Chapter 17: The Expected Value and Standard Error

1. What do the following experiments have in common? How can you analyze them?

   a) Toss a coin n times and count the number of heads.

   b) Spin a roulette wheel n times and each time bet on red and count how much you win or lose.

   c) Randomly draw a sample of voters. From this sample, estimate the percentage of Democrats who voted. Estimate the accuracy of your estimate.

   d) Roll a pair of dice 100 times and count the number of double sixes.

Box Models

2. Draw 100 times with replacement from the box [3 , 3 , 5 , 13].

   a) How small can the sum of the draws be? How large?

   b) How many times do you expect to draw a 3 ?

   c) What do you expect the sum of the draws to be?

   d) If the sum of the draws = expected value + chance error, how big is the chance error likely to be? What does it depend on?

Expected Value and Standard Error

Suppose you randomly draw n times with replacement from a box and consider the sum of the draws:

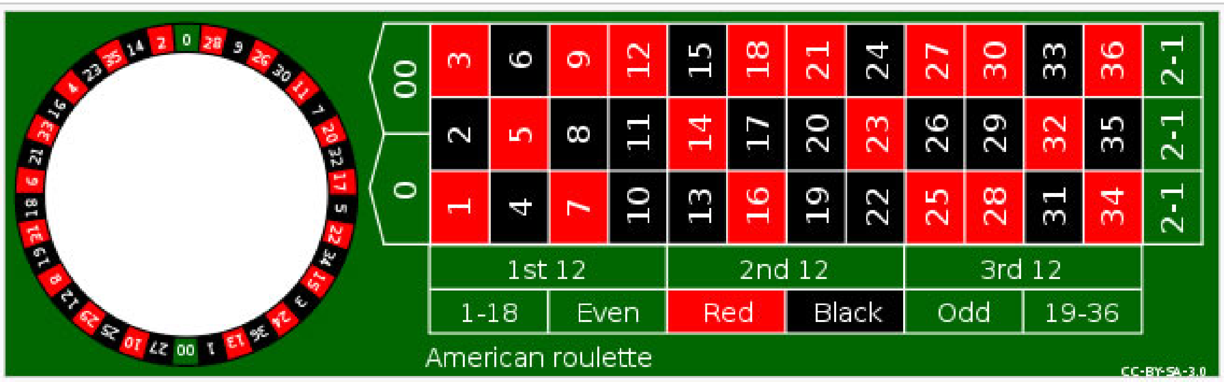
EV for the sum of the draws = Box AV x n

The sum of the draws = EV for the sum + chance error

SE for the sum of the draws = Box SD x $\sqrt{n}$

The likely size of the chance error is the standard error or SE. The SE applies to chance variability; the SD applies to a list of numbers.

3. A multiple choice test has 100 questions. Each question has 5 possible answers, one of which is correct. Suppose you guess on each question. How many do you expect to get right? How should it be scored so that your expected score is zero?



American roulette layout and wheel, with single and double zero

4. You bet $ 10 on red in 100 spins of the roulette wheel and compute how much you win. What is the box model? How much should you expect to win?

5. A pair of dice are rolled. If you get two "aces" you win $ 44 ; otherwise you lose $ 1. You repeat this game 360 times. Determine the appropriate box model. How much do you expect to win or lose?

6. Consider the sum of 25 draws from the box [0 , 2 , 3 , 4 , 6] .

a) Find the box AV.

b) Find the box SD.

c) Find the EV for the sum of the draws.

d) Find the SE for the sum of the draws.

e) The sum of the draws = EV for the sum + chance error. Find the likely size of the chance error.

Short-Cut Formulas for SDs:

1. Zero-One Boxes: Suppose the box contains only 1s and 0s and the proportion of 1s is equal to p and the proportion of 0s is equal to $1-p$.

   Box SD $=\sqrt{p \times (1-p)}$.

2. L -S Boxes: Suppose the box contains only two different numbers, Ls and Ss with $L > S$, and the proportion of Ls is equal to p and the proportion of Ss is equal to $1-p$.

   Box SD $=\left(L-S\right)\sqrt{p \times (1-p)}$.

Example: You play a game in which you roll a die 10 times. Each time a "6" occurs, you win $10, otherwise you lose $1. How much do you expect to win? Give or take how much?

Example: A multiple-choice test has 20 questions, each with 4 possible choices. Each correct answer is worth 5 points, and for each incorrect answer you lose 2 points. If you guess all the answers, what do you expect your test score to be? Give or take how much?

Example: Bet $1 on "17" 100 times in roulette. How much do you expect to win? Give-or-take how much?

Classifying and Counting

• If we want to count how many times something happens, the box has 0's and 1's. The 1's represent the thing we are counting, and the 0's represent everything else.

• Here is the box for rolling a die and counting the total number of spots we get:

1	2	3	4	5	6

• Here is the box for rolling a die and counting how many "6"s we get:

0	0	0	0	0	1

The Normal Curve

- When drawing at random with replacement from a box, the sum of the draws will follow the normal curve, even if the contents of the box do not. The sum of the draws must be put into standard units and the number of draws must be reasonably large.

- 68% of the time the **sum of the draws** will be within one *standard error* of the *expected value*.

- 95% of the time the **sum of the draws** will be within two *standard errors* of the *expected value*.

- We can use the normal curve to find the **chance** that the **sum of the draws** is between two values. We use *expected value* and *standard error* to get standard units.

- It does not matter what is in the box.
- The histogram for the tickets in the box does not have to follow the normal curve.
- The **sum of the draws** will follow the normal curve, even if the tickets in the box are 0’s and 1’s!
- In fact, we don’t even need to know what is in the box, we just need to know the average and the SD of the box.

Example: Bet $1 on "red" 100 times in roulette. What is the chance you win more than $10? What is the chance you lose more than $10? What is the chance you come out ahead?

Sampling without replacement from a LARGE population is just like sampling with replacement.

Example: A large crop of apples has an average weight of 4.3 oz with an SD of 1.5 oz. You choose 100 apples at random. What's the chance the total weight is less than 25 pounds?

Example: Suppose 10% of people in a large population are "underweight". If we take a random sample of 1000 people from this population, what is the chance that more than 103 will be "underweight"?