Overview: Analyzing Randomized Control Trials - Quiz

Bookmark this page

Question 1

0 points possible (ungraded)

How did Fisher contribute to the use of randomized control trials? (Select all that apply)

- [ ] Statistical background ✔
- [ ] How we should set up small scale experiments
- [ ] Foundation for thinking about randomized control trials ✔
- [ ] Father of how we think about randomized control trials today

Explanation

Fisher gave us a lot of statistical background that we use today and was the first one who thought strenuously about randomized control trials. However, his thinking about analyzing randomized control trials was a bit different to the way we tend to think about analyzing them today. His approach is probably most appropriate for thinking about very large scale experiments, not small scale experiments.

The Average Treatment Effect - Quiz

Bookmark this page

Finger Exercises due Oct 26, 2020 19:30 EDT **Past Due**

Question 1

0.0/1.0 point (graded)

This is the formula for the average treatment effect:

$$\hat{\tau} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs} - \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{obs}$$

Let

$$A = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs}$$

and

$$B = \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{obs}$$

What do A and B represent?

What is A?

- The outcome for the average individual in the control group
- The sample mean for the control group
- A vector of outcomes for the treatment group
- The sample mean for the treatment group ✔

What is B?

- A vector of outcomes for the treatment group
- The sample mean for the control group ✔
- The outcome for the average individual in the treatment group
- The sample mean for the treatment group

Explanation

In the equation above, the term labeled A represents the sample mean for a specific outcome for the individuals assigned to the treatment group, and the term labeled B represents the sample mean for the individuals assigned to the control group.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 2

1 point possible (graded)

The variance of the average treatment effect is given by

$$V(\hat{\tau}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{tc}^2}{N}$$

.

What does the term $\frac{S_t^2}{N_t}$ represent?

- The average of the variance of Y_i in the treatment group and in the control group
- The variance of the outcomes Y_i in the treatment group ✔

- [ ] The average outcomes Y_i in the treatment group
- [ ] The standard deviation of the outcomes Y_i in the treatment group

Explanation

The term $\frac{S_t^2}{N_t}$ represents the variance of the outcomes Y_i in the treatment group.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 3

0.0/1.0 point (graded)

When is the covariance term S_{tc}^2 equal to zero?

- [ ] When treatment is randomly assigned ✔

Finger Exercises due Oct 26, 2020 19:30 EDT **Past Due**

Question 1

1 point possible (graded)

Suppose you are interested in estimating T_s, the percentage of university students who receive some type of financial aid. To this goal, you administer a survey to poll some randomly selected sample of students. You do some work in R, then you obtain an estimate T_x^* of the fraction of students receiving financial aid from your data. By default, R returns a $(1-\alpha)$ level confidence interval for your estimand T_s (default is $\alpha = 5\%$). Which of the following is true? (Select all that apply)

- [ ] You are 95% sure of your estimate of the exact, correct value of the true effect.
- [ ] If you compute the interval in repeated samples, then $(1-\alpha)$-percent of the time, the interval will bracket the true fraction T_s. ✔
- [ ] α is the significance level of the test. ✔

Explanation

Since the interval is constructed based on your random sample, your interval is also random. You assume a particular sampling distribution for your estimator (in practice, we often use critical values from a Normal distribution), and based on that and your sample properties you construct a confidence interval for your estimand. But think about it: your sample is random, so by virtue of that, your interval is as well. That is, in repeated samples, say your significance level is 5%, then your confidence interval will bracket the true fraction 95% of the time. If you increase your level, your interval becomes narrower, so chances of the real fraction lying in that interval will be smaller.

Question 2

0.0/1.0 point (graded)

Say a scholarship is randomly assigned to a subset of students (treated) and not to others (control). The two groups are ex-ante identical and the sample size is 10,000. Assume this sample size is large enough such that a normal approximation is appropriate. The average score of the treated group is 10 and the average score of the control group is 9.2. The variance of difference between the treated and control is 0.25 and the 95% critical value for normal approximation is 1.96

Compute the 95% confidence interval of the effect of the scholarship on student performance.

What is the upper bound?

Answer: 0.8+(0.5*1.96)

What is the lower bound?

