

QUANTITATIVE POLICY EVALUATION RANDOMIZED EXPERIMENTS (II).

Eduardo Fé

Fisher's Randomization Inference.

We are interested in the effect of a treatment T on an outcome Y .

1. SUTVA is assumed throughout, so that $Y_i(\mathbf{T}) = Y_i(T_i)$
2. We focus on completely randomized experiments.
3. Unlike in Neyman's superpopulation approach, in Fisher's framework there is an finite population of units.
4. As a result, the distribution of our tests and statistics cannot be obtained on the basis of asymptotic approximations (central limit theorems), as N does not converge to ∞ .
5. Instead, we use the fact that, in a randomized experiment we know the structure of the assignment mechanism in order to obtain an 'exact' distribution of any test.

Fisher's Randomization Inference.

Before proceeding further, a reminder.

In any Randomized Experiment, we know the structure of the assignment mechanism.

Specifically, in a Completely randomized experiment, we allocate N_t units to $T_i = 1$ and $N - N_t$ units to $T_i = 0$.

It turns out, that there are $\binom{N}{N_t}$ of assigning treatments to N units, $\mathbf{T} = (T_1, T_2, \dots, T_N)'$.

Therefore, the probability of any assignment with exactly N_t units allocated to the active treatment ($T_i = 1$) is $Pr(\mathbf{T}) = \binom{N}{N_t}^{-1}$.

Fisher's Randomization Inference.

We are interested in the effect of a treatment T on an outcome Y .

In Fisher's framework, an ineffective treatment has no effect on any unit. From this perspective, we can define a **sharp** null hypothesis,

$$H_0 : Y_i(1) = Y_i(0), \text{ for } i = 1, \dots, N \quad (1)$$

- ▶ Under¹ H_0 , the observed Y_i is such that $Y_i = Y_i(1) = Y_i(0)$ for all i .
- ▶ Because the population is finite (and equal to the sample), we cannot use standard asymptotic arguments to obtain the distribution of estimators and test
- ▶ However, randomization of treatment allows us to obtain the 'exact' distribution of any test or any statistic under the **sharp null hypothesis**

¹That is, assuming that H_0 is true.

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

$T = 1$ if unit (child) given buckwheat honey; $T = 0$ if child does not receive the honey; Y is cough severity (a scale from 0 -least severe- to 6 -most severe). Let's focus on just $N = 6$ of the 105 children.

$N_t = N_c = 3$ allocated via a Completely Randomized Experiment.

	Cough severity		Observed variables	
Unit	$Y(1)$	$Y(0)$	Y_i	T_i
1	3	?	3	1
2	5	?	5	1
3	0	?	0	1
4	?	4	4	0
5	?	0	0	0
6	?	1	1	0

... for each unit, one potential outcome is missing...

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

$T = 1$ if unit (child) given buckwheat honey; $T = 0$ if child does not receive the honey; Y is cough severity (a scale from 0 -least severe- to 6 -most severe). Let's focus on just $N = 6$ of the 105 children.

$N_t = N_c = 3$ allocated via a Completely Randomized Experiment.

	Cough severity		Observed variables	
Unit	$Y(1)$	$Y(0)$	Y_i	T_i
1	3	?	3	1
2	5	?	5	1
3	0	?	0	1
4	?	4	4	0
5	?	0	0	0
6	?	1	1	0

the selection problem again!

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

$T = 1$ if unit (child) given buckwheat honey; $T = 0$ if child does not receive the honey; Y is cough severity (a scale from 0 -least severe- to 6 -most severe). Let's focus on just $N = 6$ of the 105 children.

$N_t = N_c = 3$ allocated via a Completely Randomized Experiment.

	Cough severity		Observed variables	
Unit	$Y(1)$	$Y(0)$	Y_i	T_i
1	3	?	3	1
2	5	?	5	1
3	0	?	0	1
4	?	4	4	0
5	?	0	0	0
6	?	1	1	0

But, what if we force the **sharp null hypothesis**?

