Recitation Week 2: Introduction to Statistical Inference

Licheng Liu

Massachusetts Institute of Technology

Causal Inference (6.S059/15.C08/17.C08)
February 23, 2024

1. Point and Interval Estimation
   - Point Estimation
   - Interval Estimation

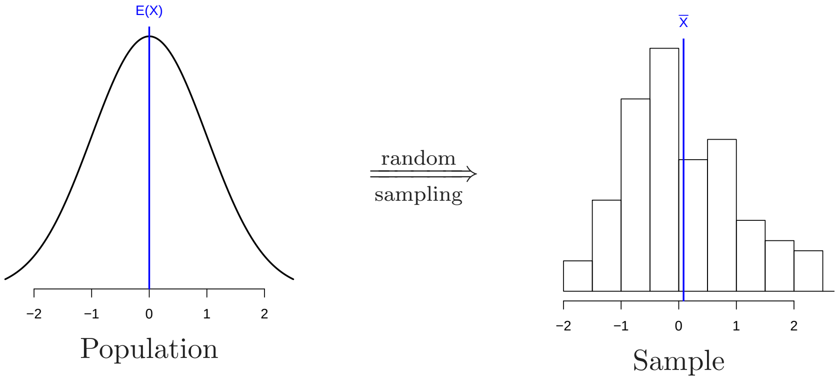
2. What is Hypothesis Testing?

3. Theory and Mechanics of Hypothesis Testing
   - t-Test Procedure
   - *p*-Values and Confidence Intervals
   - Practical versus Statistical Significance

4. Hypothesis Testing for Randomized Experiments

Sample and Population

Recall: Our typical goal is to *make inference about the population from a sample*.



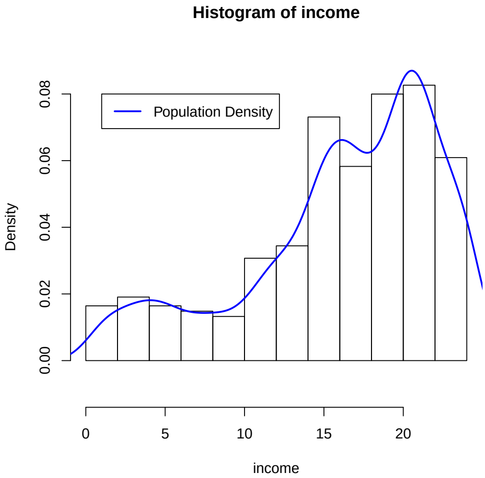
Population mean: $\mathbb{E}[X]$
Population variance: $\mathbb{V}(X)$

Sample mean: $\overline{X}$
Sample variance: S_X^2

Estimands, Estimators, and Estimates

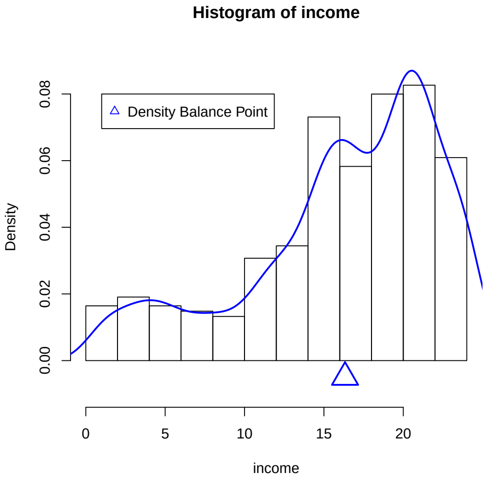
- The goal of statistical inference is to learn about the unobserved population distribution, which can be characterized by parameters.
- Estimands are the parameters that we aim to estimate (e.g. $\mathbb{E}[X]$, ATE).
- Estimators are functions of sample data which we use to learn about the estimands (e.g. $\overline{X}$, sample difference in means).
- Estimates are particular values of estimators that are realized in a given sample dataset.

Example: Income of the U.S. Households in 2008



Population Distribution with mean $\mathbb{E}[Y]$ and variance $\mathbb{V}(Y)$

Defining the Estimand



Let's consider the problem of using sample data from this population to learn about its mean (i.e. the true mean), μ.

Estimators for the Population Mean

Candidate estimators:

1. $\widehat{\mu}_1 = Y_1$ (the first data observation)
2. $\widehat{\mu}_2 = 4$ (the month I was born)
3. $\widehat{\mu}_3 = \overline{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n)$ (the sample average)
4. $\widehat{\mu}_4 = \widetilde{Y}_n \equiv \frac{1}{n+5}(Y_1 + \cdots + Y_n)$ (an "adjusted" sample average)

How do we choose between these estimators?

