

450A Political Methodology I

Lecture 2: Univariate Statistical Inference

3. Hypothesis Testing

Stanford University

What is Hypothesis Testing?

Hypothesis Testing with a Single Random Variable

Hypothesis Testing: The Trial Analogy

Often, our primary goal is to test a **hypothesis** about the value of a parameter.

The statistical **decision theory** underlies such hypothesis testing.

The trial analogy:

Suppose we must decide whether to convict or acquit a defendant based on evidence presented at a trial. There are four possible outcomes:

		<i>Defendant</i>	
		Guilty	Innocent
<i>Decision</i>	Convict	Correct	Type I Error
	Acquit	Type II Error	Correct

We could make two types of errors:

- Convict an innocent defendant (**type-I error**)
- Acquit a guilty defendant (**type-II error**)

Our goal is to limit the probability of making these types of errors.

However, for a fixed test and sample size, making rejection harder lowers the type-I error probability but raises the type-II error probability. We therefore need to balance them.

Two Types of Errors in Hypothesis Testing

		<i>Defendant</i>	
		Guilty	Innocent
<i>Decision</i>	Convict	Correct	Type-I error
	Acquit	Type-II error	Correct

Now, suppose that we have a statistical model for the probability of convicting and acquitting, conditional on whether the defendant is actually guilty or innocent.

Then, our decision-making rule can be characterized by two probabilities:

- $\alpha = \Pr(\text{type-I error}) = \Pr(\text{convict} \mid \text{innocent})$
- $\beta = \Pr(\text{type-II error}) = \Pr(\text{acquit} \mid \text{guilty})$

The probability of making a correct decision is therefore $1 - \alpha$ (if innocent) and $1 - \beta$ (if guilty).

Hypothesis testing follows an analogous logic, where we want to decide whether to **reject** (= convict) or **fail to reject** (= acquit) a **null hypothesis** (= defendant is innocent) using sample data.

Steps of Hypothesis Testing

		Null Hypothesis (H_0)	
		False	True
Decision	Reject	$1 - \beta$	α
	Fail to Reject	β	$1 - \alpha$

- 1 Specify a **null hypothesis** H_0 (e.g. the defendant = innocent)
- 2 Pick a value of $\alpha = \Pr(\text{reject } H_0 \mid H_0)$ (e.g. 0.05). This is the maximum probability of making a type-I error we decide to tolerate, and called the **significance level** of the test.
- 3 Choose a **test statistic** T , which is a function of sample data and related to H_0 (e.g. the count of testimonies against the defendant)
- 4 Assuming H_0 is true, derive the **null distribution** of T (e.g. standard normal)

Steps of Hypothesis Testing

		Null Hypothesis (H_0)	
		False	True
Decision	Reject	$1 - \beta$	α
	Fail to Reject	β	$1 - \alpha$

- 5 Using the **critical values** from a statistical table, evaluate how unusual the observed value of T is under the null hypothesis:
- If the probability of drawing a T at least as extreme as the observed T is less than α , we reject H_0 .
(e.g. there are so many testimonies against the defendant that they would be very unlikely if she were innocent, so reject the hypothesis that she was innocent.)
 - Otherwise, we fail to reject H_0 .
(e.g. there is not enough evidence against the defendant, so give her the benefit of the doubt.)

What is Hypothesis Testing?

Hypothesis Testing with a Single Random Variable

- One-Sample t-Test

- p -values and Confidence Intervals

- Practical versus Statistical Significance

What is Hypothesis Testing?

Hypothesis Testing with a Single Random Variable

- One-Sample t-Test

- p -values and Confidence Intervals

- Practical versus Statistical Significance

Biden FTS Example: Test about Population Mean

Let μ_{jb} be the expected value of Joe Biden FTS in the population.

Suppose that we want to test whether $\mu_{jb} = 49$.

Step 1: Specify a **null hypothesis** and an **alternative hypothesis**.

