Announcements
• Midterm will be returned on Wednesday
• One Handout Today

Lecture 10 – Sampling Distributions OR the Distribution of All Possible Samples OR the population of Samples for COUNTS and PROPORTIONS (Chapter 5.1)

Binomial Distributions
• The binomial distribution is a model for some kinds of categorical variables. They typically represent the number “successes” in “n” trials.
• To use the binomial, the total number of observations “n” is fixed
• Each observation takes one of two categories “success” or “failure”
• The outcomes of all “n” observations are independent
• The “n” observations have the same probability (chance) of “success” we call it “p”
Examples of Binomial Distributions

- We survey 100 UCLA students. Each student is either of legal drinking age, (greater than or equal to 21) or under the legal drinking age (under 21). Each student has either had an alcoholic drink or not.
- We examine 50 wells in a city, each well is either contaminated or it is not.
- We interview 200 women, each woman has experienced discrimination on the job or she has not

The Binomial, formally

- A binomial distribution for the count of $X$ “successes” in “$n$” observations can be written as a function of two parameters, $n$ and $p$: $B(n, p)$
- Where: $n$ is the total number of observations (and is fixed in advance), $p$ is the probability of “success” on each observation
- The count of successes $X$ can be any whole number (discrete) between 0 and $n$

Formal Examples

- A coin is flipped 10 times, each outcome is either a head or a tail. Let $X$ be the number of heads in 10 flips (our count of successes). On each flip, the probability of success (head) is 0.5. The number $X$ of heads in 10 flips is binomial $B(10, 0.5)$
- We call 100 adults at random, each adult is either a supporter of the war in Iraq or not. Let $X$ be the number of supporters in 100 calls. Suppose the probability of support is 0.40. The number of supports in 100 calls is binomial $B(100, 0.40)$
Why Binomial

• Binomials are useful when we want to know about the presence of a particular outcome and not its magnitude.
  – How many supporters, not the strength of support
  – How many drinkers, not the number of drinks they have had
  – How much discrimination, not the exact nature

Sampling Distribution: REVIEW

• The sampling distribution is a theoretical/conceptual/ideal probability distribution of a statistic.
• A theoretical probability distribution is what the outcomes (i.e., statistics) of some random process (e.g., drawing a sample from a population) would look like if you could repeat the random process over and over again and had information (that is statistics) from every possible sample.
• Note that a sampling distribution is the theoretical probability distribution of a statistic. The sampling distribution shows how a statistic varies from sample to sample and the pattern of possible values a statistic takes.
• We do not actually see sampling distributions in real life, they are simulated. They exist in theory.

Sampling Distribution of a count

• A population contains some proportion $p$ of “successes”. If we choose to sample from this population, the count $X$ of “successes” in a random sample of size $n$ will be approximately binomial $B(n,p)$.
• If the population is at least 20 times larger than the sample (a rule of thumb) the $n$ observations will be independent and the BINOMIAL SAMPLING DISTRIBUTION FOR COUNTS exists.
The mean and standard deviation of a binomial sampling distribution for a COUNT

- The mean and standard deviation of the binomial sampling distribution for a count X are:
  \[ \mu = np \quad \text{and} \quad \sigma = \sqrt{np(1-p)} \]
- Example: If the probability of a “head” in a coin toss is p=.50, then in ten tosses, \( np = (10 \times .50) = 5 \) and \( \sigma = \sqrt{10 \times .5(1-.5)} = 1.581 \)

Sampling Distribution of a proportion

- Proportions can be more informative than counts. Example: it’s easier to understand “.85 (or 85%) of likely voters are interested in Tuesday’s results than to say 1700 out of 2000 likely voters are interested”
- In statistics, the sample proportion of successes \( \hat{p} \) (p-hat) is used to estimate the true population proportion, p.
- For any SRS of size n, the sample proportion of successes is
  \[ \hat{p} = \frac{\text{count of successes in a sample}}{n} = \frac{X}{n} \]

Examples

- In a SRS of 200 UCLA students, 60 are age 21 or older. \( \hat{p} = \frac{60}{200} = .30 \)
- In a SRS 500 of bone fragments from a mass grave, forensic anthropologists were able to identify 275 of them as belonging to humans under the age of 10. \( \hat{p} = \frac{275}{500} = .55 \)
The mean and standard deviation of a binomial sampling distribution for a PROPORTION

- The mean and standard deviation of the binomial sampling distribution for a proportion are:
  \[ \mu_p = p \quad \text{and} \quad \sigma_p = \sqrt{\frac{p(1-p)}{n}} \]
- Note: the variability of \( \hat{p} \) around the true mean \( p \) decreases as \( n \) (the sample size) increases. This tells us that larger sample give closer estimates of the population proportion, \( p \).

The Normal Approximation

- When \( n \) is large and \( p \) is not too close to either 0 or 1, the binomial distribution can be approximated by the normal distribution where and  
  \[ \mu = np \quad \text{for counts} \quad \sigma = \sqrt{np(1-p)} \]
- And for proportions
  \[ \mu_p = p \quad \sigma_p = \sqrt{\frac{p(1-p)}{n}} \]
- Generally, you can use the normal approximation for a binomial when \( np \geq 10 \) AND \( n(1-p) \geq 10 \)

Application of the normal approximation: the sampling distribution of \( \hat{p} \)

- The sampling distribution of \( \hat{p} \) is never perfectly normal, but as your sample size \( n \) increases, it becomes approximately normal.
- The normal approximation works best for larger sample sizes and when \( p \) is near .5. It works poorly for small samples and when \( p \) is near 0 or 1.
Example

- A recent survey of 827 likely special election voters, 46% said they support Proposition 74 on probationary periods for public school teachers. If the governor is correct in his thinking that Proposition 74 will win with 51% of the vote, what is the probability that a survey of this size could get a result of 46% or fewer?

Answer

- Treat the 51% as the mean of the population “μ”, the governor thinks he knows the truth.
- The 46% is p-hat, from the sample of 827
- To use the normal, we need to convert this information to a Z score:

\[ Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.46 - .51}{\sqrt{\frac{(.51)(.49)}{827}}} \approx -2.88 \]

Then read table A, the area to the left of −2.88 is .002

Graphics

- This is what a binomial population looks like:
• This is what a single sample of 827 might look like:

• Theory tells us that the population of all possible samples of n=827 would look normal and the \( \hat{p} \)-hats from each sample would center on \( p=0.51 \). Here is a graph of 10,000 different random samples of size 827 with \( p=0.51 \).