# PSY 503: Foundations of Psychological Methods
# Lecture 14: Hypothesis Testing II

Robin Gomila

Princeton

October 21, 2020

# One-sample vs. two-sample tests

# One-sample vs. two-sample tests

- **One-sample tests**
  - $H_0$ : population mean equals a specific value $\mu_X$
  - e.g., the average self-reported level of happiness in New Jersey on a scale from 0-100 is 70; the average amount of sleep in the U.S. is 7 hours per night

# One-sample vs. two-sample tests

- **One-sample tests**
  - $H_0$ : population mean equals a specific value $\mu_X$
  - e.g., the average self-reported level of happiness in New Jersey on a scale from 0-100 is 70; the average amount of sleep in the U.S. is 7 hours per night
- **Two-sample tests**
  - $H_0$ : the means of two populations equal each other
  - e.g., test whether the observed difference between treatment and control groupps is likely to arise by random chance alone

# One-sample tests

# Proportion of U.S. voters who support Biden

- Suppose $H_0 : \mu_X = 0.5$

# Proportion of U.S. voters who support Biden

- Suppose $H_0 : \mu_X = 0.5$
- Let the *alternative hypothesis* $H_1 : \mu_X \neq 0.5$

# Proportion of U.S. voters who support Biden

- Suppose that we asked a random sample of 1018 U.S. voters if they support Joe Biden

# Proportion of U.S. voters who support Biden

- Suppose that we asked a random sample of 1018 U.S. voters if they support Joe Biden
- Suppose that in this sample, 550 respondents expressed support for Joe Biden
  - All other respondents do not support Biden

# Proportion of U.S. voters who support Biden

- Suppose that we asked a random sample of 1018 U.S. voters if they support Joe Biden
- Suppose that in this sample, 550 respondents expressed support for Joe Biden
    - All other respondents do not support Biden
- Proportion of Biden's supporters in the sample is $\bar{X}_n = \frac{550}{1018} = 0.54$

# Proportion of U.S. voters who support Biden

- Suppose that we asked a random sample of 1018 U.S. voters if they support Joe Biden
- Suppose that in this sample, 550 respondents expressed support for Joe Biden
    - All other respondents do not support Biden
- Proportion of Biden's supporters in the sample is $\bar{X}_n = \frac{550}{1018} = 0.54$
- Question is: is this difference between sample proportion and hypothesized proportion statistically significant?

# Proceed with hypothesis testing

- Test statistic:

# Proceed with hypothesis testing

- Test statistic: sample proportion $\bar{X}_n$

# Proceed with hypothesis testing

- Test statistic: sample proportion $\bar{X}_n$
- $\alpha = 0.05$

# Proceed with hypothesis testing

- Test statistic: sample proportion $\bar{X}_n$
- $\alpha = 0.05$
- What does the central limit theorem tell us about the distribution of $\bar{X}_n$?

# Proceed with hypothesis testing

- Test statistic: sample proportion $\bar{X}_n$
- $\alpha = 0.05$
- What does the central limit theorem tell us about the distribution of $\bar{X}_n$?
    - $X_n \sim N(0.5, \frac{0.5(1-0.5)}{1018})$

# Proceed with hypothesis testing

- Test statistic: sample proportion $\bar{X_n}$
- $\alpha = 0.05$
- What does the central limit theorem tell us about the distribution of $\bar{X_n}$?
  - $X_n \sim N(0.5, \frac{0.5(1-0.5)}{1018})$
  - That's just because for Bernoulli rvs: $\mathbb{V}[X] = \theta(1-\theta)$

# Proceed with hypothesis testing

- Test statistic: sample proportion $\bar{X}_n$
- $\alpha = 0.05$
- What does the central limit theorem tell us about the distribution of $\bar{X}_n$?
  - $X_n \sim N(0.5, \frac{0.5(1-0.5)}{1018})$
  - That's just because for Bernoulli rvs: $\mathbb{V}[X] = \theta(1-\theta)$
  - If the variable was not Bernoulli, we would have estimated the variance of $\bar{X}_n$ using the sample variance

# Proceed with hypothesis testing

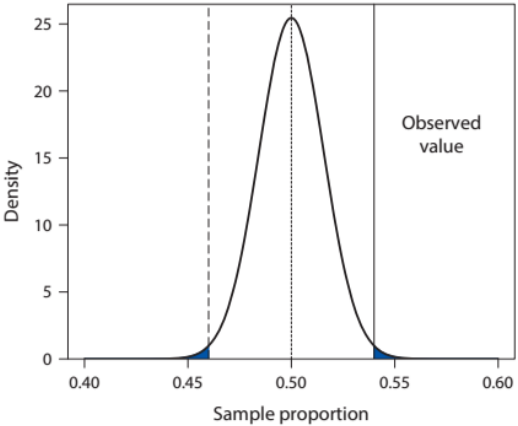
- What is the two-sided p-value? (in words)