Estimators and Sampling Distribution

- Estimators are evaluated in terms of their behavior over hypothetical repeated sampling from the population.
    - If we did the sampling over and over again, would our estimator get our target estimand correctly on average? (unbiasedness)

      Mathematically, $\mathbb{E}[\hat{\mu}] = \mu$?
    - If we did the sampling over and over again, how much would our estimator vary from sample to sample? (efficiency)

      Mathematically, $\mathbb{V}(\hat{\mu}_1) < \mathbb{V}(\hat{\mu}_2)$?
- Distribution of an estimator over repeated sampling is called the sampling distribution.
- Standard deviation of the sampling distribution is called the standard error. Efficient estimators have small standard errors.
- Good estimators are unbiased (or approximately so) and efficient compared to other possible estimators.

Simulating Sampling Distributions

Let's simulate in R the sampling distribution of our four candidate estimators for the population mean:

1. $\widehat{\mu}_1 = Y_1$
2. $\widehat{\mu}_2 = 4$
3. $\widehat{\mu}_3 = \overline{Y}_n \equiv \frac{1}{n}(Y_1 + \cdots + Y_n)$
4. $\widehat{\mu}_4 = \frac{1}{n+5}(Y_1 + \cdots + Y_n)$

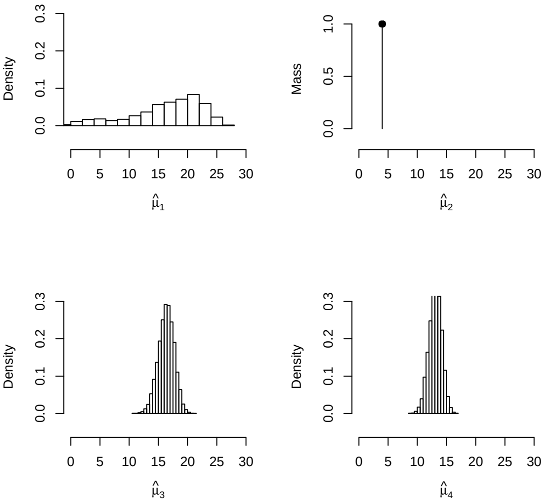
Algorithm to simulate the sampling distribution of $\widehat{\mu}$:

1. Randomly sample m datasets of size n with replacement from the population:

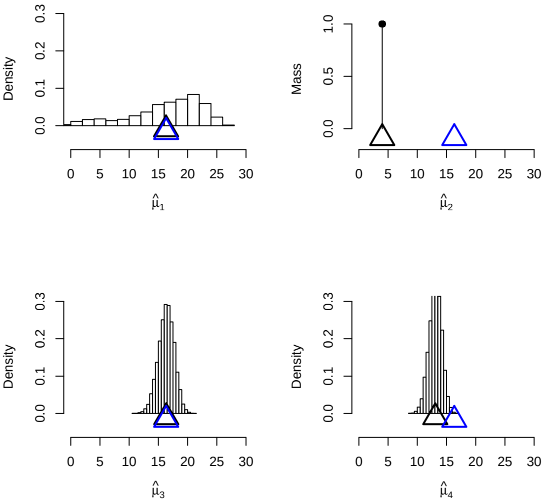
   ```
   replicate(m, sample(population, size=n, replace=TRUE))
   ```

   (m should be large, choose n as desired)
2. Calculate the estimates $\widehat{\mu}_1$, ..., $\widehat{\mu}_4$ for each sampled dataset
3. Plot the distribution of estimates with a histogram

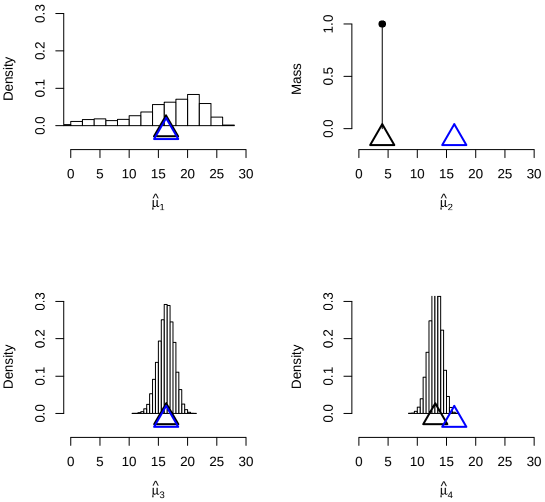
Sampling Distributions ($m = 10000$; $n = 20$)



Which Estimators Are Unbiased?



Which Estimators Are More Efficient Than Others?

