Lectures 20 - 21
Introduction to Significance Tests

Review: Estimation

Parameter	Point Estimate	Standard Error	Standard Score	Confidence Interval	Sample size
Population Proportion p	Sample Proportion $\hat{p}$	$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$z \sim N(0,1)$	$\hat{p} \pm z \times SE(\hat{p})$	$n = \frac{p\,(1-p) \times z^2}{m^2}$
Population Mean μ	Sample Mean $\bar{x}$	$SE(\bar{x}) = s/\sqrt{n}$	$z \sim N(0,1)$	$\bar{x} \pm z \times SE(\bar{x})$	$n = \frac{\sigma^2 \times z^2}{m^2}$
Population Mean μ	Sample Mean $\bar{x}$	$SE(\bar{x}) = s/\sqrt{n}$	$t \sim T(n-1)$	$\bar{x} \pm t \times SE(\bar{x})$	

Assumptions of estimators:

- The CI for a population proportion is only valid under large sample sizes
- The CI using the z-score for the population mean is valid only for very large sample sizes (greater than 100)
- The CI using the t-score for the population mean is valid for all sample sizes, but requires the population distribution of the variable x to be normal
- Simple random sampling with replacement

Warm Up

Paying Higher Prices To Protect the Environment – As part of a larger study about environmental sentiment in the U.S, the General Social Survey (GSS) surveyed 1,361 Americans to ask whether they would be willing to pay fuel higher prices ($7 or more dollars a gallon) to protect the environment. Of the adult Americans who responded, 637 reported that they were willing to do so.

- Find and 95% confidence interval for the proportion of adult Americans willing to pay more at the pump for the environment

- Suppose the researchers conduct a new study at the same confidence level. What sample size would be needed to estimate the proportion of Americans willing to pay more at the pump with a margin of error of 0.02%

Comparing populations

- In many situations, we want to compare two population parameters

- Do the statistical results in our data support certain statements of conclusions?

Ex.) An experiment to see if a certain medication can reduce risk of heart attack in elderly patients

  - Patients are randomized to either a placebo pill or the medication

  - We are interested inferring whether the proportion of elderly patients who suffered a heart attack is “**statistically**” higher in the population taking the placebo

Another example

Ex.) We may be interested in comparing the mean amount of campaign donations of between voters who register as Democrat vs those who register as Republicans

- We would sample individuals from both parties and compute their respective sample averages for charitable donations
- If there is strong evidence that amount of charitable donations between the two political affiliations differs, we say that the two groups **differ significantly with respect to the population parameters**

Statistical Significance

- For given statistical study, **statistical significance** means the result is one that is decidedly not due to “ordinary variation” in the data (i.e., not due to chance or not a coincidence).

- **Statistical tests** (aka significance tests (also called hypothesis tests) are how we decide whether an observed result is statistically significant

Example:

- The outcome of a coin flip has the following population distribution.

x	$P(x)$
Heads	p
Tails	$1-p$

- A coin is flipped n times. The value of the population parameter p implies something about the coin:
  1. If $p = 0.5$, the coin is fair
  2. If $p \neq 0.5$ the coin is not fair

Example Continued

- Assume we do not know the value of p. We flip the coin 30 times to produce a sample of $n = 30$ observations. It comes up heads 20 times, so $\hat{p} = 20/30 \approx 0.67$. What might we decide about p?

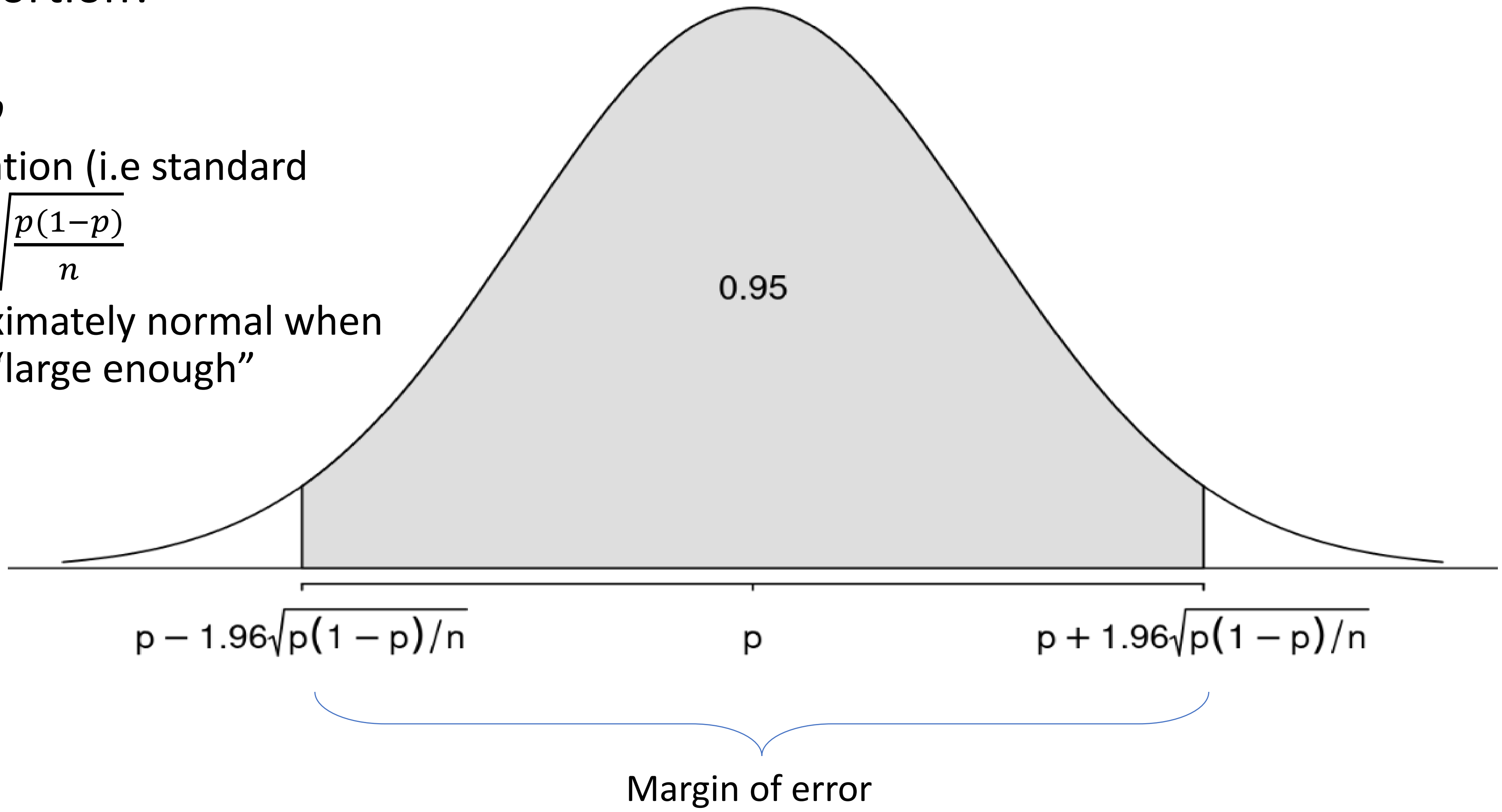
1. Conclude that $p = 0.5$: The result that $\hat{p} = 2/3$ is not statistically significant.
2. Conclude that $p \neq 0.5$: The result that $\hat{p} = 2/3$ is statistically significant.

