Statistical Inference

October 19, 2022

Today

- Statistical Inference
  - Interested in making inferences about some population parameter:
    - Mean, proportion, etc.
    - Difference in mean, proportion, etc.
  - Generally, we only observe sample statistics
  - What does a sample statistic tell us about the population?
- Sampling Distributions
  - Normal Distribution
  - Binomial Distribution
- Hypothesis Testing
- Statistical Significance and Confidence

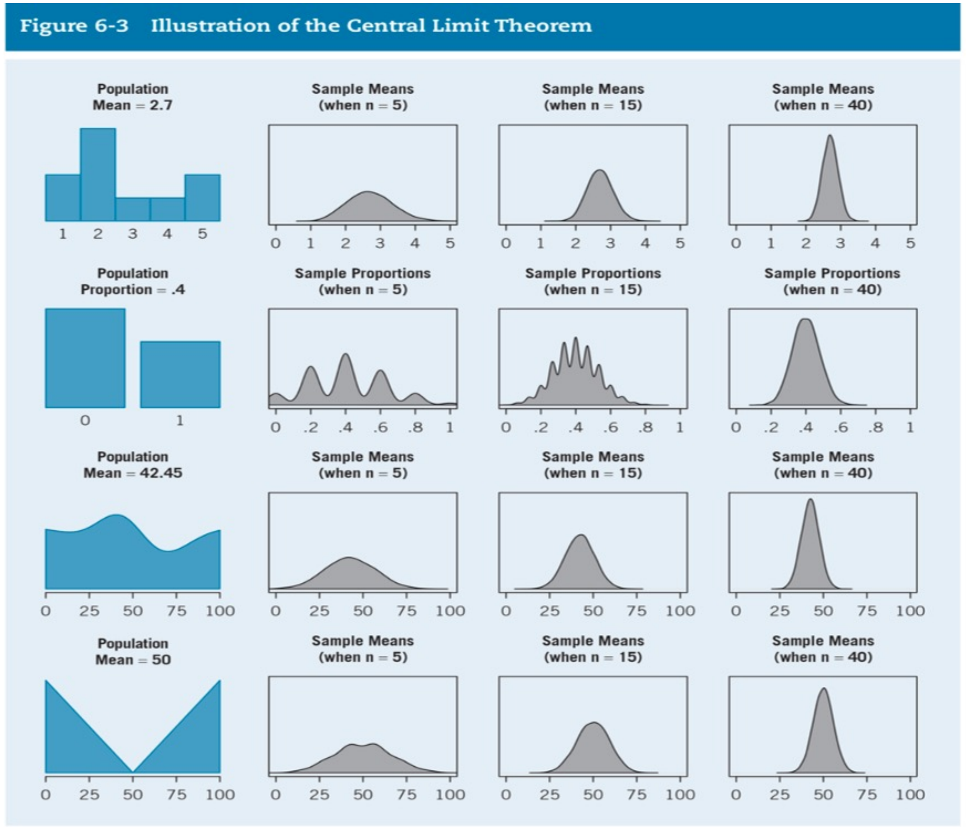
Population and Sample

- New York Times/Sienna Poll (conducted October 9-12, 2022)—A random sample of 792 registered voters (conducted via telephone)—39% approve of the job President Biden is doing as President (margin of error is +/- 4 points)
  - What does this tell us about the population of adults and their approval of Biden as president?
- Recall, a "parameter" is a fixed number in the population
  - Percentage of all adults in the U.S. that approve of the job Biden is doing as president—*we DON'T know this*
- A "statistic" is a number describing some quantity using a sample
  - 39% of people in the sample approve of the job Biden is doing as President—*we DO know this*
- *A key point—we know what the distribution of samples would look like if we took a lot of random samples*

Sampling Distribution

- Sampling Variability: the value of a statistic varies in repeated random sampling (due to random sampling error)
  - But, if we took infinite random samples, we would obtain a distribution of sample statistics
- Sampling distribution—distribution of values taken by the statistic in all possible samples of the same size from the same population
- Bias: a statistic is *unbiased* if the mean of its sampling distribution equals the true value of the population parameter
  - Assured through random sampling
- Variability: described by the spread of the sampling distribution.
  - If sample size is larger, less variability (all else equal)

