STAT 325 – Handout 9
Expected Value, Variance (2.3, 2.4)

Example 9-1: Random babies (cont.)
Recall again the "random babies" process, with $n = 4$. I just used R to simulate this process a million times, with the following results for number of correct matches:

Number of matches	0	1	2	4
Number of repetitions producing that many matches	374,844	333,184	250,296	41,676

a) Determine the average number of matches in these 1,000,000 simulated repetitions of this random process. [*Hint*: As you would calculate any average, add up the 1,000,000 values and divide by 1,000,000.]

b) What value would you guess for the limit of this average, as the number of repetitions approaches infinity?

Let the random variable X = number of correct matches. The probability mass function of X is:

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

We can use this (exact) probability distribution to find the (exact) value of the long-term limit of the average number of matches.

- The **expected value** of a discrete random variable X with pmf $p(x)$, denoted by E(X) and also by μ, is defined as: E(X) = $\sum_x x \times p(x)$.
  - Interpreted as the long-run average value of the random variable.
  - Sum of product of possibilities times probabilities
  - Not necessarily "expected" in usual meaning of word

c) Use this definition to calculate the expected value of X = number of correct matches with n = 4 items.

d) What is the probability that X = E(X) in this case? Is E(X) more likely to occur than not? Is it even the most likely outcome?

Example 9-2: Rolling dice (cont.)
Roll a fair die once. Let the random variable X = number of dots on the side that lands up.

a) Determine E(X).

b) Determine Pr(X = E(X)).

c) Is X likely to equal its expected value? Is it even possible?

d) Interpret what E(X) represents in this context.

Now suppose that I offer to play a game with you, where your winnings will depend on the outcome of the roll of a die. Let W represent your winnings, with a negative value indicating a loss, where W is defined by $W = [3 - (4-X)^2]$.

e) Determine the pmf of W.

f) Based on this pmf, calculate E(W).

g) Interpret what E(W) says about how this game would turn out in the long run.

There's another way to find the expected value of a function of a random variable:

- $E[h(X)] = \sum_x h(x) \times p(x)$
    - Sometimes called "Law of the Unconscious Statistician"

h) Verify that this expression produces the same value for E(W).

Some more rules of expected values:

- $E(aX + b) = a\,E(X) + b$ for any constants a and b, any random variable X
- $E(X + Y) = E(X) + E(Y)$ for any random variables X and Y
    - $E(a_1X_1 + a_2X_2 + \ldots + a_kX_k) = a_1E(X_1) + a_2E(X_2) + \ldots + a_kE(X_k)$

i) Suppose that I offer you a choice between rolling one fair die and doubling the points, or rolling two fair dice and adding the points. Use the appropriate rules to determine the expected value of each random variable.

Example 9-3: Composite blood testing

Suppose that 10 people need to be given a blood test for a certain disease. Assume that each person has a .1 probability of having the disease, independently from person to person. Consider two different plans for conducting the tests:

- Plan A: Give the blood test individually to each person.
- Plan B: Pool blood from all 10 people into one composite sample, test that.
    - If at least one person has the disease, then test result will be positive, and then all 10 people will have to be tested individually.
    - If nobody has the disease, then the test result will be negative, and no future tests will be needed.

Let the random variable X represent the total number of tests needed with plan A. Similarly, let Y represent the total number of tests needed with plan B.

a) Determine the pmf of X and E(X).

b) Determine the pmf of Y and E(Y).

c) Interpret what E(Y) means in this context.

d) Which plan has the smaller expected number of tests?

e) If you implement plan B once, what is the probability that the number of tests will be smaller than it would be with plan A?

f) Now let *p* represent the probability that an individual has the disease. For what values of *p* is the expected number of tests lower with composite testing (in a group of 10 people)?

Example 9-4: St. Petersburg Paradox
Suppose that I toss a fair coin repeatedly, and let the random variable X be the number of tosses required to obtain Heads for the first time.

a) Derive an expression for the pmf of X.

