4/7/2022

More Binomial Examples, mean, standard deviation
Continuous Distributions (Chapter 5)
Review for Exam #1

Three properties:

1) Two outcomes: success vs. failure (number of successes x or r)
2) Fixed number of trials (n)
3) The probability of success is fixed between trials (p, probability of failure is 1-p=q)

Expected value = mean
For the binomial distribution, the mean (expected number of successes) $\mu = np$
For the coin flip example, p=0.5, and n=10 flips, so the mean is 10*0.5=5

For standard deviation: $\sigma = \sqrt{npq} = \sqrt{np(1-p)}$
For the coin flip: $\sqrt{10 * 0.5 * 0.5} \approx 1.58$ …

Example.
In a recent study of seat-belt usage, it was determined that 90% of passengers use a seat belt in their car when they drive. A sample of 12 people was selected and put in a car simulator.

a. What is the probability that exactly 10 people put on their seatbelt?
   Number of trials: n=12, probability of success: 90%=0.9, number of successes: r=10
   $$B(12,0.9,10) = \binom{12}{10}(0.9)^{10}(0.1)^2 = 23\%$$

b. What is the probability that exactly 12 people will put on their seatbelt?
   R=12

$B(12,0.9,12) = \binom{12}{12}(0.9)^{12}(0.1)^0$ =0.28242954

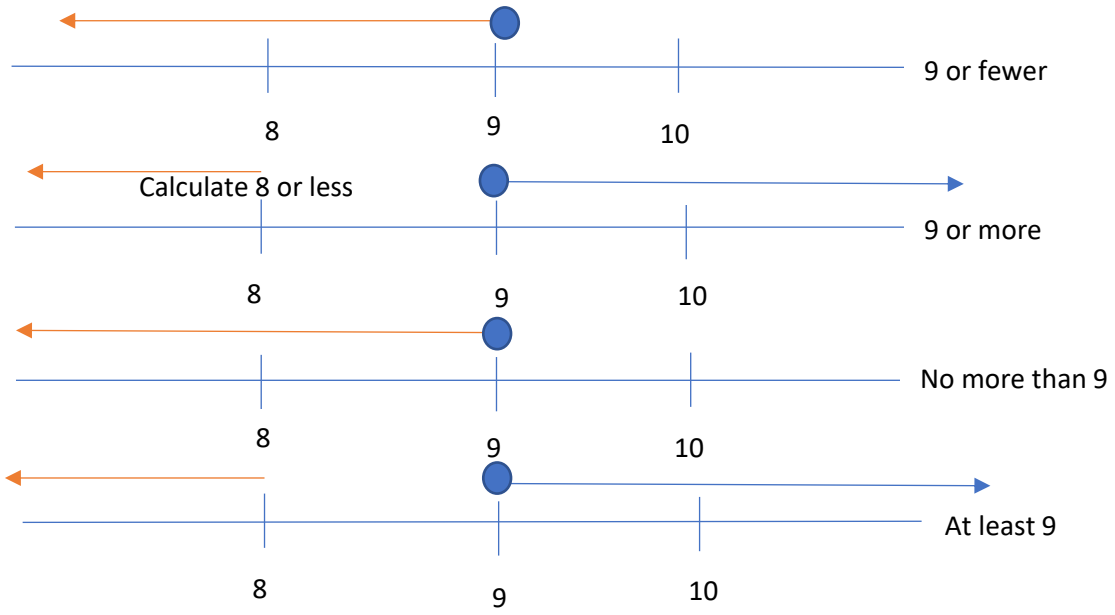
If you use 0.09, you get 2.8243E-13=$2.82 \times 10^{-13} = 0.000000000000282$

This is a very small number! Nearly 0. If you report the probability to be 2.82 you are wrong! 2.82 is not a probability.

c. What is the probability that 9 people or fewer will put on their seatbelt?
   If a problem uses phrases like “or fewer”, “or more”, “at least”, or “no more than” you will need a cdf (cumulative to be TRUE in Excel).

   The cdf/cumulative function always adds up the number included (up to an including the number specified). It’s always a less than or equal to. If we need MORE THAN, use 1-cumprob. The number where the cut-off stops, is not necessarily the number of successes in the problem. See graphs below.

   Since we want 9 or fewer, use the cumulative function stopping at 9.
   About 11%



d. What is the probability that at least 9 people put on their seatbelt. (9, 10, 11, 12 people did)
It is valid to calculate the probability that 9 people did, that 10 people did, that 11 people, that 12 people did, separately, and then add them up.
The faster way is to calculate the probability of the complement; that 8 people or less did NOT put on their seatbelt, and then subtract from 1. This is around 97%

Alternatively: 12 people put on their seatbelt = 0 failures, if 11 people put on their seatbelt = 1 failure, 10 people put on their seatbelt = 2 failures, 9 people put on their seatbelt = 3 failures.
Flip "success" and "failure".
What is the probability that 3 or fewer people will not put on their seatbelt?
P=0.1, cumulative up to 3
Still 97%

e. What is the probability that between 8 and 10 people (inclusive) put on their seatbelt?
Option 1: calculate exactly 8, exactly 9, exactly 10 and then add up.
Option 2: cumulative up to 10 (maximum), and then subtract off the cumulative less than 8 (minimum= less than or equal to 7) about 33.7%
Option 3: use the between function in Excel BINOM.DIST.RANGE

Continuous Distributions (chapter 5)
Much more likely to see an equation rather than a table of values.
In the continuous, the probability is thought as the "area" under the curve.
We can't talk about the probability at a given point, but we can talk about the probability of between 2 values. Always specify the top and bottom of a range. Always cumulative when calculating probabilities.

The simplest continuous distribution is the uniform distribution.
Every outcome is equally likely. The total probability is 1. That means area under the curve (of the pdf function) is 1.

$$A = lw = \left(\frac{1}{k}\right)k$$

Think of the width of the rectangle as k (some range of values of a certain width), then the height is (1/k) to make the area equal 1.

Suppose I have a uniform random variable that has outcomes between 10 and 20. The range is 20-10 = 10., so the height of the function (graph) = 1/10.

For the probability of an outcome between 12 and 15: multiply the height of the function (1/10) times the width of the range in the probability question.
Range: 15-12=3: Area = probability = 3*1/10 = 3/10= 0.3

Rand() = uniform between 0 and 1. Since the height is 1, the range equals the probability in this case.

The other example in textbook is the exponential distribution. Related to the Poisson distribution.
Poisson distribution counts the number of events in a given time period.
Exponential distribution measures the time between events.
The exponential distribution has a nice formula for the cumulative distribution.

Review for Exam #1

- Classify variables
- Vocabulary, explanations of concepts
- Sampling, experiments
- Frequency tables
- Constructing and interpreting graphs
- Calculating descriptive statistics
- Probability questions
- Discrete distributions, properties, expected value