# Proceed with hypothesis testing

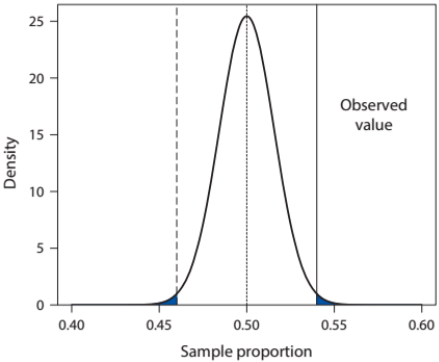
- What is the two-sided p-value? (in words)
  - The probability that, under the null hypothesis, we observe a value that is more extreme that the observed value $\bar{X}_n = 550/1018$

# Proceed with hypothesis testing

- What is the two-sided p-value? (in words)
  - The probability that, under the null hypothesis, we observe a value that is more extreme that the observed value $\bar{X}_n = 550/1018$

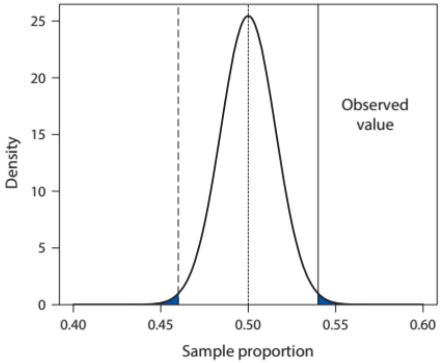


# Calculate p-value



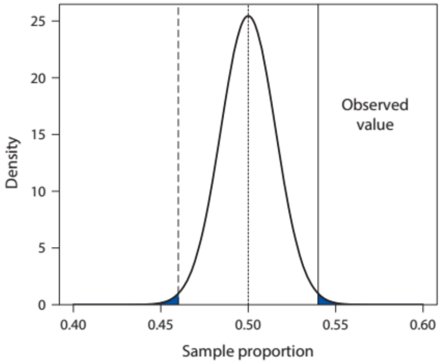

- The two-sided p-value corresponds to the sum of the two blue shaded areas

# Calculate p-value




- The two-sided p-value corresponds to the sum of the two blue shaded areas
- Let's calculate it in R!

# Calculate p-value




- The two-sided p-value corresponds to the sum of the two blue shaded areas
- Let's calculate it in R!
- $p = .01017$

# Different alternative hypothesis

- What if our alternative hypothesis was $\mu_X > 0.5$ instead of $\mu_X \neq 0.5$?

# Different alternative hypothesis

- What if our alternative hypothesis was $\mu_X > 0.5$ instead of $\mu_X \neq 0.5$?
- No need to consider the possibility of extremely small values
- $p = \frac{\text{two sided p-value}}{2} = .005085$

# Test Statistics

# z-statistic (z-test)

# z-statistic (z-test)

- We commonly use the z-score as a test statistic when the central limit theorem kicks in
  - i.e., if large enough sample, z-score can be used with any distribution of the variable of interest
- What's a z-score?

# z-statistic (z-test)

- We commonly use the z-score as a test statistic when the central limit theorem kicks in
  - i.e., if large enough sample, z-score can be used with any distribution of the variable of interest
- What's a z-score?

$$Z = \frac{\bar{X}_n - \mu_0}{\text{standard error of } \bar{X}_n} \sim N(0,1)$$

## z-statistic (z-test)

- We commonly use the z-score as a test statistic when the central limit theorem kicks in
  - i.e., if large enough sample, z-score can be used with any distribution of the variable of interest
- What's a z-score?

$$Z = \frac{\bar{X}_n - \mu_0}{\text{standard error of } \bar{X}_n} \sim N(0,1)$$

- Using the z-score instead of the sample mean for the Biden study yields the exact same result
  - Let's try this out in R

# t-statistic (t-test)

# t-statistic (t-test)

- If the variable of interest $X$ is normally distributed (e.g., IQ) in the population of interest, the same test statistic $Z_n$ follows the t-distribution with $n - 1$ degrees of freedom

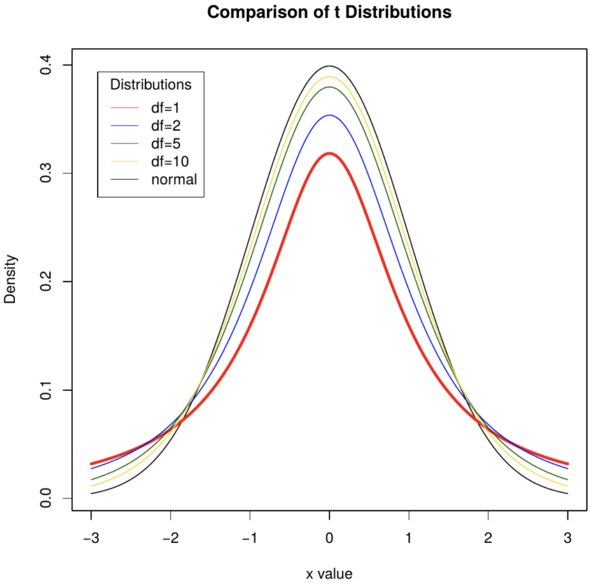
# t-statistic (t-test)