- How do we decide? (more on this in a minute)

The sampling distribution of $\hat{p}$

- What do we know about the sampling distribution of a proportion?

1. The mean of $\hat{p} = p$
2. The standard deviation (i.e standard error) is $\mathrm{SE}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$
3. It's shape is approximately normal when the sample size is "large enough"



Converting the sampling distribution of $\hat{p}$ to the standard normal

- We convert the sampling distribution of $\hat{p}$ to a standard normal distribution via

$$z = \frac{observation \ - mean}{SE} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

- But we don't know the value of p! ☹

- We set the value of p according to what we think it might or should be

Steps of a hypothesis test

- A hypothesis test has five steps:
  1. Assumptions: A hypothesis test makes certain assumptions or has specific conditions under which it applies.
     - Firstly, all hypothesis tests assumes the data is produced under randomization.
     - Other assumptions may be about sample size or the shape of the distribution

  2. Hypotheses: Each significance test has two hypotheses about a population parameter a $null\ hypothesis$ and an $alternative\ hypothesis$

     The **null hypothesis $\boldsymbol{H_0}$** is usually the hypothesis of “no effect” or that “nothing interesting is happening”. The null hypothesis is usually that the population parameter equals some value
     $H_0: \theta = a$ for some constant a

     The **alternative hypothesis $\boldsymbol{H_A}$** is the hypothesis of “effect” or that “something interesting has happened”. The alternative hypothesis is usually that the population parameter falls in some range of values

$$H_A: \theta > a$$
$$H_A: \theta < a$$
$$H_A: \theta \neq a$$

Example:

- A coin is flipped n times. We can consider the observation of each flip to be a random variable with the following distribution.

x	$P(x)$
Heads	p
Tails	$1-p$

1. If $p = 0.5$ the coin is fair
2. If $p \neq 0.5$ the coin is not fair

What assumptions do we have for a hypothesis test for the value of p?

What is the null hypothesis?

What is a possible alternative hypothesis?

Steps of a hypothesis test continued

Step 3. The **test statistic** measures the discrepancy (distance) between the point estimate of the parameter and the hypothesized value of the parameter.

- The discrepancy is typically measured in number of standard errors between the point estimate and the value of the parameter

- <u>A test statistic is computed under the assumption that the null hypothesis is true.</u>

The statistic z is a test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Example

- $z_{obs} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ is a test statistic

- What is the value of the test statistic for our example about the fair coin?

$$n = 30$$
$$\hat{p} = 20/30$$
$$H_0: p = 0.5$$

Steps of a hypothesis test continued

Step 4. The **P-value**: We look for "evidence against the null" by computing a quantity called the p-value. We do this using probability in the following way

- We assume that null hypothesis H_0 is true, since the burden of proof is on the alternative hypothesis
- We consider the sorts of values we might get for the test statistic when H_0 is true by considering its sampling distribution under H_0.
- If the test statistic we compute falls well out of the margin of error when H_0 is true, we take this as evidence against the null

The **P-value** is the probability of observing a value of a value of the test statistic that is as extreme or more extreme than the observed value given that the null hypothesis is true

$$P(Z > Z_{obs} | H_0 \text{ true})$$

Example

- $z_{obs} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$ is a test statistic

- What is the value of the test statistic for our example about the fair coin?

$$n = 30$$
$$\hat{p} = 20/30$$
$$H_0: p = 0.5$$
$$H_A: p > 0.5$$