Problem Statement:

The game of *Unlucky 7* is one of the most famous (or notorious) of all gambling games played with dice. In this game, the player rolls a pair of six-sided dice (assumed to be fair), and the SUM of the numbers that turn up on the two faces is noted. If the sum is 7 or 11 on the first roll, then the player wins immediately. If the sum is 2, 3, or 12 on the first roll, then the player loses immediately. If any other sum appears, then the player continues to throw the dice until the first sum is repeated (the player wins) or the sum of 7 appears (the player loses).

Requirements:

1. Compute the theoretical probabilities for each sum of a dice roll (rounded to the nearest thousandth decimal place). Use for the simulation model.
   - Since the random variable is the sum of the two fair die, we need to create a probability table that correlates to the probability of the sums 2-12, inclusive.
   - From prior knowledge, I know that this particular probability distribution is binomial, and the total number of ways is 6^2 because there is six possibilities on a fair dice, and since we are rolling two at the same time, it is the exponent.

RV = SUM	P(SUM)	Decimal Approximation
2	1/36	.028
3	2/36	.056
4	3/36	.083
5	4/36	.111
6	5/36	.139
7	6/36	.166
8	5/36	.139
9	4/36	.111
10	3/36	.083
11	2/36	.056
12	1/36	.028

   - The decimal approximation for P(SUM = 7) is rounded down one thousandths place so the sum of all the decimal approximation values are equal to one, and not above one.
   - Hence, now the total of the decimal approximations are 1.
2. Design a five-step simulation experiment for the game of *Unlucky 7*. Describe each of the five steps completely. Conduct 20 trials. Record the results.
   - Problem:
     i. To simulation 20 trials of the game *Unlucky 7*.
   - Key Components:
     i. Upon the first roll, if the player receives a 7 or an 11, then they win immediately. If they receive a 2, 3, or 12, then they lose immediately.
     ii. If any other sum appears than the ones listed above, the player must keep on playing until the first sum is repeated, in which they win, or the sum of 7 appears, in which case they lose.
     iii. All trials will be of different lengths.
     iv. The sums are ranged from 2 to 12, inclusive.
   - Underlying Assumptions:
     i. The dice are fair.
     ii. The person playing and the person running the game are not cheating.
   - Simulation Model:

i. Based upon the theoretical decimal approximations, I will assign the numbers 000-999 a certain sum value.
ii. SUM = 2 → 000-027
iii. SUM = 3 → 028-083
iv. SUM = 4 → 084-166
v. SUM = 5 → 167-277
vi. SUM = 6 → 278-416
vii. SUM = 7 → 417-582
viii. SUM = 8 → 583-721
ix. SUM = 9 → 722-832
x. SUM = 10 → 833-915
xi. SUM = 11 → 916-971
xii. SUM = 12 → 972-999

- Conclusions:
  i. Conclusions will be made in the next few requirements.
- Trials:
  i. Trial 1
    - 482—7
    - WIN
  ii. Trial 2
    - 266—5
    - 967—11
    - 571—7
    - LOSE
  iii. Trial 3
    - 564—7
    - WIN
  iv. Trial 4
    - 589—8
    - 169—5
    - 96—4
    - 807—9
    - 891—10
    - 78—3
    - 869—10
    - 619—8
    - WIN
  v. Trial 5
    - 98—4
    - 846—10
    - 304—6
    - 000—2
    - 955-11
    - 246—5
    - 326—6
    - 882—10
    - 965—11
    - 839—10
    - 818—9
    - 762—9
    - 704—8

- 833-10
- 95—4
- WIN

vi. Trial 6
   - 984—12
   - LOSE

vii. Trial 7
   - 795—9
   - 901—10
   - 471—7
   - LOSE

viii. Trial 8
   - 480—7
   - WIN

ix. Trial 9
   - 125—4
   - 837—10
   - 498—7
   - LOSE

x. Trial 10
   - 764—9
   - 801—9
   - WIN

xi. Trial 11
   - 941—11
   - WIN

xii. Trial 12
   - 621—8
   - 082—3
   - 628—8
   - WIN

xiii. Trial 13
   - 167—5
   - 566-7
   - LOSE

xiv. Trial 14
   - 987—12
   - LOSE

xv. Trial 15
   - 224—5
   - 038—3
   - 390—6
   - 969—11
   - 943—11
   - 810—9
   - 911—10
   - 633—8
   - 608—8
   - 427—7
   - LOSE

xvi. Trial 16

        - 549—7
        - WIN
    xvii. Trial 17
        - 949—11
        - WIN
    xviii. Trial 18
        - 726—9
        - 790—9
        - WIN
    xix. Trial 19
        - 709—8
        - 171—5
        - 215—5
        - 249—5
        - 277—5
        - 178—5
        - 590—8
        - WIN
    xx. Trial 20
        - 440—7
        - WIN
  - There was 13 WINS and 7 LOSSES.

