1. RANDOM VARIABLES

Definition -- usually denoted as X or Y or even Z and it is the numerical outcome of a random process,

Example random process: The number of heads in 10 tosses of a coin

Example: The number "5" ratings given to a company by its customers on a scale from 1 to 10.

Example: The time it takes for commuters to and from the San Fernando valley to get home on Fridays

Example: We manufacture beer. We draw one bottle at random from the conveyor belt. The contents could be measured (e.g. volume 0 to 300ml, taste scale of 1 to 5) and these numerical variables describe the characteristics of the randomly selected bottle. These variables could be thought of as random variables, and they might vary with each bottle chosen.

Can you think of other examples? Random variables can be DISCRETE or CONTINUOUS

2. Discrete Random Variables (p.303)

A discrete random variable is countable and finite. Recall some variables are naturally discrete such as you can't roll a 3.1 or you can't have 351.7 employees at a firm or you can't have 3.91 customer complaints today. Discrete variables jump from one whole value to the next. So one can "list" the outcomes of the known possible values a random variable X can take. And one can list the probabilities (or percentage of occurrences, or chance) of each possible outcome (value).

Discrete Random variables have probability distributions -- they are just a way of organizing outcomes and representing them graphically. A table or a graph might suffice. There are only 2 requirements:

- probabilities must be greater than or equal to zero
- the sum of the probabilities must be 1.

The customer ratings, the number of heads in 10 tosses, the number of times a "6" is rolled in 25 rolls of a fair die are all discrete examples.

3. Continuous Random Variables

A continuous random variable can assume an infinite number of values in a given interval of values. So for example, our beer bottles can contain any amount of beer in an interval between 0 and 300 ml.

The most commonly observed continuous random variable in this class is the NORMAL distribution (Chapter 6). A probability model can be described by a curve and the probability of any given event can be described by the area under the curve. We are always interested in the probability for an interval (or range) rather than the probability of an exact value when working with a continuous random variable. This is simply because the area under the curve at any exact point will be zero.

Notation: Greek Letter mu or μ symbolizes the mean of the normal distribution, Greek letter sigma, or σ, the standard deviation. In the Standard Normal table that we use, the mean is

μ =0 and the standard deviation σ=1.

Example: Suppose an automobile manufacturer claims their new SUV has mean in-city mileage of 16 miles per gallon. Suppose you write to the manufacturer and you find out that the standard deviation around that mean is 2 miles per gallon. This information allows you to formulate a probability model. So you think that the random variable "in city gas mileage" can be approximated by a normal distribution with a mean of 16 and a standard deviation of 2.

> *What is does the distribution of the in-city gas mileage look like for the population of these vehicles? What percentage do we expect to be between 14 and 18 miles per gallon? What percentage do we expect to be between 12 and 20 miles per gallon?*

Example: Suppose you work for a magazine that tests new autos and trucks. If you were to test this SUV, what is the probability that the one you purchased averages less than 13 miles to the gallon? What is the probability that you would purchase one that gets more than 20 miles? Suppose you were to purchase one that gets better than 20 miles per gallon, is your probability model necessarily wrong?

Example: Suppose you are thinking about investing some money in a mutual fund. Past data shows that the fund returned a mean of 19.8% with a standard deviation of 13.40%. Suppose we know it is normally distributed. Based on this information, what is the probability that you will experience a loss (get a return of less than zero). This year-to-date, the fund has returned -9.07 (a loss). What is the probability of getting a return that low or lower? Is the model necessarily wrong?

4. Random Variables have means too…

The mean of a list of numbers like 1, 2, 3, 4, 5 or like 1,1,2,2,3,3,4,4,5,5, gives every value in the list equal weight. The mean of a random variable is also an average, but slightly different, it assigns probabilities to the outcomes (values the variable can take) and these outcomes do not need to be equal.

The examples in part 3 above give a mean (and standard deviation) for a random variable.

Generally, the mean of random variables are written μ_x pronounced "mu sub x" to represent the mean of any random variable x, not just normal ones. What would the symbol μ_y mean to you? Another way it might be written is E(X) or E(Y)

5. The mean of a discrete random variable: The Expected Value (p. 304-305)

Example: Suppose it is known that a salesman typically makes 3 phone calls a year to each home in his region and that his chance of making 3 sales (one for each call) is 5/100 (or 5% or .05), 2 sales is 15/100 (or 15% or .15) , 1 sale is 30/100 (or 30% or .30) and no sale is 50/100 (or 50% or .50). If he earns $200 per sale, how much can he EXPECT to earn in a year (or what would the MEAN of this random variable be)?

We know that in 3 phone calls, he faces 4 possible outcomes, but what is "most likely to happen?" or “what is EXPECTED to happen?”

To find the mean of this random variable X, multiply each possible outcome by its probability and add up the products:

The formula (see page 305) $\boldsymbol{\mu_x = x_1p_1 + x_2p_2 + x_3p_3 + \ldots + x_np_n}$

They write it this way: $\mu = E(X) = \sum x \bullet P(X = x)$

So for the example above, μ_x is (0*.50) + (1*.30) + (2*.15) + (3*.05) = .75 sales

Note several things

1. The outcomes are "listable", there are 4 outcomes (0, 1, 2, 3)
2. The probabilities are values between 0 and 1, their sum is 1.0
   a. .50+.30+.15+.05 = 1.0
   b. .50 is 50% or 50/100, .30 is 30% or 30/100 etc.
3. The mean is .75 sales, this suggests that if he makes 3 phone calls, he can expect to make less than one sale. In other words, if he make 300 phone calls, he would expect to make 75 sales (that's .75 * 100). If the company is paying him $200 per sale, he can expect to make $15,000 in a year.

6. The mean of a continuous random variable

You usually need some calculus to calculate the mean for a continuous random variable unless it comes from a very simply symmetric distribution, such as a uniform distribution (it looks like a brick). In this class, it is generally given to you as the mean of a normal distribution. Remember, the normal distribution is a continuous probability distribution and normal random variable is a continuous random variable. So in the examples on the SUVs and the mutual funds in point #3 above, you would be expected to make statements about the distribution based on information given about the mean and standard deviation of the variable.