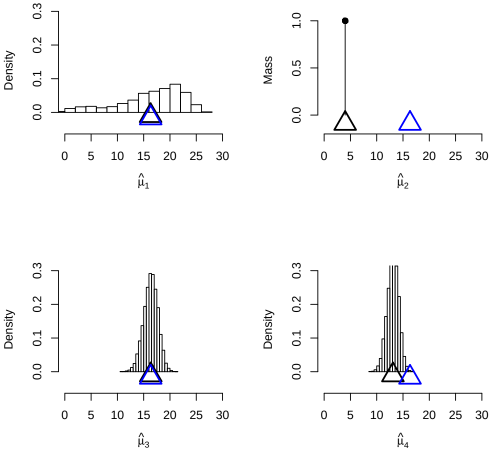


Estimators and Asymptotic Properties

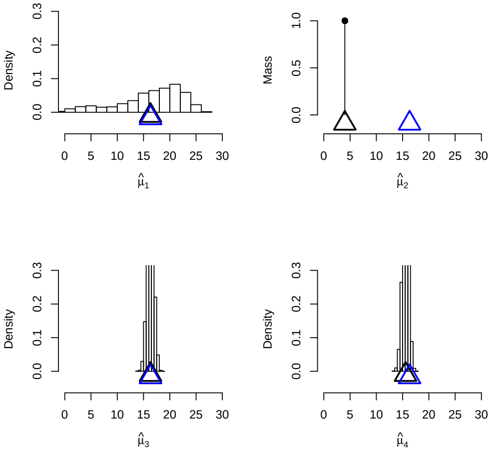
- Another important way of evaluating estimators is via their asymptotic properties, i.e. what happens when the sample size grows to infinity.
  - If $n \to \infty$, would our estimator get closer and closer to the estimand? (consistency)
  - If $n \to \infty$, would the distribution of our estimator become closer and closer to the normal distribution? (asymptotic normality)
- Good estimators are consistent and asymptotically normal.
- We have seen that the sample mean is:
  - consistent for the population mean (LLN!), and
  - asymptotically normal (CLT!), so

  we can conclude that the sample mean is a good estimator for the population mean.

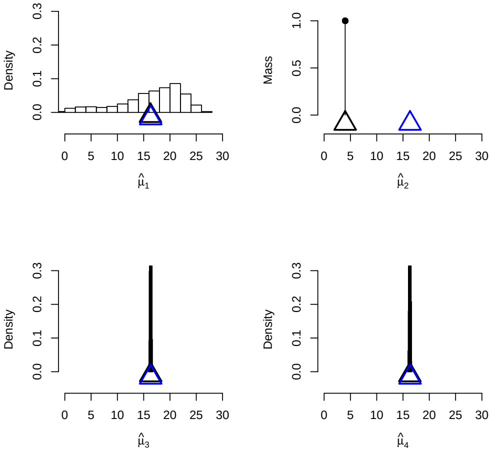
Sampling Distributions of Estimators ($N = 20$)



Sampling Distributions of Estimators ($N = 100$)



Sampling Distributions of Estimators ($N = 3000$)



Therefore,

$\widehat{\mu}_1$: Unbiased but inconsistent $\widehat{\mu}_2$: Biased and inconsistent
$\widehat{\mu}_3$: Unbiased and consistent $\widehat{\mu}_4$: Biased but consistent

Interval Estimation

- So far we have talked about point estimators, i.e., a single number (scalar) that represents our best guess about the parameter.
- However, because we are dealing with a random sample, we might also want to report uncertainty in our estimate.
- An interval estimator for θ takes the following form:

$$[\hat{\theta}_{lower}, \hat{\theta}_{upper}]$$

where $\hat{\theta}_{lower}$ and $\hat{\theta}_{upper}$ are constructed from data and therefore vary from sample to sample.

- The interval represents the range of possible values within which we estimate the true value of θ to fall.

Confidence Interval

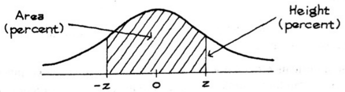
- The most common type of interval estimator is confidence intervals.
- A confidence interval (CI) is formed around a point estimate $\hat{\theta}$ and takes the following form:

$$[\hat{\theta} - 1.64 \times SE(\hat{\theta}), \quad \hat{\theta} + 1.64 \times SE(\hat{\theta})] \quad (90\% \text{ CI})$$

$$[\hat{\theta} - 1.96 \times SE(\hat{\theta}), \quad \hat{\theta} + 1.96 \times SE(\hat{\theta})] \quad (95\% \text{ CI})$$

$$[\hat{\theta} - 2.58 \times SE(\hat{\theta}), \quad \hat{\theta} + 2.58 \times SE(\hat{\theta})] \quad (99\% \text{ CI})$$

  where $SE(\hat{\theta})$ is the standard error of $\hat{\theta}$
- 1.64, 1.96, 2.58 — these numbers are called critical values
- Where do they come from? Recall the normal table...



