Math 214
Applied Statistics

Laboratory Project #3
Due: Wednesday February 27

Part I: Using R to SIMULATE Experiments

Although Einstein said that God does not play dice, R can play dice (and cards) rather easily. In this Lab assignment we will learn how to make R roll die and select cards. We can then have R quickly and conveniently play games MANY, MANY times. The output of the games can be analyzed statistically.

The main R command we will use is `sample`. For example, suppose we wanted to roll one die 10 times. Since the output of any die roll is a number between 1 and 6, we could try:

```
> sample(1:6,10,replace=T)    %Select a number from 1-6, ten times w/replacement
[1] 2 1 5 2 1 6 4 3 3 6
```

or we could write a function:

```
> Roll1Die = function(n) sample(1:6,n,rep=T)
> Roll1Die(20)
 [1] 2 3 3 5 4 2 1 3 6 1 4 6 2 1 3 2 1 5 1 4
```

Now we can 'Roll1Die' as many times as we want.

Suppose we want to flip 3 coins and count the number of heads that appear. How would you tell R to flip a coin 15 times? Try

```
> sample(0:1,15,rep=T) #Pick either 0=Tails or 1=Heads)
 [1] 1 1 0 1 0 1 1 1 0 0 0 1 1 1 0
```

or, perhaps more useful, write a function:

```
> Flip1Coin = function(n) sample(0:1,n,rep=T)
> Flip1Coin(30)
 [1] 1 1 1 1 0 0 0 1 0 0 0 0 1 0 0 1 1 0 1 0 0 0 1 0 1 1 1 1 1 1
```

How many Heads appeared in these 30 experiments? (use R's `sum` command) Is this number what you would expect? Try flipping 1,000 coins. How many heads occurred? How does this relate to the 'Empirical Definition' of probability?

Part II: Probability Distributions and Histograms

We know that a Probability Distribution (or Probability Function) assigns a probability to each value of the Random Variable. Now we want to investigate how well a Theoretical Probability Function fits with actual data. For example, suppose our experiment consists of rolling two die and adding the results. We can figure out the probability of getting any of the possible outcomes: P(2) = 1/36, P(3) = 2/36, ..., P(7) = 6/36, P(8) = 5/36 ...

We can also easily *simulate* this experiment using R. One way to get the results from one hundred experiments would be

```
> TwoDie = Roll1Die(100) + Roll1Die(100)
```

To look at the results, try a histogram

```
> hist(TwoDie,breaks=c(1.5:12.5)) #Centers intervals on 1,2,....12
```

What percentage of the 100 experiments resulted in 7? How does this compare with the Probability Function?

We can easily make an Empirical Probability Function from the data, all we have to do is tell R to plot the relative frequency instead of the frequency on the histogram.

```
> hist(TwoDie,breaks=c(1.5:12.5),prob=T) #Plot Relative Frequency
```

To Do:

1. Use R to conduct the experiment 50 TIMES. (ie Roll two dice fifty times)

2. Plot and print the resulting Relative Histogram.

3. On the histogram, sketch the theoretical Probability Function, P(X).

4. How well do the two compare?

5. Repeat the process using 500 experiments. Comment.

6. Repeat the process using 5000 experiments. Comment.

Part III: Mean of a Random Variable

We know that the Probability Function, $P(X)$ can be used to calculate the *mean* of a Random Variable. For Discrete Probability Distributions, the mean of a random variable X is written as

$$\mu = \sum X_i P(X_i) \quad .$$

For example, compute the mean (or expected value) in the Two Dice experiment

$$\mu = 2 \cdot 1/36 + 3 \cdot 2/36 + \ldots + 12 \cdot 1/36 = 7$$

To Do: Find the experimental (sample) mean for the Two Dice experiment using R. For example, to roll two dice and add up the results 100 times:

```
> TwoDie = Roll1Die(100) + Roll1Die(100)
```

The experimental mean value is simply:

```
> mean(TwoDie)
[1] 7.32
```

1. Try this several times - Does the experimental mean change?

2. Compare the mean value as computed above to the mean of the simulated data when you take 50, 500, 5000 *realizations* of the experiment in R. Comment on what happens as the number of experiments increases.

3. *Extra Credit:* Pick one value of n, say $n = 50$ and run 50 experiments 10 or 20 times. Record the value of the sample mean each time. Plot a histogram of these mean values. What is the mean of the sample means? Explain this odd statement:

"The sample mean is a Random Variable."

Problems - Questions

1. Four fair coins are tossed simultaneously and the number of HEADS is recorded each time.

    (a) Use your knowledge of Probability to construct the Probability function for this experiment. (What is the Random Variable? What is its Range?) Make a chart showing X and $P(X)$.

    (b) Use R to simulate conducting this experiment 50 times.

    (c) Compare your theoretical $P(X)$ to the relative frequency histogram of the simulated experiments.

    (d) Use R to simulate 1,000 four-coin experiments. Again, compare the relative frequency histogram you get to $P(X)$.

    (e) Calculate the theoretical mean, μ. Compare this to the sample mean when you find when simulating 50, 100 and 1,000 realizations of the experiment.

2. A card shark you know proposes playing the following game with you: Each card in the deck has a *value*: Aces are worth One, Number cards have the value of the number, Picture cards (Jack, Queen, King) are worth 10 points. She will deal you one card from a deck of 52. If the card has a *value* greater than, or equal to 8, you must pay $1.00. If the card has a value less than 8, she will pay you $1.00.

    You decide to think about it before playing and saunter home to your R-console to do some simulations.

    (a) Use Probability to construct the Probability Function for this game. (Ie: Find the probability of being dealt a card with value=1,2,3,....,10).

    (b) Find the theoretical probability of Winning/Losing.

    (c) Use R to simulate playing the game. Have R play the game 1,000 times for you. How many times did you win? How many times did you lose?

    (d) Plot the relative frequency histogram of the simulation and compare it to $P(X)$.

    (e) Use $P(X)$ to find the mean value (expected value) of the Game. Compare this to the sample mean of the 1,000 experiments you simulated.

    (f) Would you play the Game?

    (g) Explain how you would decide whether to play the game if the dealer paid you $2.00 every time you won and only took $1.00 when you lost.