. 20% - 40%   c. 40% - 60%   d. 60% - 80%   e. over 80%

We simulate this probability by generating 10 classes of 25 students each. We then compare the number of classes with multiple birthdays to the total number of classes.

OUTLINE for objective II.

Represent birthdays using integers from 1 to 365 (Jan 1 to Dec 31). You will need to:

A. Generate a column of random birth dates for each of the 25 students in a class. Then generate 9 similar columns representing 9 other classes for a total of 10 classes.

B. Determine if there are duplicate birthdays. This is much easier if you sort the numbers. Examine only the whole part (left of the decimal) of the numbers. The answer is 'Yes' or 'No' for each class.

C. Based on this simulation, determine how many of the classes have duplicate birthdays compared to the total number of classes. This is the probability that a class of 25 students will have duplicate birthdays.

PROCEDURE for objective II:

Using the dice simulation as a model, **describe how you *intend* to accomplish steps A, B, and C above**

Step A

Step B

Step C

1. To begin, you will need to click on 'Sheet 2' near the bottom of the Excel worksheet to obtain a clean sheet for generating the birthday data.

2. Using your outline above as a guide, perform the simulation of duplicate birthdays.

   Note: when sorting, check 'Continue with the current selection' before you sort.

CONCLUSION for objective II:

   The probability that a class of 25 students will have duplicate birthdays is: _____.

   Does this probability agree with the estimate you made above? _____.

The theoretical probability of duplicate birthdays is 0.5687. Solving the duplicate birthday problem is somewhat challenging and is not part of this lab assignment. However the theoretical tools needed to solve this problem are covered in the probability section of the text. See if you can do it!

SUBMITTING YOUR PROJECT:

Submit your completed worksheet. Your instructor will also need to see your Excel spreadsheet. If you are completing this project in a computer lab, make sure to show the Excel spreadsheet to your instructor before shutting down your computer. If you are not completing this project in a computer lab, ask your instructor how he/she would prefer that you turn in your Excel spreadsheet.