ISCAM 2: CHAPTER 0 EXERCISES

1. Random Ice Cream Prices

Suppose that an ice cream shop offers a special deal one day: The price of a small ice cream cone will be determined by rolling a pair of ordinary, six-sided dice. The price (in cents) will be the larger number followed by the smaller number. So, rolling a 2 and 5 results in a price of 52 cents, rolling a 4 and 3 produces a price of 43 cents and so on. Use R or Minitab to conduct a simulation analysis of this random process, and then perform an exact analysis of this random process.

(a) Use R or Minitab to simulate 10,000 repetitions of rolling a pair of fair, six-sided dice:

> *R:*   `> die1 = sample(1:6,10000, replace = TRUE)`
> `> die2 = sample (1:6,10000, replace = TRUE)`
>
> *Minitab:* Select Calc > Random data > Integer
> Number of rows to generate: 10000; Store in column(s) die1 die2;
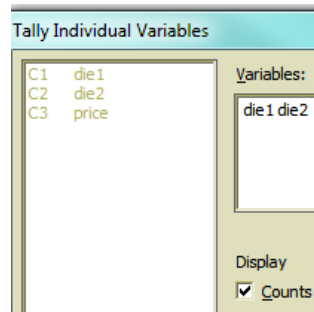> Minimum value 1; Maximum value: 6 > OK.
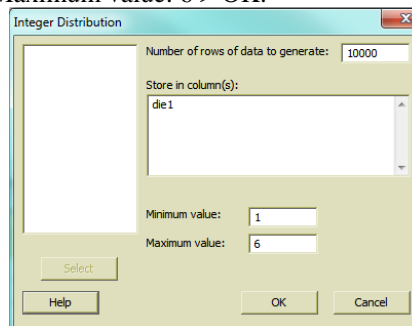
Tally the results of the 10,000 simulated rolls for each die.

> *R:*   `> table (die1); table(die2)`
>
> *Minitab:* Select Stat > Tables > Tally Individual Variables
> Variables: die1 die2 > Check the Counts box under Display. Click OK.

Report the tallies and comment on whether they appear to be consistent with 1–6 being equally likely results of a die roll.

(b) Use software to calculate the 10,000 ice cream prices:

> *R:*   `> price = 10*pmax(die1,die2)  + pmin(die1,die2)`
>
> *Minitab:* Select Calc > Calculator > Store result in variable price;
> Expression: `10*RMAX(die1,die2) + RMIN(die1,die2)` then click OK.

Explain what the command does and why it calculates the price correctly.

(c) Produce a histogram of the 10,000 prices:

> *R:* `> hist(price)`
>
> *Minitab*: Select Graph > Histogram > Simple > OK. Graph variables: price, then click OK.

Submit this graph, and comment on what it reveals about the distribution of these prices.

(d) Calculate and report the average of these 10,000 prices:

> *R:*   `> mean(price)`
>
> *Minitab:*   `MTB> desc 'price'`

Now suppose that you walk into the ice cream shop having only two quarters and no other money.
(e) Approximate the probability that the price is 50 cents or less by determining the proportion of these 10,000 prices that are 50 cents of less:

      R: `> table(price)`
    Minitab: Select Stat > Tables > Tally Individual Variables
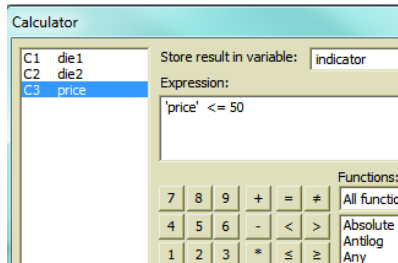   Variables: price
         Check the Cumulative counts box under Display. Click OK.

Or, create an indicator variable that equals 1 if the condition is satisfied, 0 if not, as follows:

      R: `> indi50 = (price <= 50); table(indi50)`
    Minitab: Select Calc > Calculator > Store result in variable indicator;
         Expression: `'price' < = 50` then click OK.

