Ec140 - Randomized Controlled Trials and Statistical Inference

Fernando Hoces la Guardia
07/06/2022

Today's Lecture

- Finish RCTs

- Review of Statistical Inference

  - Standard deviation of the sample mean
  - Distribution of the sample mean
  - Distribution for the difference in means
  - Hypothesis testing
  - P-values
  - Confidence intervals
  - P-hacking

- How about a population that is more relevant to current policy debates (in the US)?
- Expanding Medicaid leads to less costs? Does it improve health?
- Oregon implemented an RCT unintentionally when they decided to expand Medicaid to a broader population.
- This expansion of the Oregon Health Plan (OHP) was later studied to learn about use of medical services and health outcomes.

Example #3: Orengon Health Plan (OHP) RCT 2/2

- Year: 2008
- Population:
  - Residents of Oregon
  - Under the poverty line and not eligible for Medicaid (non-disabled, non-children, non-pregnant)
  - $n = 75,000$; $30,000$ into an "invitation" treatment.

TABLE 1.5

OHP effects on insurance coverage and health-care use

	Oregon		Portland area	
	Control mean	Treatment effect	Control mean	Treatment effect
Outcome	(1)	(2)	(3)	(4)
A. Administrative data				
Ever on Medicaid	.141	.256 (.004)	.151	.247 (.006)
Any hospital admissions	.067	.005 (.002)		
Any emergency department visit			.345	.017 (.006)
Number of emergency department visits			1.02	.101 (.029)
Sample size	74,922		24,646	
B. Survey data				
Outpatient visits (in the past 6 months)	1.91	.314 (.054)		
Any prescriptions?	.637	.025 (.008)		
Sample size	23,741			

Notes: This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on insurance coverage and use of health care. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard errors are reported in parentheses.

TABLE 1.6
OHP effects on health indicators and financial health

	Oregon		Portland area	
	Control mean	Treatment effect	Control mean	Treatment effect
Outcome	(1)	(2)	(3)	(4)
A. Health indicators				
Health is good	.548	.039 (.008)		
Physical health index			45.5	.29 (.21)
Mental health index			44.4	.47 (.24)
Cholesterol			204	.53 (.69)
Systolic blood pressure (mm Hg)			119	−.13 (.30)
B. Financial health				
Medical expenditures >30% of income			.055	−.011 (.005)
Any medical debt?			.568	−.032 (.010)
Sample size	23,741		12,229	

Notes: This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on health indicators and financial health. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard

- First: not all who won the lottery got insurance. So the first thing to look at is the effect of winning the lottery on getting insurance (Medicaid).
- Second, the results show higher utilization of healthcare ss. Problematically, one of the most expensive ones, like emergency visits. After a couple of years since the invitation. It also shows improvements on health, particularly on mental health.
- Both the HIE and OHP suggest no causal effect of HI on physical health in the short run. Both show more utilization. OHP shows improvements on mental health and financial stability (also in the short run). Two, or more, studies finding similar results are much more persuasive than any single study showing a particular result.
- One final issue with the second RCT is that not everybody who was invited ended up receiving the most relevant treatment (HI). Hence the effect of winning on utilization and health are basically pooling a bunch of zeros for those invited that did not get HI, and a larger effect (both in emergency use and in mental health) over those invited that did receive the health insurance treatment. We will learn how to separate these two effects once we study Regression and Instrumental Variables.

TABLE 1.6
OHP effects on health indicators and financial health

	Oregon		Portland area	
	Control mean	Treatment effect	Control mean	Treatment effect
Outcome	(1)	(2)	(3)	(4)
A. Health indicators				
Health is good	.548	.039 (.008)		
Physical health index			45.5	.29 (.21)
Mental health index			44.4	.47 (.24)
Cholesterol			204	.53 (.69)
Systolic blood pressure (mm Hg)			119	−.13 (.30)
B. Financial health				
Medical expenditures >30% of income			.055	−.011 (.005)
Any medical debt?			.568	−.032 (.010)
Sample size	23,741		12,229	

Notes: This table reports estimates of the effect of winning the Oregon Health Plan (OHP) lottery on health indicators and financial health. Odd-numbered columns show control group averages. Even-numbered columns report the regression coefficient on a dummy for lottery winners. Standard

RCTs: Final Considerations

- Sometimes impractical

- Sometimes unethical. The role of informed consent and freedom of participants.