- If the variable of interest $X$ is normally distributed (e.g., IQ) in the population of interest, the same test statistic $Z_n$ follows the t-distribution with $n-1$ degrees of freedom
  - In small samples, the t-distribution looks like the normal distribution but has fatter tails

# t-statistic (t-test)

- If the variable of interest $X$ is normally distributed (e.g., IQ) in the population of interest, the same test statistic $Z_n$ follows the t-distribution with $n - 1$ degrees of freedom
  - In small samples, the t-distribution looks like the normal distribution but has fatter tails
  - In large sample, t-distribution becomes the normal distribution, therefore t-statistic and z-statistic yield the same p-values

# t-statistic (t-test)



# t-statistic (t-test)

- Statistical software always use the t-statistic

# t-statistic (t-test)

- Statistical software always use the t-statistic
- In large samples: no matter the distribution of $X$, z-test and t-test produce same answer

# t-statistic (t-test)

- Statistical software always use the t-statistic
- In large samples: no matter the distribution of $X$, z-test and t-test produce same answer
- In small samples: t-test can be used (but not z-test) **if and only if** $X$ is normally distributed

# p-values vs. confidence intervals

# P-values vs. CIs for hypothesis testing

- P-values and CIs yield the exact same conclusion:
  - Reject the null if CI (confidence level $1-\alpha$) does not include $\mu_0$
  - Reject the null if $p < \alpha$

# Advantages of reporting CIs (instead of or in addition to p-values)

# Advantages of reporting CIs (instead of or in addition to p-values)

- CIs carry information about the magnitude of the effect
- p-values don't
- CIs superior in this sense!

# one-sample t-tests in R

Back to R studio

# Two-sample tests

# Two-sample tests are very similar to one-sample test

- All the properties of the sampling distribution of the sample mean also applies to the sampling distribution of the difference in means between two groups (see code from previous lecture)

# Two-sample tests are very similar to one-sample test

- All the properties of the sampling distribution of the sample mean also applies to the sampling distribution of the difference in means between two groups (see code from previous lecture)
  - CLT kicks in with large enough samples

# Two-sample tests are very similar to one-sample test

- All the properties of the sampling distribution of the sample mean also applies to the sampling distribution of the difference in means between two groups (see code from previous lecture)
  - CLT kicks in with large enough samples
  - z-statistic of the difference in mean is normally distributed with mean 0 and variance 1

# Two-sample tests are very similar to one-sample test

- All the properties of the sampling distribution of the sample mean also applies to the sampling distribution of the difference in means between two groups (see code from previous lecture)
  - CLT kicks in with large enough samples
  - z-statistic of the difference in mean is normally distributed with mean 0 and variance 1
  - if the random variable of interest is normally distributed in the two groups, the z-statistic becomes a t-statistic and follows the t-distribution in small samples

# Two-sample tests are very similar to one-sample test

- All the properties of the sampling distribution of the sample mean also applies to the sampling distribution of the difference in means between two groups (see code from previous lecture)
    - CLT kicks in with large enough samples
    - z-statistic of the difference in mean is normally distributed with mean 0 and variance 1
    - if the random variable of interest is normally distributed in the two groups, the z-statistic becomes a t-statistic and follows the t-distribution in small samples
- Open R Studio

# Statistical Power and Power Analysis

# Statistical power

- The **power** of a statistical test is defined as the probability of rejecting the null hypothesis, **conditional on** the null hypothesis being false

# Statistical power

- The **power** of a statistical test is defined as the probability of rejecting the null hypothesis, **conditional on** the null hypothesis being false

$$\text{power } = 1 - P(\text{type II error})$$

# Statistical power

- The **power** of a statistical test is defined as the probability of rejecting the null hypothesis, **conditional on** the null hypothesis being false

$$\text{power } = 1 - P(\text{type II error})$$

- This implies that the more statistical power we have, the greater the probability that the study will detect a departure from the null hypothesis, if the null hypothesis is false

# Power analysis

- Statistical power is a function of the sample size and the variance of $X$

# Power analysis

- Statistical power is a function of the sample size and the variance of $X$
- Power analysis is used to determine the smallest sample size necessary to estimate the parameter with enough precision that its observed value is distinguishable from the parameter value assumed under the null hypothesis

# Power analysis: Applications

- e.g., determine the sample size of an experiment to detect an effect of a certain size to reject the null that there is no effect
- Power analyses in pre-registrations: a priori decision about sample size
- Power analyses often require for research grant applications to justify budget

# Power analysis:

- Let's run a power analysis in R