A NORMAL TABLE

z	Height	Area	z	Height	Area	z	Height	Area
0.00	39.89	0	1.50	12.95	86.64	3.00	0.443	99.730
0.05	39.84	3.99	1.55	12.00	87.89	3.05	0.381	99.771
0.10	39.69	7.97	1.60	11.09	89.04	3.10	0.327	99.806
0.15	39.45	11.92	1.65	10.23	90.11	3.15	0.279	99.837
0.20	39.10	15.85	1.70	9.40	91.09	3.20	0.238	99.863
0.25	38.67	19.74	1.75	8.63	91.99	3.25	0.203	99.885
0.30	38.14	23.58	1.80	7.90	92.81	3.30	0.172	99.903
0.35	37.52	27.37	1.85	7.21	93.57	3.35	0.146	99.919
0.40	36.83	31.08	1.90	6.56	94.26	3.40	0.123	99.933
0.45	36.05	34.73	1.95	5.96	94.88	3.45	0.104	99.944
0.50	35.21	38.29	2.00	5.40	95.45	3.50	0.087	99.953
0.55	34.29	41.77	2.05	4.88	95.96	3.55	0.073	99.961
0.60	33.32	45.15	2.10	4.40	96.43	3.60	0.061	99.968
0.65	32.30	48.43	2.15	3.96	96.84	3.65	0.051	99.974
0.70	31.23	51.61	2.20	3.55	97.22	3.70	0.042	99.978
0.75	30.11	54.67	2.25	3.17	97.56	3.75	0.035	99.982
0.80	28.97	57.63	2.30	2.83	97.86	3.80	0.029	99.986
0.85	27.80	60.47	2.35	2.52	98.12	3.85	0.024	99.988
0.90	26.61	63.19	2.40	2.24	98.36	3.90	0.020	99.990
0.95	25.41	65.79	2.45	1.98	98.57	3.95	0.016	99.992
1.00	24.20	68.27	2.50	1.75	98.76	4.00	0.013	99.9937
1.05	22.99	70.63	2.55	1.54	98.92	4.05	0.011	99.9949
1.10	21.79	72.87	2.60	1.36	99.07	4.10	0.009	99.9959
1.15	20.59	74.99	2.65	1.19	99.20	4.15	0.007	99.9967
1.20	19.42	76.99	2.70	1.04	99.31	4.20	0.006	99.9973

$$[\hat{\theta} - 1.64SE(\hat{\theta}),\ \hat{\theta} + 1.64SE(\hat{\theta})]$$
$$[\hat{\theta} - 1.96SE(\hat{\theta}),\ \hat{\theta} + 1.96SE(\hat{\theta})]$$
$$[\hat{\theta} - 2.58SE(\hat{\theta}),\ \hat{\theta} + 2.58SE(\hat{\theta})]$$

- The normal table is in terms of the standard unit. Imagine the same table, but on the original unit on the x-axis.
- If $\hat{\theta}$ has a sampling distribution that is approximately normal, then the XX% CI contains the true θ about XX% of the time over repeated sampling.
- *Note that it is confidence intervals that randomly vary from sample to sample, not the true θ*

Random Intervals, A Fixed Parameter

The last point from the previous slide is worth reiterating:

> WRONG: "My 95% CI is -2 to 4. Therefore, the true effect is greater than -2 or less than 4 with 0.95 chance."
>
> CORRECT: "My 95% CI is -2 to 4 based on the sample I happen to get this time. If I did the sampling again, I would get another 95% CI, say -1.5 to 3.8. Or -3 to 2.2. Regardless, the way I calculate my CI guarantees that 95% of those specific CIs would contain the true effect."

Still don't quite get it? Here you go:

http://rpsychologist.com/d3/CI/

Confidence Interval for the Mean

- A simple random sample of size n, $\{X_1, X_2, ..., X_n\}$
- Target parameter: $\mathbb{E}[X]$ (population mean)
- Estimator: $\overline{X}$ (sample mean)

If we have a large sample, $\overline{X}$ is approximately normal, so the 95% CI is

$$[\overline{X} - 1.96 \times SE(\overline{X}), \quad \overline{X} + 1.96 \times SE(\overline{X})]$$

What is $SE(\overline{X})$? Let's derive:

$$SE(\overline{X}) = \sqrt{\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i)} = \sqrt{\frac{n\mathbb{V}(X)}{n^2}} = \frac{SD(X)}{\sqrt{n}}$$

Problem: We do not know the true $SD(X)$, so we need to also estimate it.

The formula for a 95% CI for $\mathbb{E}[X]$ with a large sample is therefore:

$$[\overline{X} - 1.96 \times S_X/\sqrt{n}, \quad \overline{X} + 1.96 \times S_X/\sqrt{n}],$$

where S_X is the sample standard deviation of the variable X.

Confidence Interval for the Mean: Variants

The basic formula:

$$[\overline{X} - 1.96 \times S_X/\sqrt{n}, \quad \overline{X} + 1.96 \times S_X/\sqrt{n}]$$

- What if we want a different confidence level, e.g. 90%?
  $\rightarrow$ Change the critical value: $[\overline{X} - 1.64 \times S_X/\sqrt{n}, \ \overline{X} + 1.64 \times S_X/\sqrt{n}]$

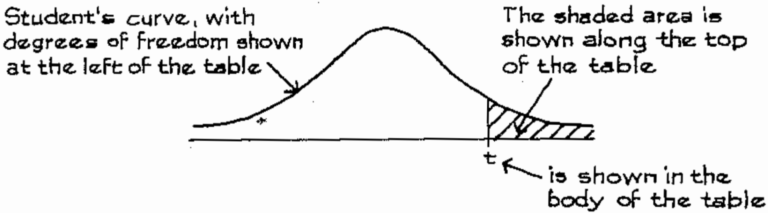
- What if we only have a small sample size?
  $\rightarrow$ Estimation impossible, because we can no longer use the CLT...
  However, if X itself is normally distributed, we can replace the normal table with the t table and construct a CI:

  $$[\overline{X} - t_{(n-1,2.5\%)} \times S_X/\sqrt{n}, \quad \overline{X} + t_{(n-1,2.5\%)} \times S_X/\sqrt{n}],$$

  where $t_{(n-1,2.5\%)}$ is the critical value from the t table with $n-1$ degrees of freedom (df).