- Sometimes the most ethical option.

- Always a good frame of reference to think about other research designs.

Review of Statistical Inference

- How can we tell if this difference of $198 is due to some observations that happen to appear in our random sample?
- For example: maybe the HIE happen to sample individuals from the general population that are very high spenders, and maybe, just due to chance, those individuals were assigned into the treatment group.

	Means	Differences between plan groups			
	Catastrophic plan (1)	Deductible − catastrophic (2)	Coinsurance − catastrophic (3)	Free − catastrophic (4)	Any insurance − catastrophic (5)
A. Health-care use					
Face-to-face visits	2.78 [5.50]	.19 (.25)	.48 (.24)	1.66 (.25)	.90 (.20)
Outpatient expenses	248 [488]	42 (21)	60 (21)	169 (20)	101 (17)
Hospital admissions	.099 [.379]	.016 (.011)	.002 (.011)	.029 (.010)	.017 (.009)
Inpatient expenses	388 [2,308]	72 (69)	93 (73)	116 (60)	97 (53)
Total expenses	636 [2,535]	114 (79)	152 (85)	285 (72)	198 (63)

- Reminder: a random variable Y_i has a sample variance $(S^2(Y_i))$:

$$s_Y^2 = \frac{\sum_{i=1}^{n} \left(y_i - \overline{Y}\right)^2}{n}$$

- Intuition for sample variance: average of squared deviations from its mean.

- The population variance, or just variance $(V(Y_i))$:

$$Var(Y_i) = \sigma^2 = \mathbb{E}\big((Y_i - \mu)^2\big)$$

- Where μ is defined as the population mean $(E(Y_i))$.

- Both μ and σ^2 represent fixed numbers (not variables).

- We want to use statistical inference to say something about the sample mean $(\overline{Y})$, which is itself a random variable that sums a collection of random variables and divides them by the population size.

- Let's start by giving it a name: the sampling variance.

- Using independence, we can now derive formula for sampling variance.

$$Var(\overline{Y}) = Var\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

- Hence, the sampling variance (the variance of the sample mean) is equal to the variance of the underlying data (σ^2), divided by sample size (n).

- To distinguish between the standard deviation of the sample mean $\left(\frac{\sigma}{\sqrt{n}}\right)$, and the standard deviation of the underlying data (σ), we call the standard deviation of the sample mean the standard error

- In addition to sample mean, we will call standard error, any standard deviation of a statistic that aggregates data.

- **Important:** I was wrong when implying that MM gets confused about this, calling the standard deviation of some statistics "standard error" in the tables of Ch1, and "standard deviation" other statistics.

- For example, let's look at Table 1.1. The standard deviation of square brackets refers to the variation in the underlying data. The standard deviations in parentheses refer to variation in sample means.

Health and demographic characteristics of insured and uninsured couples in the NHIS

	Husbands			Wives		
	Some HI (1)	No HI (2)	Difference (3)	Some HI (4)	No HI (5)	Difference (6)
A. Health						
Health index	4.01 [.93]	3.70 [1.01]	.31 (.03)	4.02 [.92]	3.62 [1.01]	.39 (.04)
B. Characteristics						
Nonwhite	.16	.17	−.01 (.01)	.15	.17	−.02 (.01)
Age	43.98	41.26	2.71 (.29)	42.24	39.62	2.62 (.30)
Education	14.31	11.56	2.74 (.10)	14.44	11.80	2.64 (.11)
Family size	3.50	3.98	−.47 (.05)	3.49	3.93	−.43 (.05)
Employed	.92	.85	.07 (.01)	.77	.56	.21 (.02)
Family income	106,467	45,656	60,810 (1,355)	106,212	46,385	59,828 (1,406)
Sample size	8,114	1,281		8,264	1,131	

Notes: This table reports average characteristics for insured and uninsured married couples in the 2009 National Health Interview Survey (NHIS). Columns (1), (2), (4), and (5) show average characteristics of the group of individuals specified by the column heading.

