Math 114 **ACTIVITY 6:** The Binomial distribution and the Empirical Rule

Why

The binomial family the most basic family of distributions for considering sampling and repeated experiments. The "pure" binomial applies directly to repeated experiments and to experiments on groups of subjects; sampling without replacement involves the "near binomial"—calculations give "close enough" results when we sample form a large enough population. It is necessary to understand the ideas behind the calculations to see how inference base on categorical variables are handled in statistics. One particular feature for today's work is the bell-shaped character of the binomial distribution for large values of n

LEARNING OBJECTIVES

1. Be able to use the formula and the binomial probability table (for nice cases) to find binomial probabilities.
2. Be able to find and understand the generic parameters μ and σ for a binomial variable
3. Recognize when the Empirical rule applies to a binomial distribution and be able to draw appropriate conclusions.

CITERIA

1. Success in working as a team and in fulfilling the team roles.
2. Success in involving all members of the team in the conversation.
3. Success in completing the exercises.

RESOURCES

1. Your text, section 6.2 and the binomial table (Table II) page A-2
2. The team role desk markers (handed out in class for use during the semester)
3. Your class notes from Monday and Wednesday
4. Your calculator
5. 40 minutes

PLAN

1. Select roles, if you have not already done so, and decide how you will carry out steps 2 and 3 (5 minutes)
2. Work through the exercises given here - be sure everyone understands all results & procedures(30 minutes)
3. Assess the team's work and roles performances and prepare the Reflector's and Recorder's reports including team grade (5 minutes).

DISCUSSION

1. A variable **X** is a *binomial random variable* with n trials and probability p of success if it counts the number of "successes" ("Yes" results) in a sequence of n Bernoulli trials with probability p of success ("Yes") each time.
2. The possible values for a binomial variable with n trials and probability p of success are $0, 1, \ldots, n$ and for each number k in this range, $P(X = k) = nCk * p^k * (1 - p)^{(n-k)}$. [$nCk$ counts the number of orders in which k successes could occur in n trials; in each order, the k successes have probability p^k and the $(n - k)$ failures have probability $(1 - p)^{(n-k)}$]
3. There are tables which give these values for "nice" n and p; Table III (p. A-3 in the appendix) gives binomial probabilities for values of n up to 12 and for $n = 15$ and $n = 20$, simplifying our calculations. For each n there is a block in the table— the column heading gives p, and there is a row for each possible x. For example, if we have $n = 10$ trials and $p = .15$, we use the block on page A-6 (labeled with $n = 10$) and the column headed .15. Looking in the row for $x = 2$ tells us the probability $P(X = 2)$ is .2759 (rounded to 4 places). Combining appropriate values (from the rows) lets us find that if $n = 10, p = .15$, then $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = .1969 + .0374 + .2759 = .5102$.
4. There are very nice (special) formulas for the mean and standard deviation of a binomial variable:
mean $\mu_X = np$
standard deviation $\sigma_X = \sqrt{n * p * (1 - p)}$.

5. The distribution of X will be symmetric if $p = .5$, skewed right if $p < .5$, skewed left if $p > .5$ (with stronger skew for p further from .5 and for smaller values of n). However, if n is large enough and p not too far from .5 (Rule of thumb: $np(1-p) \geq 10$), then the distribution of values is approximately bell-shaped, and we can apply our *Empirical Rule* which says that approximately 68% of the values are within one standard deviation of the mean (between $\mu - \sigma$ and $\mu + \sigma$ - about half below μ and half above) and about 95% of the values are within two standard deviations of the mean (between $\mu - 2\sigma$ and $\mu + 2\sigma$); in particular, any values less than $\mu - 2\sigma$ or greater than $\mu + 2\sigma$ are *unusual* observations - which would occur less than 5% of the time by chance.