       Then tally the values in the indicator column.

Exact Analysis
(f) To determine the exact probabilities, start by listing all 36 possible and equally likely outcomes (ordered pairs) of rolling two six-sided dice.
(g) Determine the (exact) probability that the price is 50 cents or less by counting how many outcomes produce such a price and dividing by the total number of possible outcomes. Is the approximate probability from your simulation close to this exact probability?
(h) Suppose that I offer to pay for your ice cream cone if the price turns out to be an odd number. Determine the (exact) probability that I pay for your ice cream cone.

Different Dice
You will now investigate how much difference it would make to use a pair of fair, *ten*-sided dice that contain the numbers 0–9 on the ten sides. The rule for determining the price is the same: larger number followed by smaller number (in cents).
(i) Use software to simulate 10,000 repetitions of rolling a pair of fair, *ten*-sided dice containing the numbers 0–9 on the ten sides. (*Hint*: Use essentially the same instructions as before, with one or two small adjustments.) Produce and submit a histogram of the resulting prices. Also determine and report the average price. Comment on how this average price has changed from using six-sided dice. Is the change what you would have expected? Explain briefly.
(j) Use your simulation results to approximate the probability that the price is 50 cents or less. Then determine the exact, theoretical probability that the price is 50 cents or less. (*Hint*: You do not have to submit a complete list of all possible outcomes, but you do need to consider all possible outcomes. Be sure to explain/show how you determine these probabilities.) Comment on how this probability has changed from using six-sided dice.

2. Consecutive Lottery Numbers
The California Lottery's SuperLottoPlus game involves picking five numbers between 1 and 47.
Many people who play the lottery refuse to choose consecutive integers among their five number choices, believing that it's very unlikely for consecutive numbers to appear? But how unlikely is it? We will use simulation to approximate this probability.
(a) First, make a *guess* for the probability that consecutive numbers appear.

Simulation Analysis:
(b) Use R or Minitab to conduct a simulation of the California game by randomly selecting 5 numbers (without replacing) from the numbers 1 – 47, repeating this process 25 times.

> *R:*   `winner <- sample(1:47,5)`
> *Minitab:* To sample *without replacement* in Minitab, you need to create a column of the values you want to sample from:
> ```
>         MTB> set c1
>         DATA> 1:47
>         DATA> end
> ```
> Then you can randomly sample from that column, saving the results into another:
> ```
>          MTB> sample 5 c1 c2
> ```

Repeat this process a total of 25 times, (you can simply copy-and-paste the commands above over and over), keeping a running tally of whether or not the (sorted) set includes consecutive integers.
Report the proportion of these 25 drawings in which the numbers include at least one set of consecutive integers.
(c) Now we want to automatic this process to run it a very large number of times.

> *R:*
> ```
> > winner = 0; y =0; diff=0
> > for (j in 1:1000){
>     winner <- sample(1:47,5)
>     sortwin = sort(winner)
>     for (i in 1:4) diff[i] = sortwin[i+1]-sortwin[i]
>     y[j] = (min(diff)==1)
> }
> table(y)
> ```

> *Minitab*: Save the worksheet you had been using (with C1) into a specific directory. Save the `lottery.mac` file from the textbook page into the same directory. (You can view this file using a text editor.) The main commands are:
> ```
>         Do k1=1:1000
>             sample 5 c1 c2
>             sort c2 c3
>             diff c3 c4
>             let c5(k1)=(min(c4)=1)
>         endDo
> ```
> Then in Minitab, you should be able to execute the macro:
> ```
>         MTB> %lottery.mac
>         MTB> tally c5
> ```

Write a description of what each commands does, and explain how the computer achieves the goal of determining whether a drawing produces at least one set of consecutive integers.
(d) From your results in (c), report the estimate of the probability that this lottery game produces at least one set of consecutive integers. How does this compare to your prediction in (a)?