- To make things clearer: they could have also reported on the standard error of the first sample mean. To do this we just needed to take the standard deviation of the underlying data (e.g., .93 for column 1) and divide it by the square roof of its sample size $(\frac{0.93}{\sqrt{(8,114)}}) = 0.01$.

- Notice that the standard error goes to zero as n grows, but not the standard deviation of the underlying data

Health and demographic characteristics of insured and uninsured couples in the NHIS

	Husbands			Wives		
	Some HI (1)	No HI (2)	Difference (3)	Some HI (4)	No HI (5)	Difference (6)
A. Health						
Health index	4.01 [.93]	3.70 [1.01]	.31 (.03)	4.02 [.92]	3.62 [1.01]	.39 (.04)
B. Characteristics						
Nonwhite	.16	.17	−.01 (.01)	.15	.17	−.02 (.01)
Age	43.98	41.26	2.71 (.29)	42.24	39.62	2.62 (.30)
Education	14.31	11.56	2.74 (.10)	14.44	11.80	2.64 (.11)
Family size	3.50	3.98	−.47 (.05)	3.49	3.93	−.43 (.05)
Employed	.92	.85	.07 (.01)	.77	.56	.21 (.02)
Family income	106,467	45,656	60,810 (1,355)	106,212	46,385	59,828 (1,406)
Sample size	8,114	1,281		8,264	1,131	

Notes: This table reports average characteristics for insured and uninsured married couples in the 2009 National Health Interview Survey (NHIS). Columns (1), (2), (4), and

- Standard errors may be complicated but the idea is simple: they summarize variability in an estimate due to sampling variability.

- The standard error needs to be estimated, hence its estimated version is called... Estimated Standard Errors

- Usually we forget to say the "estimated part" but that is what we are measuring.

- Beyond the names and their eternal confusions, the key idea that I want you to take away is that when looking the sample mean (and other statistics later on) we need to remember that there are two types of standard deviations in them: the standard deviation of the underlying data (σ) and the standard deviation of the sample mean, or any statistic that aggregates this data, called standard errors $\left(\frac{\sigma}{\sqrt{n}}\right)$. One does not shrink to zero as we have more information, the other does.

- We have learned about the mean and standard deviation of the sample mean. But What about its distribution?

- By the CLT, we know that the distribution of the sample mean $(\overline{Y})$ is normal with mean μ and standard deviation $\sigma/\sqrt{n}$. Denoted as $N(\mu, \sigma/\sqrt{n})$

- Let's look again at the normal distribution in Seeing Theory. Particularly, how to move from any $N(\mu, \sigma/\sqrt{n})$ to a $N(0, 1)$.

- Hence, if we take the random variable $\overline{Y}$ and substract its population mean, and divide by its standard deviation, we have a $N(0, 1)$:

$$Z = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n})} \sim N(0, 1)$$

- The reason we standarize, is that we know a lot about $N(0, 1)$:

  - Most of its mass (probability) is between -1 and 1: ~70%
  - Between -2 and 2: ~95%
  - Between -3 and 3: ~99%.

- The probability of observing any value in outside the range -2 to -2 is 5% (or 1 in 20).

- The probability of observing a "3-sigma" event is 1%.

- Everything we have done when characterizing $\overline{Y}$ we can do to a difference of two sample means: $\overline{Y_1} - \overline{Y_0}$
- First, define its population mean as μ.
- Compute its variance:

$$Var(\overline{Y_1} - \overline{Y_0}) = Var(\overline{Y_1}) + Var(\overline{Y_0})$$
$$= \frac{\sigma_Y^2}{n_1} + \frac{\sigma_Y^2}{n_0} = \sigma_Y^2 \left(\frac{1}{n_1} + \frac{1}{n_0} \right)$$

- With its corresponding standard error (SE):

$$SE(\overline{Y_1} - \overline{Y_0}) = \sigma_Y \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_0} \right)}$$

- Analogous to the single sample mean, the SE can be estimated using the estimate for the underlying standard deviation:

$$\hat{SE}(\overline{Y_1} - \overline{Y_0}) = S(Y_i)\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_0}\right)}$$

- Where $S(Y_i)$ is the standard deviation of all the underlying data (pooling Y_1 and Y_0)

- This difference in mean is also an average of (underlying) independent random variables, so CLT applies and we have:

$$(\overline{Y_1} - \overline{Y_0}) \sim N(\mu, SE(\overline{Y_1} - \overline{Y_0}))$$

- This distribution can also be standardized to obtain:

$$Z = \frac{(\overline{Y_1} - \overline{Y_0}) - \mu}{SE(\overline{Y_1} - \overline{Y_0})} \sim N(0,1)$$

- And again, this $N(0,1)$ has the same properties as above.