Sample Size & Sampling Variability



Population and Sampling Distribution

- Population Distribution—the distribution of a variable's values for all members of the population
  - Ex. Income for all adults in the United States
- Population distribution=the probability distribution for individual chosen at random from population
  - Ex. Likelihood that random individual will have specific income (or income in a specific range)

Sample Means

- Sample mean ($\bar{x}$):
  - Sample means are less variable than individual observations
  - Sample means are more "Normal" than individual observations
- If the population has mean μ, then μ is the mean of the probability distribution for each observation
- So, the mean of $\bar{x}$ (in repeated sampling) is the same as the mean of the population
  - The mean of the sample means ($\bar{x}$) is the same as the mean of the population
    - So, $\bar{x}$ is an unbiased estimator of μ if a simple random sample
- We can estimate the variance of the sampling distribution as well:
  - $\sigma^2{}_{\bar{x}} = \frac{\sigma^2}{n}$ (note: σ^2 is the variance of the population distribution)
  - Standard deviation: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Sampling distribution of sample mean

- So, the sample mean is an unbiased estimator of the population mean
  - i.e., the center of the probability distribution of $\bar{x}$ is μ
- And, the accuracy of estimation is affected the sample size
  - Standard deviation of the probability distribution of $\bar{x}$ (which we will call the "standard error") is $\frac{\sigma}{\sqrt{n}}$
- What about the shape of the distribution?
  - If a random sample is large, the probability distribution is approximately normal
    - This is the Central Limit Theorem

Distribution for Counts and Proportions

- In sum, the probability distribution for the sample mean of a quantitative variable is normal
  - Mean of the sample means is the same as the population mean
  - Spread affected by variance in the population & sample size
- However, for categorical variables—a different distribution
  - For a random variable with only two possible categories (like a coin flip): binomial distribution
  - Sample proportion: $\hat{p} = \frac{x}{n}$ (note: the "hat" signifies a sample statistic)
    - Example: Sample proportion of adults approving of Biden=301/792=0.38

Binomial Distribution

- The Binomial Setting:
  - There is a fixed number of observations (*n*)
  - The *n* observations are all independent
  - Each observation falls into one of just two categories, can label these "success" and "failure"
  - The probability of a success (*p*) is the same for each observation
- Binomial Distribution:
  - The distribution of the count x of successes with parameters n and p.
    - n = number of observations
    - p = probability of a success on any one observation
    - x is $B(n,p)$

Binomial Probabilities

- If the population is much larger than the sample (at least 20 times):
  - The count *x* of successes in an SRS of size *n* has approximately the binomial distribution *B(n,p*)
- The mean (μ_x) of the *B(n,p*) distribution is *n*p*
  - The number of observations * the probability of obtaining a success
- The standard deviation of the *B(n,p)* distribution is:
  - σ_x=square root of np(1-p)
- Sample proportion, remember, is
  - $\hat{p} = \frac{x}{n}$
  - Mean of sample proportion: μ_p=p
  - Standard deviation (i.e., standard error) of sample proportion: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Normal Approximation

- When the number of observations is large, the sampling distribution of both $\bar{x}$ and $\hat{p}$ are approximately normal
  - This is because of the Central Limit Theorem
- The normal distribution is useful because we can use it to determine the likelihood that some specific proportion in our sample is different from some other proportion

Statistical Inference

- If we know the distribution for the population, can simulate drawing many random samples and demonstrate all of this in those samples
- But, we don't normally know anything about the population
  - Size
  - Mean
  - Standard Deviation
- But, we can use the same logic developed here to make estimates of the population parameter and to discuss how confident we are about those estimates

Statistical Inference

- We want to estimate a population parameter from a sample statistic
- **Point estimate**: a single number that is our "best guess" for the population parameter
- **Interval Estimate**: An interval of numbers within which the parameter value is believed to fall (with some specified degree of confidence)
- A point estimate does not tell us how close the estimate is likely to be to the parameter
- An interval estimate is more useful
  - Margin of error helps to gauge the accuracy of the point estimate
- Example:
  - President Biden's approval rating is 38%, give or take 4%

