Activity: Mathematical Expectation

Concepts: Mathematical Expectation for discrete random variables. Includes expected value and variance.

Prerequisites: The student should be familiar with discrete random variables and with some of the standard distributions of discrete random variables.

Definition: The *expected value* of a discrete random variable X is denoted and defined by $\mu = E(X) = \Sigma x_i P(X=x_i)$, 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. Note that this is a special case of the definition on page 119 of your text.

Scenario: College Committee Formation II

Recall our college committee formation scenario: A college professor found herself to be one of two women on a six-person committee. The committee needed to choose two representatives to speak for the committee to other groups, and both women were selected. Previously we investigated how unusual this would be if the two representatives had in fact been chosen at random. Now let the random variable X be the number of women chosen when two people are selected from a committee of four men and two women.

(a) How is X distributed? You should state the name of the distribution, give all relevant parameters, and determine the formula for the probability mass function.

(b) Find the values of the probability mass function for all values of X (i.e. determine the probability distribution of X). Fill in the table below with these values.

X	0	1	2
$f(x)=P(X=x)$			

(c) Determine the expected value of X.

(d) 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.

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

(f) Does E(Z)=5E(X)-2?

(g) 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.

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

<u>Definition</u>: The expected value of a function h(X) of a random variable X is E[h(X)] = $\Sigma h(x_i)P(X=x_i)$. Note that this is the definition given on page 119 of your text.

(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. See if you can use the properties in Theorem 3.2-1 on page 121 of your text to prove that E(aX+b)=aE(X)+b.

<u>Scenario:</u> 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.

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

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

(c) Determine the probability distribution of Y

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

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

Definition: The *variance* of a random variable is denoted and defined by $\sigma^2 = V(X) = E[(X-\mu)^2]$. It can be calculated as $\Sigma(x_i-\mu)^2P(X=x_i)$. It measures the spread (variation) that we expect to see in the outcomes. The *standard deviation* SD(X) is the square root of V(X).

(f) Use the properties of mathematical expectation to show that a shortcut formula for calculation purposes is $V(X) = E(X^2)-[E(X)]^2$ where X is an arbitrary random variable. Hint: Expand $(X-\mu)^2$ in $E[(X-\mu)^2]$ and use the properties of expected value in Theorem 3.2-1 of your text. Double Hint: This is shown on the bottom of page 123 of your text. Try to get as far as you can before you look at the answer.

(g) Show that $V(aX+b) = a^2V(X)$ where X is an arbitrary random variable. Hint: Use the properties of expected value in Theorem 3.2-1 of your text. Double Hint: This is shown on the bottom of page 125 of your text. Try to get as far as you can before you look at the answer.

(h) Calculate the variance of the net winnings for each type of bet in the Roulette scenario. Which is larger? Explain why this makes sense. Also calculate the standard deviation of the net winnings for each type of bet.

Roulette Simulation:

To investigate the meaning of these expected values and standard deviations, you will perform a Minitab simulation of betting 1000 times with these strategies.

(i) 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 1000 repetitions of this bet:

```
MTB> random 1000 c3;
SUBC> discrete c1 c2.
MTB> name c3 'netwincolor'
```

Note that the above commands are simulating playing Roulette 1000 times and each time making a "color" bet. 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.:

(j) 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?

(k) To look at how the mean changes over the 1000 repetitions, calculate the cumulative sums of the net winnings and the means after each repetition:

```
MTB> parsum c3 c4
```

This gives the total amount won (i.e. the cumulative amount of the net winnings on each play) after the nth play of the game for n up to 1000.

```
MTB> set c5
DATA> 1:1000
DATA> end
MTB> let c6=c4/c5
MTB> name c6 'cumavgcolor'
```

This gives the average amount won per each play of the game for n games. This goes from n=1 to 1000.

```
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.

(l) Now open a new worksheet in this project (File|New|Worksheet). Note: We are doing this because the student version of Minitab can only handle 5000 entries in a worksheet. Repeat (i)-(k) for the "number" bet by first putting its values into c1 and the corresponding probabilities into c2 of Worksheet 2. In particular, comment on how the net winnings vary for the "number" bet versus the "color" bet and why this makes sense.

(m) Lastly, open another worksheet (Worksheet 3) and copy columns 5 and 6 from Worksheet 1 into c1 and c2, respectively, of Worksheet 3 and copy columns 5 and 6 from Worksheet 2 into c3 and c4, respectively, of Worksheet 3. Finally, compare the two bets in terms of their cumulative mean winnings by plotting both on the same scale (make sure your curser is in Worksheet 3 when you execute these commands):

```
MTB> plot c2*c1 c4*c3;
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.]

<u>Caution:</u> It's important to be able to distinguish between a *distribution of data* and a *probability distribution*, including sample mean $\bar{x}$ vs. expected value E(X) and sample variance s^2 vs. V(X).

(n) Explain the relationship of the "mean" computed in part (i) to the "mean" computed in part (b) using terms in the Caution above.

<u>Disclaimer</u>: If you have any thoughts of gambling, check out:
http://math.ucsd.edu/~anistat/gamblers_ruin.html