Hypothesis Testing: Main Idea

- We want to ask if the statistic we observe ($\overline{Y}$ or $(\overline{Y_1} - \overline{Y_0})$) is consistent with some underlying truth, represented by a theoretical distribution.

- Our working hypothesis, or null hypothesis, is that this statistic does come from such truth. Let's define the population mean of that theoretical distribution, as μ_0.

- Assuming that our hypothesis is true, we can again standardize the statistic:

$$\frac{(\overline{Y_1} - \overline{Y_0}) - \mu_0}{SE(\overline{Y_1} - \overline{Y_0})} = t(\mu_0) \sim N(0, 1)$$

- This is called the t-statistic for the null hypothesis μ_0 (given that in small samples as a t-distribution, but in large sample is normal).

- One of the most common null hypothesis is that of no effect $(\mu_0 = 0)$, in this case the t-statistic becomes the ratio of the estimate by its standard error.

- Remember that this statistic is distributed $N(0, 1)$, now assume that we observe $t = -0.15$ what is the probability of observing this statistic or something larger (in absolute value), assuming that the null is true.

(Check out this great explanation on hypothesis testing by Nick Huntington-Klein)
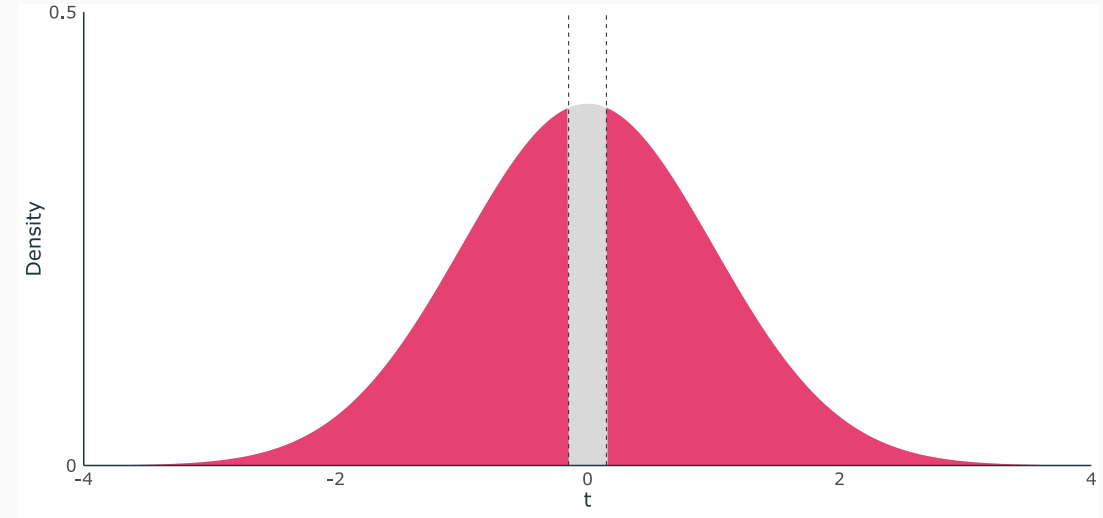
The **p-value** is the probability of observing a t-statistic at least as extreme as the one we observe, given that the hypothesis is true, is $p = 0.88$. So the statistic that we observe seems to be consistent with our null hypothesis. More

- But when does a t-statistic, and its corresponding p-value, stops being consistent with the null hypothesis?

- A **convention** is that if the p-value should be less than 0.05, then it is said to be statistically significant.

- This corresponds to a t-statistic of around 2 in absolute value. Hence the rule of thumb of dividing the estimate by its standard error anc checking if its bigger than 2 in absolute value.