Answer: 0.8-(0.5*1.96)

Show answer

Submit You have used 0 of 2 attempts

What does the hypothesis test below test?

$$H_0 : \frac{1}{N} \sum_{i=1}^{N} Y(1) - Y(0) = 0$$

$$H_a : \frac{1}{N} \sum_{i=1}^{N} Y(1) - Y(0) \neq 0$$

- The null hypothesis that there is no treatment effect for any individual i versus the alternate hypothesis that the treatment effect is nonzero
- The null hypothesis that the treatment effect on any individual is zero versus the alternate hypothesis that the treatment effect on any individual is nonzero
- The null hypothesis that the average treatment effect is zero versus the alternate hypothesis that the average treatment effect is nonzero ✔
- The null hypothesis that the average treatment effect is zero versus the alternate hypothesis that the average treatment effect is positive

Explanation

The test shown above is the Neyman hypothesis, which tests the null hypothesis that the average treatment effect is zero against the alternate hypothesis that the average treatment effect is nonzero.

In the Neyman hypothesis test described in the previous question, the test statistic, t, is calculated as shown below. The numerator refers to _______________ and the denominator refers to _____________________ is for a specific outcome. What does the term in the denominator refer to?

$$t = \frac{\bar{Y_t^{obs}} - \bar{Y_c^{obs}}}{\sqrt{\hat{V_{Neyman}}}}$$

What does the numerator refer to?

- ◯ The estimated difference in outcomes in pairs of units in the treatment and control groups, averaged
- ◯ The difference in sample means between the treatment and control groups ✔
- ◯ The difference in outcomes in pairs of units in the treatment and control groups, averaged
- ◯ The actual population difference in means between the treatment and control groups

What does the denominator refer to?

- ◯ The square root of the actual variance within the treatment group

- ( ) The square root of the actual population variance
- ( ) The square root of the actual variance within the control group
- ( ) The square root of the estimated variance of the difference between treated and control groups ✔

Explanation

The test statistic for a Neyman hypothesis test can be calculated as the difference in sample means between the treatment and control groups divided by the square root of the estimated variance.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 3

0.0/1.0 point (graded)

Suppose you are given a table that contains the following information from a randomized experiment. The table contains the following: the mean and standard deviation for the treatment group, the mean and standard deviation for the control group, the difference in means between both groups, and the standard error of the difference. Suppose you don't have a calculator or any other resources, and you want to test the hypothesis that the average treatment effect is 0 at the 5% significance level. Which of the following can you use as a "rule of thumb for large N"?

- Reject the null if mean(control group)/sd(control group) is greater than 1.96 .
- Reject the null if mean(difference)/se(difference) is less than 1.96 .
- Reject the null if mean(treatment group)/sd(treatment group) is less than 1.96 .
- Reject the null if mean(difference)/se(difference) is greater than1.96 . ✔
- Reject the null if mean(treatment group)/sd(treatment group) is greater than 1.96 .

Explanation

An estimate of the average treatment effect is given by the difference in means between the treatment and control group. The natural t-statistic for this estimator ($\hat{\tau}$) is simply the difference in means divided by the estimated variance; and follows a t-distribution with N-1 degrees of freedom, which converges to a normal distribution for large N. So you can use the general rule of thumb by comparing your t-statistic with the t-critical value for the 5% significance level for a standard normal distribution, which as you saw is 1.96.

Show answer

The next two questions refer to the Oregon Health Lottery Experiment discussed in the lecture, the researchers were interested in estimating the effect of OHP on health. Below is the table you saw in class:

TABLE 1.6
OHP effects on health indicators and financial health

Outcome	Oregon: Control mean (1)	Oregon: Treatment effect (2)	Portland area: Control mean (3)	Portland area: Treatment effect (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 errors are reported in parentheses.

Question 1

0.0/1.0 point (graded)

What is the mean systolic blood pressure for the control group in the Portland area?