Two possibilities:

- ① $H_0 : \mu_{jb} = 49$ and $H_1 : \mu_{jb} \neq 49$ (two sided alternative)
- ② $H_0 : \mu_{jb} = 49$ and $H_1 : \mu_{jb} > 49$ (one sided alternative)

Step 2: Fix the **significance level** of the test at $\alpha = 0.05$

Source: American National Election Studies. 2021. ANES 2020 Time Series Study Full Release [dataset and documentation]. February 10, 2022 version. electionstudies.org/data-center/2020-time-series-study. Pre-election wave; $n = 8,060$ valid ratings. FTS = feeling thermometer score (0–100) for Joe Biden, the 2020 Democratic presidential candidate. Unweighted: respondents treated as an i.i.d. sample, ignoring the survey weights and sampling design.

One-Sample t -Statistic

Step 3: As the **test statistic**, we use the **one-sample t -statistic** or t -score, i.e. the standardized sample mean:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

where $\bar{X}$ and S are the sample mean and sample standard deviation ($\bar{X}_n$ and S_n in Lecture 2.2), and μ_0 is the “null” value of the parameter μ_{jb} . In our case $\mu_0 = 49$.

Being a function of the sample data, the t -statistic has a sampling distribution under the null hypothesis. This distribution is called the **null distribution**.

We use this distribution to find the probability of a value of T at least as extreme as the observed value t .

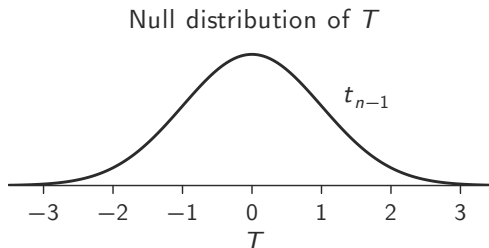
The Null Distribution for the t -Statistic

Step 4: Now let's derive the **null distribution** of the t -statistic.

Setting the null value of μ_{jb} at 49, what is the sampling distribution of T :

- 1 If $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu_{jb}, \sigma^2)$? $\rightarrow T \sim t_{n-1}$
- 2 If $X_1, \dots, X_n$ is i.i.d. with mean μ_{jb} and variance σ^2 , but not normal?
 $\rightarrow T \stackrel{approx.}{\sim} N(0, 1)$ if n is large enough (by CLT).

Commonly we use t_{n-1} , for the reasons discussed previously.



The Null Distribution and Critical Values

Step 5: Finally, obtain the **critical value** c under the null distribution.

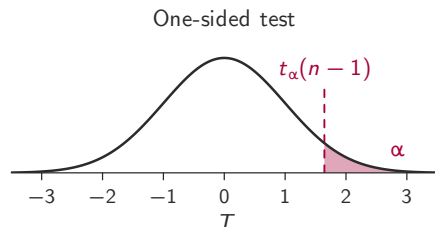
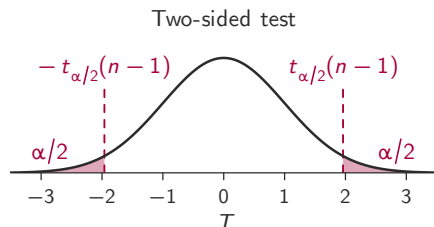
The critical value differs depending on the alternative hypothesis:

- For a **two-sided test** (i.e. $H_1 : \mu_{jb} \neq 49$), $c = t_{\alpha/2}(n-1)$, the $100(1 - \alpha/2)$ percentile of the t_{n-1} distribution.

That is, find c such that $P(|T| > c | H_0) = \alpha$.

- For a **one-sided test** (i.e. $H_1 : \mu_{jb} > 49$), $c = t_{\alpha}(n-1)$, the $100(1 - \alpha)$ percentile of the t_{n-1} distribution.

That is, find c such that $P(T > c | H_0) = \alpha$.