Properties of Point Estimates

- Property 1: a good estimator has a sampling distribution that is centered at the (population) parameter
  - An estimator with this property is *unbiased*
  - Sample mean is an unbiased estimator of the population mean
  - Sample proportion is an unbiased estimator of the population proportion
- Property 2: a good estimator has a small standard error compared to other estimators
  - Smallest variance (i.e., spread) of the sampling distribution
  - This means it tends to fall closer than other estimates to the parameter

Confidence Intervals for Proportions

- Inference about a parameter should provide not only a point estimate but also indicate its likely precision
- A **confidence interval** is an interval containing the most believable values for a parameter
- The probability that this method produces an interval that contains the parameter is called the **confidence level**.
- Remember, the sampling distribution gives the possible values for the sample proportion and their probabilities
  - Is approximately a normal distribution for large random samples
  - Has a mean equal to the population proportion
  - Has a standard deviation called the **standard error**

Confidence Intervals for Proportions

- Approximately 95% of a normal distribution falls within 1.96 standard deviations on each side of the mean
- With probability 0.95, the sample proportion falls within about 1.96 standard errors of the population proportion.
- Margin of error measures how accurate the point estimate is likely to be in estimating a parameter
- The distance of 1.96 standard errors is the margin of error for a 95% confidence interval
    - A confidence interval is constructed by adding and subtracting a margin of error from a given point estimate

Confidence Interval for Proportions

- We symbolize a population proportion by "p"
- The point estimate of the population proportion is the sample proportion
  - We symbolize the sample proportion by $\hat{p}$
  - se($\hat{p}$) = $\sqrt{\frac{p(1-p)}{n}}$
- Remember, p is typically unknown
- A 95% confidence interval for the population proportion is
  - $\hat{p}$ +/-1.96*se($\hat{p}$)

Confidence Intervals for Proportions

- A sample proportion value occurs with probability of 0.95 such that the confidence interval contains the population proportion
- The method produces a confidence interval that misses p with probability of 0.05
- In the long run (i.e., repeated sampling), about 95% of intervals would give correct results that contain the true population proportion

Example

- I flip a coin 1000 times and observe the number of heads. Assume that the coin is fair. What is a 95% confidence interval for the number of heads I will get?
- I get 522 heads. Can I be 95% confident that the coin is fair?

Confidence Intervals for Sample Mean

- Remember, the standard error for the mean of $\bar{x}$ is
  - $\frac{\sigma}{\sqrt{n}}$
- The confidence interval for the population mean is
  - $\bar{x} +/- z(\frac{\sigma}{\sqrt{n}})$
- Since we do not know σ, we estimate it using s. So:
  - $\overline{x\grave{}} +/- z(\frac{s}{\sqrt{n}})$

Example

- Assume that height is normally distributed. I take an SRS of 900 individuals from a population with a mean height of 66 inches and a standard deviation of 6 inches. What is the range within which 95% of the sample means will fall?

Choosing a sample size

- We can select a sample size to give us a given margin of error at a given level of confidence
- Confidence interval for a population mean will have a specific margin of error m when the sample size is:
  - $n = \frac{z^2(s^2)}{m^2}$
- Example: Estimate of average income for population, standard deviation probably \$10000, at 95% confidence interval, with a margin of error of \$500

Choosing a sample size

- Confidence interval for a proportion will have a specific margin of error when the sample size is:
  - $n = \frac{\hat{p}(1-\hat{p})z^2}{m^2}$
- Example: Estimate proportion of Maryland voters intending to vote for Wes Moore, at 95% confidence interval, with a margin of error of 0.02, assume a proportion around 0.6

Hypotheses

- A hypothesis is a statement about a population, usually of the form that a certain parameter takes a particular numerical value or falls in a certain range of values
- The main goal in many research studies is to check whether the data support certain hypotheses
- Confidence intervals and significance tests are used to evaluate hypotheses