Exact Analysis:
(e) How many different drawings are possible? In other words, how many ways are there to choose five numbers from 47 numbers?

(f) It can be shown that when choosing k numbers between 1 and n, the number of ways to get consecutive integers is $\binom{n}{k} - \binom{n-k+1}{k}$. Use this result to determine the number of different drawings that include consecutive integers with the California game.

(g) Use your answers to (e) and (f) to determine the theoretical probability that the five winning numbers in the California game will include consecutive integers.

(h) How does the theoretical probability compare to your simulation results in (d)?

(i) Even though picking consecutive numbers won't increase your chances of winning in any one drawing, explain another advantage to doing so considering other players' tendencies.

3. Roulette

An American roulette wheel has 38 slots: 18 contain black numbers, 18 red numbers, and 2 green numbers. The wheel is spun and the ball falls at random into one of the 38 slots. If you bet $1 on a color (either red or black) and win (the ball lands on the color you picked), you receive $2 for a net gain of $1. If you lose, your "net gain" is -$1. If you bet $1 on a number (1–36) and win, you receive $36 for a net gain of $35. Let the random variable X denote your *net* gain from one bet on a color, and let Y be your *net* gain from one bet on a number.

(a) Specify the probability distribution of X (net gain in a color bet). [*Hint*: List the (two) possible values and their probabilities (to four decimal places), where the 38 slots are equally likely to occur.]

(b) Perform a simulation of this probability distribution. First, set up the probability distribution (outcomes and probabilities) and then sample with replacement from that distribution.

R:
```
> x= c(-1,1); prob=c(.5263, .4737)
> outcomes=sample(x, 1000, replace = TRUE, prob)
```
Minitab: Enter the values -1 and 1 into C1 and their respective probabilities (.5263, .4737) in the corresponding rows in C2. Also make sure you have "enabled the command prompt" (Editor > Enable Commands).

```
MTB> random 1000 c3;        Simulates 1000 spins of roulette wheel using the
SUBC> discrete c1 c2.        probability distribution specified by c1 and c2
```

Then look at the total "gain" for these 1000 bets and the average net gain among these 1000 bets.

R:
```
>  sum(outcomes); mean(outcomes)
```
Minitab:
```
MTB> sum c3
MTB> mean c3
```
Report these values.

(c) Example the behavior of the average gain as you increase the number of spins:

R:
```
> spinnum=1:1000
> avgcolor = cumsum(outcomes)/spinnum
> plot(spinnum, avgcolor, type="l")
```
Minitab:
```
MTB> set c4
DATA> 1:1000
DATA> end
MTB> let c5=parsum(c3)/c4
MTB> name c4 'spin number' c5 'cum avg'
MTB> plot c5*c4;
SUBC> connect.
```

Include and describe the behavior of this graph. Explain what these commands have done (*Hint*: You may want to look at the cumsum or parsum output directly to see what that command does.)

(d) Now calculate the theoretical expected value of X. How does this result compare to your simulation results? [Show your work and use the proper notation, e.g., $E(X)$.]

(e) Write an interpretation of the expected value calculated in (d). (*Hint*: Use the results from (c).)

(f) In a similar manner, specify the probability distribution of *Y*, produce a simulation of 1000 (or more) spins, and examining the behavior of the average gain across the numbers of spins, and calculate the theoretical expected value of *Y*.

(g) How do the two expected values for the two bets (color vs. number) compare? Does this mean that the two bets should be considered equivalent? Explain.

(h) Now determine and compare the theoretical *variance* of the gain for the two bets. How do they compare? (*Hint*: Create a graph as in (c) and consider how the variability in the two sets of results compares.)

4. Early Die Game
In the 1600s, French nobleman Chevalier de Mere and mathematician Blaine Pascal developed formal study of probability. One of the first questions they looked at was "which is more likely, rolling at least one six in four rolls of a die, or rolling at least one pair of double sixes in 24 rolls of a die." Use simulation and/or probability rules to answer this question.