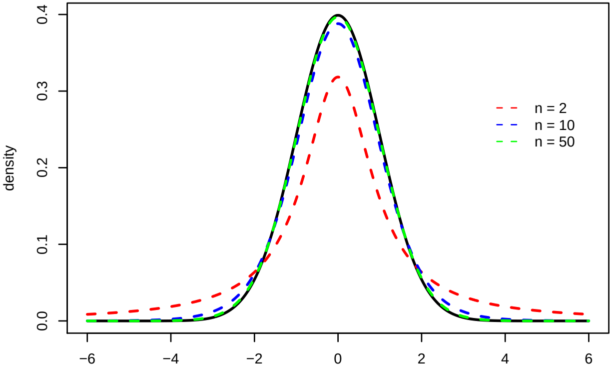
The t Table

A t-TABLE



Degrees of freedom	25%	10%	5%	2.5%	1%	0.5%
1	1.00	3.08	6.31	12.71	31.82	63.66
2	0.82	1.89	2.92	4.30	6.96	9.92
3	0.76	1.64	2.35	3.18	4.54	5.84
4	0.74	1.53	2.13	2.78	3.75	4.60
5	0.73	1.48	2.02	2.57	3.36	4.03

Student's t Distribution



- The t distribution has fatter tails than the normal distribution, and it converges to the normal distribution as df $\to \infty$.
- Therefore, it is "safer" to always use the t CI instead of the normal CI.
- R functions: `rt(n, df)`, `dt(x, df)`, `pt(q, df)`

Confidence Interval for the Difference in Means

- Now let's apply the same logic to the difference in means, our estimator of ATE in a randomized experiment.

- Target parameter: $ATE = \mathbb{E}[Y_{1i} - Y_{0i}]$
- Estimator: $\overline{Y}_1 - \overline{Y}_0$

If sample size is large, we can apply the CLT, and so a 95% CI for ATE is:

$$[(\overline{Y}_1 - \overline{Y}_0) - 1.96 \times SE(\overline{Y}_1 - \overline{Y}_0),\ (\overline{Y}_1 - \overline{Y}_0) + 1.96 \times SE(\overline{Y}_1 - \overline{Y}_0)]$$

Next, derive the SE:

$$SE(\overline{Y}_1 - \overline{Y}_0) = \sqrt{\mathbb{V}(\overline{Y}_1) + \mathbb{V}(\overline{Y}_0)} = \sqrt{\frac{\mathbb{V}(Y_{1i})}{n_1} + \frac{\mathbb{V}(Y_{0i})}{n_0}}$$

So our formula for the CI is:

$$\left[(\overline{Y}_1 - \overline{Y}_0) - 1.96\sqrt{S^2_{Y_1}/n_1 + S^2_{Y_0}/n_0},\ (\overline{Y}_1 - \overline{Y}_0) + 1.96\sqrt{S^2_{Y_1}/n_1 + S^2_{Y_0}/n_0}\right]$$

If the CI doesn't include zero, then it is said that the effect is statistically significant.

Hypothesis Testing: The Trial Analogy

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

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:

		Defendant	
		Guilty	Innocent
Decision	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, creating a decision rule which minimizes both types of errors at the same time is impossible. We therefore need to balance them.

Two Types of Errors in Hypothesis Testing

		Defendant	
		Guilty	Innocent
Decision	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 $\alpha = \Pr(\text{reject } H_0 \mid H_0 \text{ is true})$ (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	$1-\alpha$

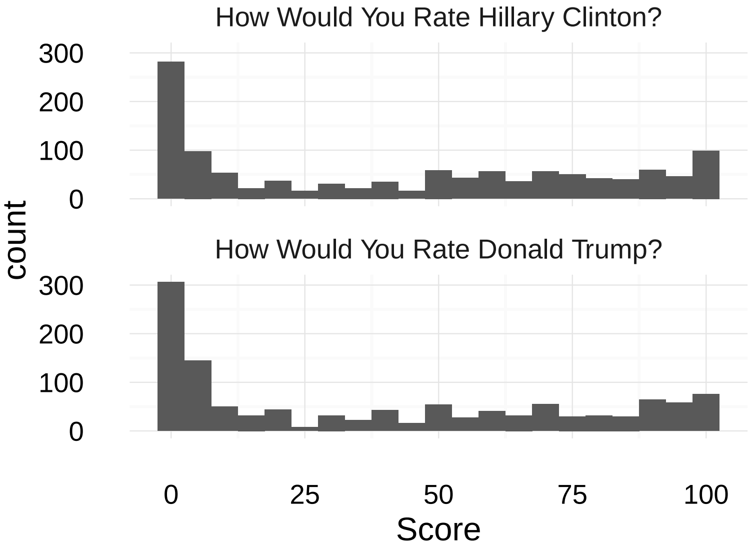
5. Using the critical values from a statistical table, evaluate how unusual the observed value of T is under the null hypothesis:
    - If the probability of drawing a T at least as extreme as the observed T is less than α, we reject H_0.
      (e.g. there are too many testimonies against the defendant for her to be 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.)