3. Create a relative frequency histogram to summarize the population distribution for the sum of the dice roll.

4. Answer the following questions.
    - What type of random variable is represented in this game?
        i. This game represents a discrete random variable. The random variable is the sum of the two fair die.
    - What is the shape of the distribution?
        i. The distribution has a standard normal shape—bell shaped, single-peaked, area under curve is 1. Furthermore, this distribution is binomial further ensuring that is symmetrical.
    - What sum appears most likely to occur? What sum appears least likely to occur?

i. The sum of 7 is most likely to appear, since it has the highest probability (6/36). The sums of 2 and 12 are the least likely to occur since they have the lowest probability (1/36).

- What is the relative frequency of winning on the first roll of the dice? What is the relative frequency of losing on the first roll of the dice?
  i. To win immediately (upon the first roll of the dice) then one must receive either the sum of 7 or 11.
     - The probability of rolling a 7 is (6/36), and the probability of rolling an 11 is (2/36).
     - Since these are independent of each other, simply add the probabilities together. (6/36) + (2/36) = (8/36) ≈.222
  ii. To lose immediately (upon the first roll of the dice) then one must receive either the sum of 2, 3, or 12.
     - The probability of rolling a 2 is (1/36), the probability of rolling a 3 is (2/36), and the probability of rolling a 12 is (1/36).
     - Since these are independent of each other, simply add the probabilities together. (1/36) + (2/36) + (1/36) = 4/36 = 1/9 ≈ .111

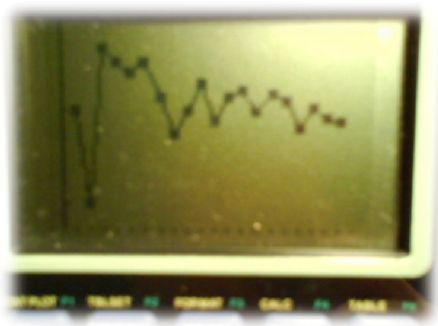
5. Calculate the mean the variance of the random variable based upon the simulation experiment. Compare the results with the theoretical mean and variance.
   - To get the probability table from the simulation experiment, count how many times the sum of 2, 3, 4… 12 were rolled.
   - In my simulation, a total of 67 rolls were thrown to complete all twenty trials.

RV = SUM	P(SUM)
2	1/67
3	3/67
4	4/67
5	10/67
6	3/67
7	10/67
8	9/67
9	9/67
10	9/67
11	7/67
12	2/67

- To calculate the mean (of the simulation experiment), use the formula $\sum x_i * p_i$. The x_i is the random variable, and the p_i is the probability associated with that random variable.
  - $\mu = 7.791$
- To calculate the variance (of the simulation experiment), use the formula $[\sum x_i^2 * p_i] - \mu^2$
  - $\sigma^2 = 4.001$
- To calculate the mean (of the actual probability table), you utilize the same formula above.
  - $\mu = 7$
- To calculate the variance (of the actual probability table), you use the same formula above.
  - $\sigma^2 = 5.833$
- My simulation experiment mean is higher than the theoretical mean, as opposed to the simulation variance which is smaller than the theoretical variance.

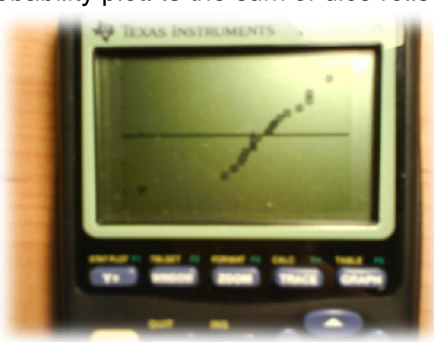
6. Create a list of sample mean sums as the number of dice rolls increase from 1 to 20. Plot a frequency polygon of observations (number of dice rolls) versus sample mean. Discuss the law of large numbers.

Roll #	**Average**	**Roll #**	**Average**
1	7	11	76/11 ≈ 6.90
2	6	12	43/6 ≈ 7.17
3	23/3 ≈ 7.67	13	94/13 ≈ 7.23
4	7.5	14	7
5	7.4	15	7.2
6	7.5	16	7.125
7	50/7 ≈ 7.14	17	116/17 ≈ 6.82
8	6.75	18	127/18 ≈ 7.06
9	7	19	132/19 ≈ 6.95
10	7.3	20	6.9



The law of large number states that as the number of trial increases, the mean (average) tend to go towards the theoretical (expected) mean. As exemplified by the frequency polygon shown above, this law holds true for this binomial probability because the frequency polygon appears to be approaching 7, and at certain times, is at seven (as shown in the table above also). Hence, as more dice are rolled, the average should eventually be extremely close to seven.

7. Construct a normal probability plot. Is the sum of dice rolls normally distributed?



- By the normal probability plot, shown below, the data appears to be relatively linear (it has a slight curvature) thus showing that the sum of dice rolls is somewhat normally distributed, but not definitively.

Laboratory #3
Random Variables

Pooja Mehta
5th period
AP/IB Statistics
Mr. LeKander
12-14-2009