Answer: 119

Explanation
The control group mean is shown in column 3.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 2

1 point possible (graded)

What is the mean systolic blood pressure for the treatment group in the Portland area? (round your answer to two decimal places)

Answer: 118.87

Explanation
You are given the mean for the control group, and the difference in means, so you can recover the mean from the treatment group by adding the estimated treatment effect to the control group mean. i.e 119-0.13 = 118.87 .

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 3

0.0/1.0 point (graded)

True or False? We can reject, at the 5% level that there is no effect of OHP on systolic blood pressure.

- True
- False ✔

Question 1

0.0/1.0 point (graded)

What is the t-statistic for the estimated treatment effect of OHP on the probability that a person spends 30% of their income or more on medical expenditures?

Answer: -2.2

Explanation

As mentioned in class, you can construct the t-statistic by taking the ratio of the difference in means to the standard error $t=-0.011/0.005$.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 2

0.0/1.0 point (graded)

What is the t-statistic for the estimated treatment effect of OHP on the probability that a person has medical debt?

Answer: -3.2

Explanation

As mentioned in class, you can construct the t-statistic by taking the ratio of the difference in means to the standard error $t=-0.032/0.01=-3.2$.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 3

0.0/1.0 point (graded)

Recall the critical value for the statistics is 1.96 at the 5% significance level and 1.64 at the 10% level. This comes from the standard normal distribution (remember, the t-distribution converges to a normal distribution for a large N).

From the above, which of the following statements can you conclude to be true? (Select all that apply)

- [ ] We can reject the null that OHP has no effect on the probability that a person spends more than 30% of their income on medical expenses, at the 5% level. ✔
- [ ] We cannot reject the null that OHP has no effect on the probability that a person spends more than 30% of their income on medical expenses, at the 5% level.
- [ ] We can reject the null that OHP has no effect on the probability that a person has medical debt, at the 5% level. ✔
- [ ] We can reject the null that OHP has no effect on the probability that a person has medical debt, at the 10% level. ✔

Explanation

Compare the absolute value of the t-statistics to the critical values. Based on your answers to the previous questions, 2.2 is greater than 1.96 so we can reject the null. Similarly, 3.2 is greater than 1.96 so we can reject the null at both the 5% and 10% levels.

Show answer

Submit You have used 0 of 2 attempts

You can find the results of the experiment here. Which of the following findings were results of the experiment? (Select all that apply)

- [ ] Medicaid reduced individual spending on medical care. ✔
- [ ] Medicaid has no impact on employment and earnings. ✔
- [ ] Medicaid reduced the number of emergency department visits.
- [ ] Medicaid decreased doctor visits.
- [ ] Medicaid increased the probability of receiving high quality care. ✔
- [ ] Medicaid increased the use of preventative care. ✔
- [ ] People were more likely to get the care they needed. ✔

Explanation

The results of the experiment found that Medicaid increased the number of emergency room visits and increased the number of doctor visits.

Finger Exercises due Oct 26, 2020 19:30 EDT Past Due

Question 1

0.0/1.0 point (graded)

If we are only interested in testing hypotheses about the sample for which we have data, which of the following forms of uncertainty do we need to account for?

- ◯ Uncertainty stemming from the fact that our sample is drawn from the population
- ◯ Uncertainty stemming from the fact that you only observe one possible permutation of the treatment assignment ✔
- ◯ Both

Explanation

When testing hypotheses only about the sample we do not need to account for sampling uncertainty, since the hypothesis will only speak to the sample and is not generalizable beyond it. However, we do need to account for uncertainty due to the fact that only one possible permutation of all possible ones is actually observed.

Show answer

Submit You have used 0 of 1 attempt

Question 2

0.0/1.0 point (graded)

True or False? Whether you are interested in drawing conclusions about a given sample, or the underlying population from which it is taken , you need to account for the uncertainty stemming from the fact that you only observe one possible permutation of the treatment assignment.

- True ✔
- False

Explanation

This is true, and explained in detail in lecture. We recommend you revisit the lecture to understand why this is the case if you have not.

Show answer

Submit You have used 0 of 1 attempt

Answers are displayed within the problem

Fisher's Exact Test - Quiz

Bookmark this page

Finger Exercises due Oct 26, 2020 19:30 EDT Past Due

Question 1

0.0/1.0 point (graded)

Fisher was interested in testing the sharp null hypothesis. The sharp null refers to the hypothesis that ___________.