Null and Alternative Hypothesis

- The value of the null hypothesis typically represents "no effect"
  - H_0
- The value of the alternative hypothesis usually represents an effect of some type:
  - H_a
- Example:
  - H_a: In comparing individuals, those who are married watch lower levels of television on average than those who are not married.
  - H_0: In comparing individuals, there is no difference in the average level of television viewed by those who are married compared to those who are not married

One/Two-sided hypotheses

- Hypotheses can be one or two-sided
- A one-sided hypothesis predicts a particular direction
  - Example: one group has a higher mean than another group
- Two-sided does not predict a direction
  - Example: one group has a different mean than another group
- How confidently we can reject the null hypothesis is affected (in part) by whether the hypothesis is one or two-sided

Test Statistic

- The hypotheses refer to a population parameter
  - Ex. Difference in means between two groups in population
- The parameter to which the hypothesis refers has a point estimate: the sample statistic
  - Ex. Difference in means between two groups in sample
- A test statistic describes how far that estimate (the sample statistic) falls from the parameter value given in the null hypothesis

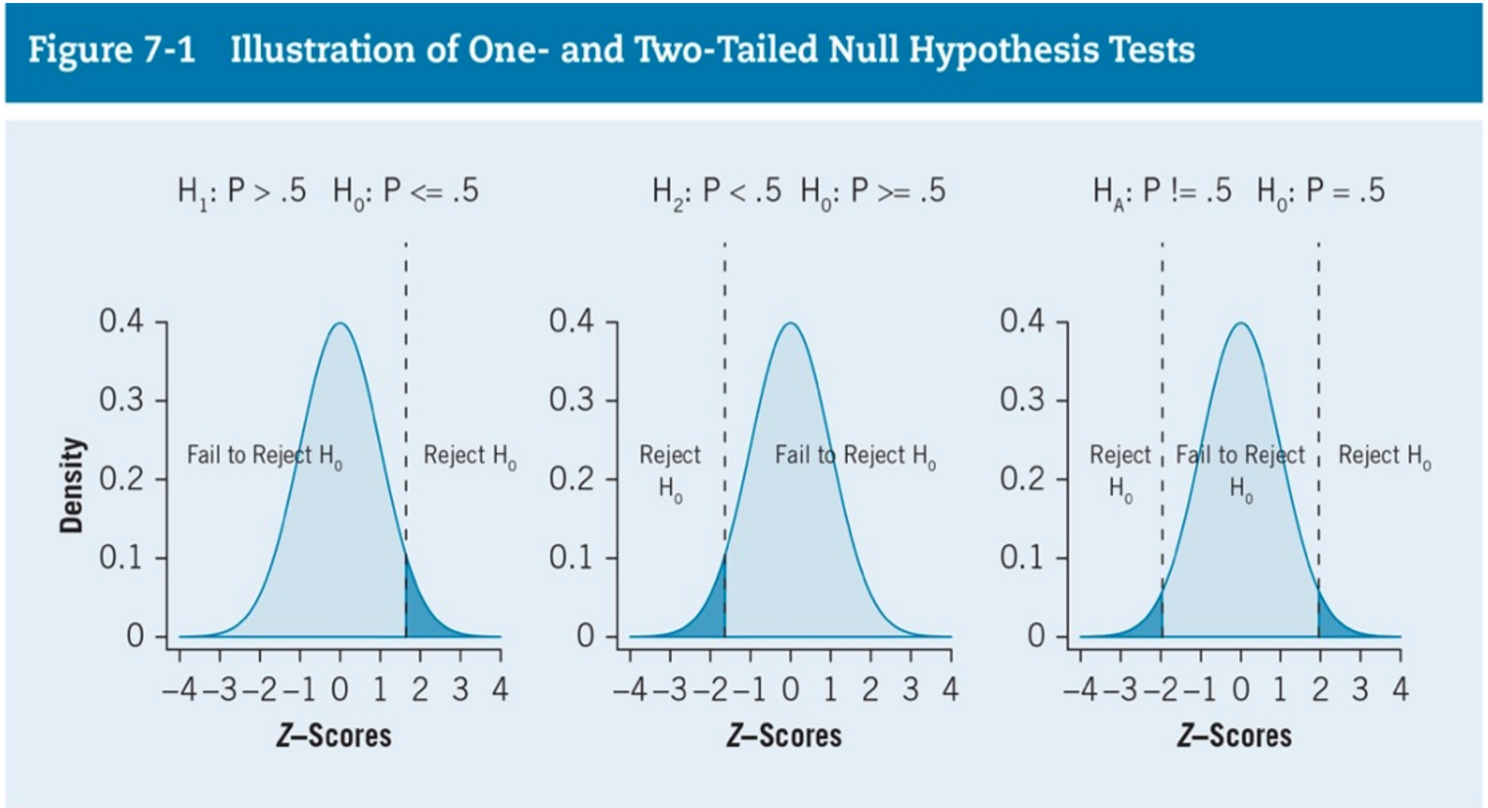
P-values

- To interpret a test statistic value, we use a probability summary of the evidence against the null hypothesis, H_0
- First, we presume that H_0 is true
- Next, we consider the sampling distribution from which the test statistic comes
- We summarize how far out in the tail of this sampling distribution the test statistic falls
  - We do this by the tail probability of that value and values even more extreme
- This probability is called a **p-value**
  - *The probability that the test statistic equals the observed value or a value even more extreme*
  - The smaller the p-value, the stronger the evidence is against H_0

Interpret p-values—Two-sided vs. One-sided Hypotheses

- A significance test analyses the strength of the evidence against the null hypothesis
- Remember—a two-sided alternative hypothesis does not predict a specific direction
- To test a two-sided null hypothesis, the p-value is the two-tail probability under the standard normal curve
- We calculate this by finding the tail probability in a single tail and then doubling it—i.e., must account for both tails of the distribution
- A one-sided significance test, by contrast, only looks to one tail of the distribution

One vs. Two-Tailed Hypothesis Testing



Significance Levels