1. Point and Interval Estimation
   - Point Estimation
   - Interval Estimation
2. What is Hypothesis Testing?
3. Theory and Mechanics of Hypothesis Testing
   - t-Test Procedure
   - *p*-Values and Confidence Intervals
   - Practical versus Statistical Significance
4. Hypothesis Testing for Randomized Experiments

ANES - Feeling Thermometers

I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person.
If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one.

Empirical Distribution of FT Scores



Trump FTS Example: Test about Population Mean

Let μ_{dt} be the expected value of Donald Trump's FTS in the population. Suppose that we want to test whether $\mu_{dt} = 38$.

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

$$H_0 : \mu_{dt} = 38 \quad \text{v.s.} \quad H_1 : \mu_{dt} \neq 38$$

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

One-Sample t-Statistic

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

$$T = \frac{\overline{X} - 38}{S_X / \sqrt{n}}$$

where $\overline{X}$ is the sample mean and S_X is the sample standard deviation.

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 determine the probability of observing a realized value of T.

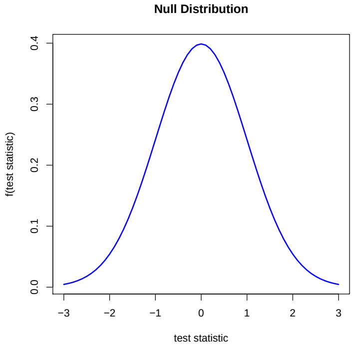
The Null Distribution for the t-Statistic

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

Recalling the previous lecture, what is the sampling distribution of T if...

1. The sample size n is large? $\rightarrow$ Normal with mean 0, variance 1.
2. n is small, but $X_1, ..., X_n$ themselves are normally distributed?
   $\rightarrow$ the t distribution with $n-1$ degrees of freedom.

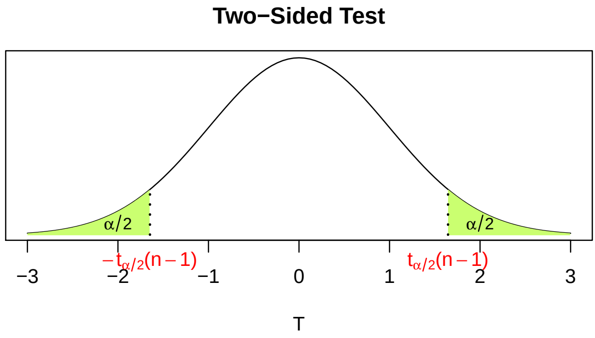
For the reason we discussed, we usually use the t distribution:



The Null Distribution and Critical Values

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

With $\alpha = 0.05$, we are looking for c such that the probability of getting T more extreme than c is 0.05.

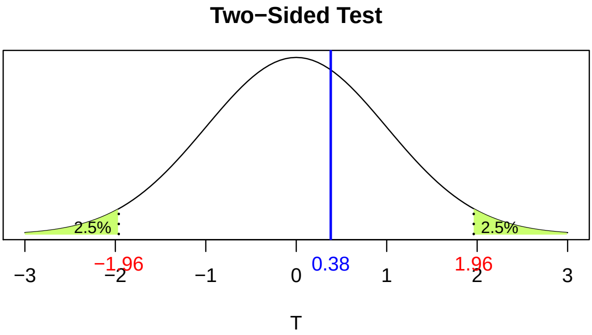


Testing H_0 for the Mean of Edwards' FT

With $\alpha = .05$, the critical value is $c = 1.96$ for a two sided test.

Plugging in $n = 1196$, $\mu_{dt} = 38$, $\overline{X} = 38.4$ and $S_X = 36.5$, we obtain:

$$T = 0.38$$



Would you reject or fail to reject H_0 against the alternative $\mu_{dt} \neq 38$?

$\rightarrow$ Fail to reject.

p-Value

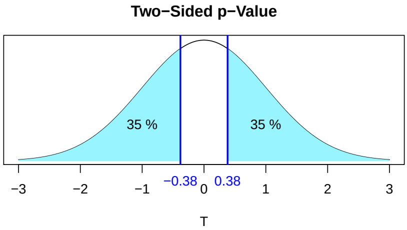
The *p-value* is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.

In other words, p-value is the largest significance level α at which we would carry out the test and still fail to reject H_0.