- ◯ The treatment has a non-zero effect on every person
- ◯ The treatment has no effect on any person ✔
- ◯ The treatment has no effect, on average, across all people
- ◯ The treatment has no effect on the majority of people

Explanation

The Fisher exact test aims to test the sharp null hypothesis, which is that the treatment has no effect on any person in the sample, with the alternate hypothesis being that the treatment has an effect on at least one person in the sample.

The Honey Cough Study: Counterfactual Distributions - Quiz

Bookmark this page

Finger Exercises due Oct 26, 2020 19:30 EDT Past Due

Question 1

0.0/1.0 point (graded)

In this segment, Professor Duflo describes the process for conducting the Fisher exact test. In this test, all possible Select an option **Answer:** treatment assignments are tested while keeping the Select an option **Answer:** outcomes fixed. This is called Select an option **Answer:** a permutation test

Explanation

In the Fisher exact test, (simulated) treatment effects are calculated under all possible treatment assignments, keeping the number of treated and control units the same as in the original assignment and holding the outcomes themselves constant. For example, we first assume that the first half of the units are treated and the second are control, and we compute the treatment effect under this assignment.

Show answer

Submit You have used 0 of 2 attempts

Shown below, is another table similar to that Prof.Duflo showed in class, but in this case, our sample contains 4 units. As in lecture, each row represents a different permutation of the treatment assignment amongst the 4 units, and the table contains all possible permutations. The “levels” column contains the difference in outcomes between treated and untreated units, assuming the given permutation. The true observed permutation is bolded and written in blue.

W1	W2	W3	W4	Levels
1	**1**	**0**	**0**	**1.00**
1	0	1	0	0.00
1	0	0	1	2.00
0	1	1	0	-2.00
0	1	0	1	0.00
0	0	1	1	-1.00

Question 1

0.0/1.0 point (graded)

Using the above table, calculate Fisher's p-value by conducting a two-sided test.

Note: Please review our guidelines on precision regarding rounding answers here.

Note: Please review our guidelines on precision regarding rounding answers here.

Answer: 2/3

Explanation

The p-value for Fisher's exact test is given by: the number of permutations for which the absolute value of the difference in outcomes between the treatment and control group that are greater than or equal to the observed effect size given your observed assignment.

Show answer

Submit You have used 0 of 2 attempts

Answers are displayed within the problem

Question 2

0.0/1.0 point (graded)

After calculating the treatment effect under all possible treatment assignments, the next step is to:

- ◯ Count the number of cases when the treatment effect is at least as large as the observed case
- ◯ Count the number of cases in which the absolute value of the simulated treatment effect is less than the treatment effect actually observed, divide by the total number of cases, this is the relevant p-value
- ◯ Average the p-values from each of the possible treatment effects
- ◯ Count the number of cases in which the absolute value of the simulated treatment effect is greater than or equal to than the treatment effect actually observed, divide by the total number of cases, this is the relevant p-value ✔

Explanation

The next step in the Fisher exact test would be to count up the number of cases in which the absolute value of the calculated treatment effect is greater than or equal to the treatment effect actually observed, and divide by the total number of cases to get the relevant p-value: under the sharp null hypothesis (the treatment effect is zero for everyone), the observed treatment effect should be somewhere in the middle of the distribution of simulated differences. The null will be rejected if the observed treatment effect is very much to the right (or the left) of the distribution.

Show answer

Submit You have used 0 of 2 attempts

Choosing the Number of Simulations - Quiz

Bookmark this page

Finger Exercises due Oct 26, 2020 19:30 EDT *Past Due*

Question 1

0.0/1.0 point (graded)

True or False? To conduct a Fisher exact test or randomization inference test, you always need to compute the p-values over all possible permutations of the treatment assignment.

- True
- False ✔

Explanation

As discussed in lecture, if your dataset is very large computing all possible permutations can be computationally costly, and p-values stabilize after a certain sample size. Therefore, you may choose to draw a fixed large number of permutations, because after a certain number of permutations the p-value becomes very precise.

Show answer