Now suppose that I offer to play a game: You win $1 if the number of tosses required is no more than 5, and you lose $25 if the number of tosses exceeds 5.

b) Determine the expected value of your profit from playing this game.

c) Interpret what this expected value means.

Now let's make the game more interesting: You win 2^X dollars if the number of tosses required (X) is no more than 5, and you lose $5 if the number of tosses exceeds 5.

d) Determine and interpret the expected value of your profit from playing this game.

Now let's consider one final variation of this game: You pay a certain amount to play, and then you receive 2^X dollars, where X is the number of tosses required.

e) Use expected value to determine how much you should be willing to pay for the privilege of playing this game. Explain what your answer means.

Example 9-5: Rolling dice (cont.)
Again suppose that I offer to pay you one of the following amounts:

V = two times the number rolled on one fair die
W = sum of the numbers rolled on two fair dice

The pmf's of these two random variables are:

v	2	4	6	8	10	12
p(*v*)	1/6	1/6	1/6	1/6	1/6	1/6

W	2	3	4	5	6	7	8	9	10	11	12
p(*w*)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

You found earlier that E(V) = 7 and E(W) = 7.

a) Interpret what these expected values mean.

b) Are these three random variables identical? Explain.

c) With which of these random variables would you have more variability in your winnings?

- The **variance** of a random variable X, denoted by Var(X) or σ^2, is defined as $E[(X-\mu)^2]$, where $\mu = E(X)$.
    - A measure of how spread out the probability distribution is.
    - The **standard deviation** of X, denoted by SD(X) or σ, is defined as the square root of the variance.

d) Use the definition to calculate the variance of the random variables V and W.

d) Do these variance calculations agree with what you anticipated in c)?

e) Calculate the 2 standard deviations.

- Short-cut formula: Var(X) = $E(X^2) - [E(X)]^2$

f) Calculate $E(V^2)$, and then use this short-cut formula to calculate Var(Y). Does the answer agree with above?

Now suppose that I charge you $5 to play this game, so your winnings would be V – 5.

g) Without doing any calculations, how do you expect the Var(V – 5) to compare to Var(V)? Explain.

h) Write our the pmf of V – 5. Then use this and the short-cut formula to determine Var(V – 5).

i) How does Var(V – 5) to compare to Var(V)? Explain why this makes sense.

As you might expect, we have some rules about variances that could have been helpful with the above calculations. Let X be a random variable and let *a* and *b* be constants. Then:

- Var(X + b) = Var(X)
- Var(aX) = a^2Var(X)
- Var(aX + b) = a^2Var(X)
- If X and Y are independent random variables, then Var(X + Y) = Var(X) + Var(Y).
  - If X and Y are not independent, then this result usually does not hold.

j) Let X be the result of a single die roll. Use the shortcut formula to calculate Var(X).

k) Use the rules above to calculate Var(2X) and Var(X_1 + X_2). Confirm that these agree with Var(V) and Var(W) as found above.

Example 9-6: Roulette

In the game of (American) roulette a wheel is spun, and a ball eventually comes to rest in one of its 38 numbered slots. The slots have colors: 18 red, 18 black, and 2 green. A gambler may either bet on a number or a color. If he bets $1 on a number and the ball lands in that number's slot, then he gets $36 back for a net profit of $35. If he bets $1 on a color and the ball lands in a slot of that color, then he gets $2 back for a net profit of $1. In both situations, he loses his original $1 bet (for a net profit of -$1) if the ball does not land in a slot of the right number/color.

Let the random variable X represent the net winnings/profit from betting $1 on a *color*, and let the random variable Y represent the net winnings/profit from betting $1 on a *number*.

a) Determine the pmf of X and the pmf of Y.

b) Determine E(X) and E(Y).

c) What do you notice about these expected values, and what does that tell you about these two kinds of bets?

d) Determine $E(X^2)$ and then Var(X) and SD(X).

e) Determine $E(Y^2)$ and then Var(Y) and SD(Y).

f) Which type of bet has the larger variance? Explain why this makes sense.