For our test with $H_1 : \mu_{dt} \neq 38$:

$$\begin{aligned} \text{p-value} &= \Pr(T \geq 0.38 \text{ or } T \leq -0.38 \mid \mu_{dt} = 38) \\ &= 2 \cdot \Pr(T \geq 0.38 \middle| \mu_{je} = 38) \ = \ 0.70 \end{aligned}$$

Interpretation: If H_0 were true, we would expect a value as distant from zero as 0.38 or -0.38 about 70% of the time.



Inverting a Hypothesis Test

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

We would "accept" the null hypothesis $\mu = \mu_0$ at the 0.05 level if:

$$-1.96 \leq \frac{\overline{X} - \mu_0}{S_X/\sqrt{n}} \leq 1.96$$

Rearrange the terms to obtain:

$$\overline{X} - 1.96 \times S_X/\sqrt{n} \leq \mu_0 \leq \overline{X} + 1.96 \times S_X/\sqrt{n}$$

This is simply the 95% confidence interval for μ!

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

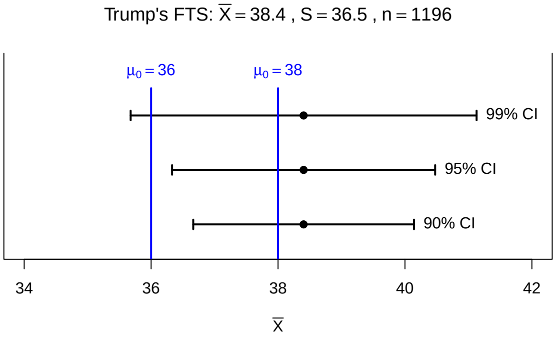
- The test fixes $\mu = \mu_0$ and asks what values of $\overline{X}$ are consistent with that fixed value (and reject μ_0 with which the observed $\overline{X}$ is not consistent)
- The confidence interval fixes $\overline{X}$ and asks what values of μ make this sample value most plausible

Using Confidence Intervals for Hypothesis Testing

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

To conduct a test at the 0.01/0.05/0.1 level:

1. Construct a 99/95/90% 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.



The t-Test in R

Testing $H_0 : \mu_0 = 38$ against $H_1 : \mu_0 \neq 38$:

R Code

```
> t.test(dtFTS, mu = 38)

        One Sample t-test

data:  dtFTS
t = 0.38149, df = 1195, p-value = 0.7029
alternative hypothesis: true mean is not equal to 38
95 percent confidence interval:
 36.33037 40.47565
sample estimates:
mean of x
 38.40301
```

Practical versus Statistical Significance

$$\text{The t-Statistic} = \frac{\overline{X} - \mu_0}{S_X/\sqrt{n}}$$

- What are the possible reasons for rejecting the null?
  1. Maybe $\overline{X} - \mu_0$ 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_X is small (the population variance is small)
- 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 effects will be significant, but the results may not be very important substantively. Always discuss both!

Testing Equality of Means

- We can use the same framework to conduct other tests.
- In particular, we can test equality of means for two independent variables:

$$Y_A \sim \text{unknown distribution with } \mathbb{E}[Y_A] = \mu_A \text{ and } \mathbb{V}[Y_A] = \sigma_A^2$$

$$Y_B \sim \text{unknown distribution with } \mathbb{E}[Y_B] = \mu_B \text{ and } \mathbb{V}[Y_B] = \sigma_B^2$$

- Y_A and Y_B may be samples from two different populations, or outcomes from two different treatment groups.
- Suppose we have groups with sizes N_A and N_B
- Assuming N_A and N_B are reasonably large,

$$\bar{Y}_A \sim \text{normal distribution with } \mathbb{E}[\bar{Y}_A] = \mu_A \text{ and } \mathbb{V}[\bar{Y}_A] = \sigma_A^2/N_A$$

$$\bar{Y}_B \sim \text{normal distribution with } \mathbb{E}[\bar{Y}_B] = \mu_B \text{ and } \mathbb{V}[\bar{Y}_B] = \sigma_B^2/N_B$$

because of the CLT.

Two-Sample t-Test

- We test H_0: $\mu_A = \mu_B$ vs. H_1: $\mu_A \neq \mu_B$
- Under H_0, we have:

$$\bar{Y}_A - \bar{Y}_B \;\sim\; \text{normal distribution}$$

with

$$\text{mean} \;=\; \mu_A - \mu_B \;=\; 0 \quad \text{and} \quad \text{variance} \;=\; \frac{\sigma_A^2}{N_A} + \frac{\sigma_B^2}{N_B}$$

- This implies the two-sample t-statistic:

$$T \;\equiv\; \frac{\bar{Y}_A - \bar{Y}_B}{\sqrt{S_A^2/N_A \;+\; S_B^2/N_B}} \;\sim\; \text{normal with mean 0, variance 1}$$

- Then as usual, reject H_0 at the 0.05 level if $T < -1.96$ or $T > 1.96$.
- For a small sample you should use an appropriate t-distribution instead of normal. This version is called Welch's two-sample t-test.

Example: Testing Equality of Means

Was Donald Trump as popular as Hillary Clinton?