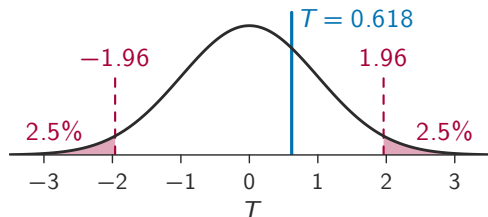


Testing H_0 for the Mean of Biden's FT

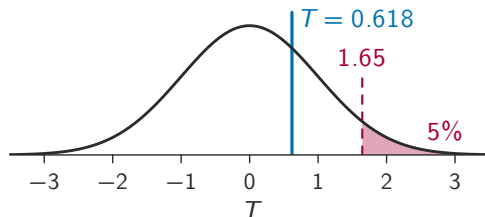
With $n = 8,060$ and $\alpha = .05$, the critical value is $c = 1.96$ for a two sided test and $c = 1.65$ for a one sided test.

Plugging in the sample mean and standard deviation, we obtain $T = 0.618$.

Two-sided test



One-sided test



Would you reject or fail to reject H_0 against the two-sided hypothesis? → Fail to reject.

Would you reject or fail to reject H_0 against the one-sided hypothesis? → Fail to reject.

What is Hypothesis Testing?

Hypothesis Testing with a Single Random Variable

- One-Sample t-Test

- p -values and Confidence Intervals

- Practical versus Statistical Significance

The p -value

The p -value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.

The alternative hypothesis defines what values of the statistic we consider “extreme”:

- For a one-sided test with $H_1 : \mu_{jb} > 49$:

$$p\text{-value} = \Pr(T \geq 0.618 \mid \mu_{jb} = 49) = 0.268$$

Interpretation: If H_0 were true, we would see a value at least as **large** as 0.618 in about 27% of samples.

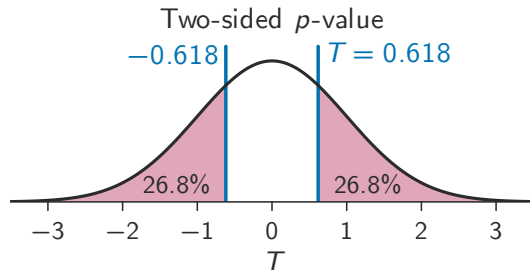
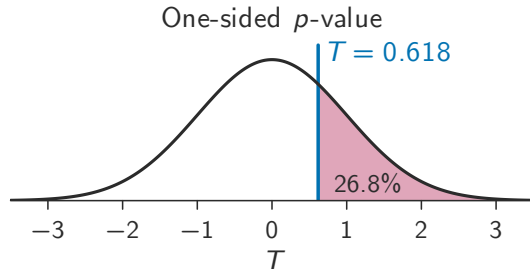
- For a two-sided test with $H_1 : \mu_{jb} \neq 49$:

$$p\text{-value} = \Pr(T \geq 0.618 \text{ or } T \leq -0.618 \mid \mu_{jb} = 49) = 2 \cdot \Pr(T \geq 0.618 \mid \mu_{jb} = 49) = 0.537$$

Interpretation: If H_0 were true, we would see a value at least as **far from zero** as 0.618 in about 54% of samples.

- More generally, p -value is the lowest significance level α for which we can reject H_0 .

The p -values for Biden's FT Test



Inverting a Hypothesis Test

As you may have guessed, there is a strong correspondence between a hypothesis test and a confidence interval.

(*Test- CI equivalence by inversion*) For the two-sided level- α t test of $H_0 : \mu = \mu_0$ (for a generic population mean μ), prove that H_0 is rejected if and only if μ_0 is not contained in the $100(1 - \alpha)\%$ t -confidence interval

$$\left[\bar{X} - t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \frac{S}{\sqrt{n}} \right].$$

Proof (Worksheet Q1)

Inverting a Hypothesis Test

Essentially, hypothesis tests and confidence intervals ask the same question, but from a different perspective:

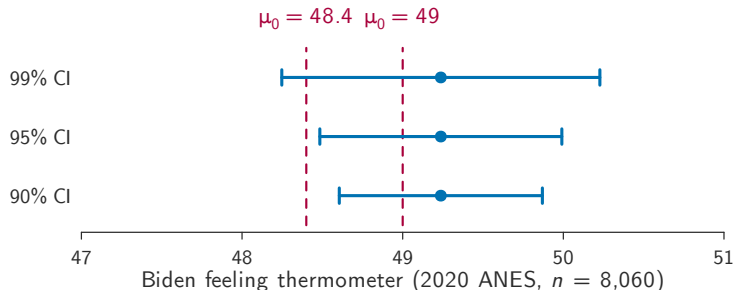
- The test fixes $\mu = \mu_0$ and asks what values of $\bar{X}$ are consistent with that fixed value (and reject μ_0 with which the observed $\bar{X}$ is not consistent)
- The confidence interval fixes the data and collects every value of μ that the test would fail to reject

Using Confidence Intervals for Hypothesis Testing

Therefore, we can conduct a hypothesis test by simply looking at the corresponding confidence interval.