MODELS

1. Suppose we select 15 people at random from a city in which 70% of the people are opposed to opening a casino. The variable X = number of people in the sample who oppose the casino is a binomial random variable with $n = 15, p = .7$
   (a) The probability we will get exactly 10 people who oppose the casino is $P(X = 10) = 15C10{*}(.7)^{10}{*}(.3)^5 = .2061$ — this probability is also easily obtained from Table III in the text (which begins on p. A-3).
   (b) The probability we will get exactly 4 people (in our sample) who oppose the casino is $P(X = 4) = 15C4 * (.7)^4 * (.3)^1 1 = .0006$
   (c) The probability we will get 10 or more people (in our sample) who oppose the casino is $P(X \geq 10) = P(X = 10) + P(X = 11) + P(X = 12) + P(X = 13) + P(X = 14) + P(X = 15) = .2061 + .2186 + .1700 + .0916 + .0305 + .0047 = .7215$ (values obtained more easily form the table than from the formula).
   (d) The probability we will get fewer than 14 (in our sample) who oppose the casino is $P(X < 14)$—which is more easily obtained using the complement rule for probabilities: $P(X < 14) = 1 - P(X \geq 14) = 1 - (P(X = 14) + P(X = 15)) = 1 - (.0305 + .0047) = .9648$.

2. Suppose we select a random sample of 150 people from a population in which 10% of the people have allergies. We will count the number of people in the sample who have allergies; our variable X = number of people in the sample with allergies is binomial, with $n = 150, p = .10$
   (a) The expected number of people in the sample with allergies is $\mu = 150 * .10 = 15$ and the standard deviation (remember X is a random variable - different samples will give different numbers of people with allergies) is $\sigma = \sqrt{150 * .10 * .9} = 3.67$
   (b) Since $n * p * (1-p) = 150 * .1 * .9 = 13.5$ and $13.5 > 10$, the distribution of X will be approximately bell-shaped. We know that about 68% of samples will give between 11.33 (=15 - 3.67) and 18.67 (= 15 + 3.67) people with allergies; about 95% will give between 7.66 (= 15 - 2*3.67) and 22.34 (= 15 + 2*3.67) people with allergies.
   (c) A sample with 30 allergy sufferers would be unusual – 30 is more than $\mu + 2\sigma$ (which is 22.34, in this case), so the probability of such a sample is less than 5%.

EXERCISE

1. For a binomial random variable with $n = 12, p = .75$
   (a) Use the formula to find $P(X = 8)$
   (b) Use the table to find $P(X > 10)$
   (c) Give μ_X and σ_X

2. (Use the table, if possible, for calculations) In March, 80% of the people on on a certain large Florida beach are from other states. Suppose we take a sample of 20 people from this beach and count the number of people from other states.
   (a) What is the probability we will find (on our sample) exactly 10 people from other states?
   (b) What is the probability we will find fewer than 18 people from other states?
   (c) What is the probability we will find 15 to 17 people from other states (pretty closely matching the real proportion 80%)?
   (d) If we drew many such samples, what would we expect as the average number of out-of state people per sample (What is the long-term mean)?
   (e) What is the standard deviation of the number of out-of-state people per sample ["average" deviation of actual number from the mean]?

3. Clarinex-D is a medication intended to reduce symptoms associated with allergies. In clinical trials, 5% of patients experienced insomnia as a side effect. If 240 users of Clarinex-D are randomly selected:

(a) Let X = number of people in the sample who experience insomnia as a side effect. Give the values of n and p for this binomial variable.

(b) What is the expected number of people in the sample who experience insomnia as a side effect?

(c) The value of p is fairly small (not very close to .5) — show that our sample is large enough for the distribution of the variable X to be (approximately) bell-shaped.

(d) Based on the empirical rule, what range of values should contain 68% of the values of X?

(e) Would a sample with 30 people suffering the insomnia side effect be considered unusual?

READING ASSIGNMENT (in preparation for next class)
Section 7.1 on continuous random variables, particularly the uniform and normal families

SKILL EXERCISES:(hand in - individually - at next class meeting Use the binomial table (Table III), when n and p are in the table, to reduce calculations) Sullivan p.315 #39–42, 48-49