R Code

```
> mean(hcFTS) [1] 42.98746
> mean(dtFTS) [1] 38.40301
```

Two sample *t*-test in R:

R Code

```
> t.test(x = dtFTS, y = hcFTS)

        Welch Two Sample t-test

data:  dtFTS and hcFTS
t = -3.0702, df = 2390, p-value = 0.002163
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.512533 -1.656364
sample estimates:
mean of x mean of y
 38.40301  42.98746
```

Experiment Example: Party Brands and Party Attachment

- How and when do people feel closer to a politicay party?
- Theory: Partisanship is like brand loyalty (Lupu 2012)
- Hypothesis: Information that makes a party more distinguishable from rival parties should strengthen attachment of its supporters
- Experiment: A face-to-face survey of Argentine voters in 2009

> This page contains some information about some political parties in today's Argentina. Please look over the information and I will then ask you some questions about these political parties. Please take your time and let me know when you have finished reading the information.

- Treatment: Type of information shown along with party names

Justicialist Party (PJ)
Leader: Daniel Scioli

- Party that implemented the economic reforms of the 1990s with the support of the UCR
- Formed a pact with the UCR in 1993 to support the constitutional reforms
- Supports government intervention in the economy
- Supports increasing the powers of the provinces
- Proposes an electoral reform for obligatory primary elections

Radical Civic Union (UCR)
Leader: Gerardo Morales

- Formed an alliance with FREPASO in the 1990s
- Signed a pact with the PJ in 1993 to support the constitutional reforms
- Currently forms part of an alliance with ARI and several other parties
- Believes in a market-based economy, with the state ensuring equality of opportunity
- Supports reducing taxes on small and medium-sized businesses
- Supports strengthening the police

Affirmation for a Republic of Equals (ARI)
Leader: Elisa Carrió

- Made up primarily of former Radicals
- Maintains alliances with the UCR and Socialist Party
- Emphasizes a more equal distribution of income
- Supports increasing the powers of the Congress
- Supports limiting the powers of the president and of the provinces

Socialist Party
Leader: Rubén Giustiniani

- Forms alliances with ARI, the UCR, and the PDP
- In 2007 allied with Southern Project to support filmmaker Pino Solanas for president
- Believes in a socialist economy
- Emphasizes the redistribution of income for a society without classes
- Opposes international financial institutions
- Emphasizes municipal autonomy
- Believes in a right to employment

Progressive Democratic Party (PDP)
Leader: Alberto Natale

- In 2009, formed alliances with ARI, the UCR, and the Socialist Party
- Supported the Radical candidate for president in 2007
- Proposes declaring a state of agricultural emergency
- Believes in a market-based economy, with the state ensuring equality of opportunity

Recreate for Growth
Leader: Esteban Bullrich

- Maintains alliances with factions of the PJ and PRO
- Believes in an economy based on the free market and competition
- Supports ensuring the autonomy of the INDEC and the Central Bank
- Emphasizes fiscal discipline and reducing government spending
- Supports strengthening judicial security

New Party
Leader: Luis Alfredo Juez

- Composed primarily of former members of the PJ and UCR
- In 2009, formed alliances with the Socialist Party and ARI
- Supports increasing the powers of the provinces
- Supports ensuring the autonomy of the INDEC
- Believes in a market-based economy, with the state ensuring equality of opportunity

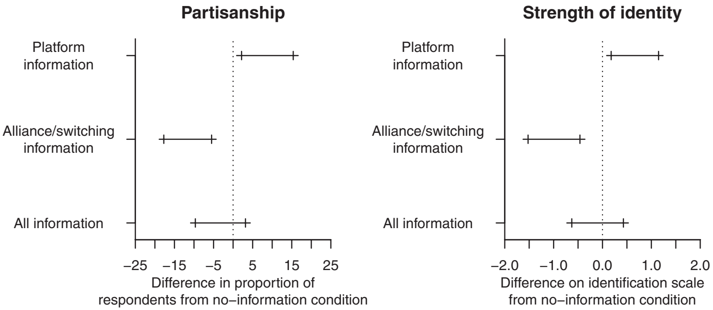
Republican Proposal (PRO)
Leader: Mauricio Macri

- Maintains alliances with factions of the PJ and Recreate
- Believes in an economy based on the free market and competition
- Emphasizes the fight against crime and supports strengthening the security forces
- Proposes limiting the use of public funds for electoral campaigns

Control: No information
Platform: Info emphasizing differences
Alliance: Info blurring differences
Both information

Result: Point Estimates and Confidence Intervals

FIGURE 2 Average Treatment Effects



Notes: Average treatment effects calculated as the difference between each treatment condition and the control (no information) condition. Values in the left panel refer to whether or not respondents said they identified with a political party in each condition. Values in the right panel refer to the average strength of partisans' identification on a 0–10 scale. Horizontal lines represent 95% confidence intervals, while vertical bars indicate 90% confidence intervals.

Assignments

Next Week:

- Feb 28 3:00 PM: Problem Set 1 Due
- Feb 28 4:30 PM: Problem Set 2 Released (due Mar 18)