To conduct a two-sided α -level test:

- 1 Construct a $100(1 - \alpha)\%$ confidence interval, using the usual procedure
- 2 If the null value of μ is outside of the confidence interval, reject H_0 .
- 3 If the null value of μ is inside of the confidence interval, fail to reject H_0 .



At $\alpha = .10$ and $.05$, we reject $\mu_0 = 48.4$ but not $\mu_0 = 49$. At $\alpha = .01$, we reject neither.

One Sample T-Test in R

Testing $H_0 : \mu_{jb} = 49$ against $H_1 : \mu_{jb} \neq 49$:

R Code

```
> t.test(x = jbFTS, mu = 49)
```

```
One Sample t-test
```

```
data:  jbFTS
```

```
t = 0.61801, df = 8059, p-value = 0.5366
```

```
alternative hypothesis: true mean is not equal to 49
```

```
95 percent confidence interval:
```

```
 48.48425 49.99069
```

```
sample estimates:
```

```
mean of x
```

```
 49.23747
```

One Sample T-Test in R

Testing $H_0 : \mu_{jb} = 48.4$ against $H_1 : \mu_{jb} \neq 48.4$:

R Code

```
> t.test(x = jbFTS, mu = 48.4)
```

```
One Sample t-test
```

```
data:  jbFTS
```

```
t = 2.1795, df = 8059, p-value = 0.02932
```

```
alternative hypothesis: true mean is not equal to 48.4
```

```
95 percent confidence interval:
```

```
 48.48425 49.99069
```

```
sample estimates:
```

```
mean of x
```

```
 49.23747
```

Let $Y_i = a + bX_i$ with $b \neq 0$. Consider testing $H_0 : \mu = \mu_0$ using the t statistic based on X_i , and testing $H'_0 : \mu_Y = a + b\mu_0$ using the t statistic based on Y_i . Show that both two-sided tests yield the same rejection decision.

Proof (Worksheet Q2)

What is Hypothesis Testing?

Hypothesis Testing with a Single Random Variable

- One-Sample t-Test

- p -values and Confidence Intervals

- Practical versus Statistical Significance

Practical versus Statistical Significance

Under H_0 (and normal data):

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

- What are the possible reasons for rejecting the null?
 - ① Maybe $\bar{X} - \mu_0$ is large (big difference between sample mean and mean assumed by H_0)
 - ② Maybe n is large (you have a lot of data so you have a lot of precision)
 - ③ Maybe S is small (the data vary little around their mean, as they do when σ is small)
 - ④ Maybe H_0 is true and we drew an unusual sample (this happens with probability α)
- We need to be careful to distinguish:
 - **practical significance** (e.g. a big effect)
 - **statistical significance** (i.e. we reject the null)
- In large samples even tiny nonzero differences are likely to be statistically significant, though they may not matter substantively. Always discuss both!

Power Depends on Effect Size, Variance, and n

The **power** of a test is $1 - \beta = \Pr(\text{reject } H_0 \mid H_0 \text{ false})$. It depends on the true value of μ .

Assume $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$ with known σ , and test $H_0 : \mu = \mu_0$ vs. two-sided H_1 using the z statistic $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, rejecting for $|Z| > z_{\alpha/2}$. Derive the power function at $\mu = \mu_0 + \delta$ and show: (i) for any fixed $\delta \neq 0$, power $\rightarrow 1$ as $n \rightarrow \infty$; (ii) for fixed n , power increases in $|\delta|$ and decreases in σ .

Worksheet Q3 (optional)