- The significance level (α) is a number such that we reject H_0 if the p-value is less than or equal to that number
- In practice, the most common significance level is 0.05
- When we reject H_0, we say that the results are statistically significant (at some level—e.g., the 0.05-level)
- Example:
  - The p-value for a two-sided test of the null hypothesis H_0: $\mu=30$ is 0.06
  - Does the 95% confidence interval include the value 30? Why?
  - Does the 90% confidence interval include the value 30? Why?

Significance Levels

- A level α two-sided significance test rejects H_0: $\mu=\mu_0$ exactly when the value μ_0 falls outside a level 1-α confidence interval around μ.
- Learning the actual P-value is more informative than learning only whether the test is "statistically significant at the 0.05 level"
- The p-values of 0.01 and 0.049 are both statistically significant in this sense, but the first p-value provides (arguably) stronger evidence against H_0 than the second

What affects significance?

- Four things affect ability to find statistical significance (all else equal):
  - Sample size—as the sample size increases, a smaller statistic in the sample is more likely to be significant
    - But, diminishing returns (especially once the sample size exceeds 1200 or so)
  - Variability—as the variability of a sample statistic increases, harder to find statistical significance
  - Distance between the sample statistic and what the population parameter would be if the null hypothesis were true
  - Level of statistical significance chosen—95% confidence (i.e., a p-value less than 0.05) is the norm
    - Which also includes decision for one vs. two-tailed test

Type I and Type II Errors

- Two types of Error:
  - Relationship in our sample, but not in population (Type I Error)
  - Relationship in population, but not in our sample (Type II Error)
- So, we can be incorrect about a relationship in the population in two ways
- In general, we are more concerned about Type I errors than Type II errors
  - Would rather mistakenly miss a relationship than mistakenly see one

Type I and Type II Errors

		Relationship in population	
		Yes	No
Relationship in our sample	Yes	Correct	**Type I error**
	No	**Type II error**	Correct

Remember Randomization

- Remember, all of this presumes randomization
- If we do not have a random sample, we introduce bias
- Bias means that center of sampling distribution not population parameter
- In that case, estimates are worse

Next Steps

- Next week, we will look specifically at hypothesis testing in three cases:
- Difference of means testing:
  - Interval DV
- Difference of proportion test:
  - Dichotomous DV
- Two-way Table:
  - Categorical IV/Categorical DV