

**Stat 321 - Day 12**  
**Expected Value and Variance**

**Example: Random Babies and College Committees (cont.)**

Recall that you found the probability distribution for the random variable  $X$  = number of matches to be:

$X$	0	1	2	4
$P(X=x)$	9/24	8/24	6/24	1/24

Earlier you found the *expected value* of  $X$  to be  $(0)(9/24) + (1)(8/24) + (2)(6/24) + (4)(1/24) = 1$ . You also found that when you simulated this process a large number of times, the average number of matches closely approximated this expected value.

The *expected value* of a discrete random variable  $X$  is denoted and defined by  $\mu = E(X) = \sum x_i P(X=x_i)$ , where the sum is taken over all possible values  $x_i$ . It is interpreted as the long-term average value that would result if the process were repeated over and over.

Recall that the probability distribution of  $Y$  = number of women chosen when two people are selected from a committee of four men and two women is:

$y$	0	1	2
$P(Y=y)$	6/15	8/15	1/15

(a) Determine the expected value of  $Y$ .

(b) Do you literally “expect” to see this many women when two people are chosen? Is it even possible? Explain what “expected value” means in this context.

(c) Consider a new random variable:  $Z=5Y-2$ , where  $Y$  is the same random variable as above. Determine the probability distribution of  $Z$ , and then find the expected value of  $Z$ .

(d) Does  $E(Z)=5E(Y)-2$ ?

(e) Consider a new random variable:  $W=Z^2-2$ , where  $Z$  is the same random variable as above. Determine the probability distribution of  $W$ , and then find the expected value of  $W$ .

(f) Does  $E(W) = [E(Z)]^2 - 2$ ?

The expected value of a function  $h(X)$  of a random variable  $X$  is  $E[h(X)] = \sum h(x_i)P(X=x_i)$ . In general it is not necessarily true that  $E[h(X)] = h[E(X)]$ , but this does always hold with a linear function:  $E(aX+b) = aE(X)+b$ .

**Example: 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 (red or black) and win, you receive \$2 for a net gain of \$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 winnings from one bet on a color, and let  $Y$  be your net winnings from one bet on a number.

(g) Determine the probability distribution of  $X$ . [*Hint*: List the possible values (there are two of them) and their probabilities.]

(h) Calculate the expected value of  $X$ , and interpret what it means in this context.

(i) Determine the probability distribution of  $Y$ .

(j) Calculate the expected value of  $Y$ , and interpret what it means in this context.

(k) How do the expected values of the two bets compare? Does this mean that the two bets are identical? Explain.

**Def:** The *variance* of a random variable is denoted and defined by  $\sigma^2 = V(X) = E[(X-\mu)^2]$ . It can be calculated as  $\sum(x_i-\mu)^2P(X=x_i)$ . It measures the spread (variation) that we expect to see in the outcomes. A shortcut formula for calculation purposes is  $V(X) = E(X^2)-[E(X)]^2$ . One can show that  $V(aX+b) = a^2V(X)$ . The *standard deviation*  $SD(X)$  is the square root of  $V(X)$ .

(l) Which type of bet do you expect to have the larger variance? Explain, without doing any calculations.

(m) Calculate the variance of the net winnings for each type of bet. [Use the shortcut formula.] Also calculate the standard deviation of the net winnings for each type of bet.

color bet:

number bet:

(n) Which type of bet has the larger variance and standard deviation? Explain why this makes intuitive sense. Does this agree with your prediction in (l)?

Simulation:

To investigate what these expected values and standard deviations mean, you will perform a Minitab simulation of betting 10,000 times with this strategy.

(o) Start with the “color” bet by putting the values -1 and 1 into c1 and their respective probabilities .5263 and .4737 into c2. Then simulate 10,000 repetitions of this bet:

```
MTB> random 10000 c3;  
SUBC> discrete c1 c2.  
MTB> name c3 'netwincolor'
```

Now look at a tally and at descriptive statistics:

```
MTB> tally c3
MTB> describe c3
```

Record the number of -1's, the number of +1's, the mean, and the standard deviation:

```
# of -1's:          # of +1's:          mean:          std. dev.:
```

(p) Are the tallies close to what the probabilities predict? Is the mean close to the expected value? Is the standard deviation close to its theoretical value?

(q) To look at how the mean changes over the 10,000 repetitions, calculate the cumulative sums of the net winnings and the means after each repetition:

```
MTB> parsum c3 c4
MTB> set c5
DATA> 1:10000
DATA> end
MTB> let c6=c4/c5
MTB> name c6 'cumavgcolor'
MTB> plot c6*c5;
SUBC: connect.
```

Comment on how the mean (of the net winnings) changes over time and indicate the value to which it appears to be converging.

(r) Repeat this simulation analysis for the “number” bet by first putting its values into c11 and the corresponding probabilities into c12. Then proceed as above, using c13-c16 instead of c3-c6. Comment on your findings.

(s) Finally, compare the two bets in terms of their cumulative mean winnings by plotting both on the same scale:

```
MTB> plot c6*c5 c16*c15;
SUBC> overlay;
SUBC> connect.
```

Comment on similarities and differences between the long-term performance of the two bets. [Be sure to identify which is which on the graph.]