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

$T = 1$ if unit (child) given buckwheat honey; $T = 0$ if child does not receive the honey; Y is cough severity (a scale from 0 -least severe- to 6 -most severe). Let's focus on just $N = 6$ of the 105 children.

$N_t = N_c = 3$ allocated via a Completely Randomized Experiment.

	Cough severity		Observed variables	
Unit	$Y(1)$	$Y(0)$	Y_i	T_i
1	3	?	3	1
2	5	?	5	1
3	0	?	0	1
4	?	4	4	0
5	?	0	0	0
6	?	1	1	0

Then $H_0 : Y_i(1) = Y_i(0)$, for $i = 1, \dots, N$

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

$T = 1$ if unit (child) given buckwheat honey; $T = 0$ if child does not receive the honey; Y is cough severity (a scale from 0 -least severe- to 6 -most severe). Let's focus on just $N = 6$ of the 105 children.

$N_t = N_c = 3$ allocated via a Completely Randomized Experiment.

	Cough severity		Observed variables	
Unit	$Y(1)$	$Y(0)$	Y_i	T_i
1	3	3	3	1
2	5	5	5	1
3	0	0	0	1
4	4	4	4	0
5	0	0	0	0
6	1	1	1	0

Then $H_0 : Y_i(1) = Y_i(0)$, for $i = 1, \dots, N$

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

Why does this matter?

1. First, we can compute any statistic under the current assignment vector $\mathbf{T} = (1, 1, 1, 0, 0, 0)'$ (this is nothing special). For example, we can compute

$$d(\mathbf{T}, \mathbf{Y}) = |\bar{Y}_1 - \bar{Y}_0| \quad (2)$$

2. Second, and critically, we can compute any statistic under any assignment vector $\mathbf{T} = (T_1, T_2, T_3, T_4, T_5, T_6)'$ **because**, under the sharp null hypothesis, we know both $Y_i(1)$ **and** $Y_i(0)$.

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

Thus, with the original assignment,

	<u>Cough severity</u>		<u>Observed variables</u>	
Unit	$Y(1)$	$Y(0)$	Y_i	T_i
1	3	3	3	1
2	5	5	5	1
3	0	0	0	1
4	4	4	4	0
5	0	0	0	0
6	1	1	1	0

$$d(\mathbf{T}, \mathbf{Y}) = \left| \frac{8}{3} - \frac{5}{3} \right| = 1 \quad (3)$$

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

and with the alternative assignment $\mathbf{T}^* = (0, 1, 0, 0, 1, 1)'$

	Cough severity		Observed variables	
Unit	$Y(1)$	$Y(0)$	Y_i	T_i^*
1	3	3	3	0
2	5	5	5	1
3	0	0	0	0
4	4	4	4	0
5	0	0	0	1
6	1	1	1	1

$$d(\mathbf{T}, \mathbf{Y}) = \left| \frac{6}{3} - \frac{7}{3} \right| = 1/3 \quad (3)$$

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

Why does this matter?

1. First, we can compute any statistic under the current assignment vector $\mathbf{T} = (1, 1, 1, 0, 0, 0)'$ (this is nothing special). For example, we can compute

$$d(\mathbf{T}, \mathbf{Y}) = |\bar{Y}_1 - \bar{Y}_0| \quad (4)$$

2. Second, and critically, we can compute any statistic under any assignment vector $\mathbf{T} = (T_1, T_2, T_3, T_4, T_5, T_6)'$ **because**, under the sharp null hypothesis, we know both $Y_i(1)$ **and** $Y_i(0)$.
3. Therefore we can compute the distribution of any statistics (such as $d(\mathbf{T}, \mathbf{Y})$) for 'all' possible assignments of units to treatment/control...

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

\mathbf{T}	\bar{Y}_1	\bar{Y}_0	$d(\mathbf{T}, \mathbf{Y})$
(0, 0, 0, 1, 1, 1)	1.667	2.667	1
(0, 0, 1, 0, 1, 1)	.334	4	3.667
(0, 0, 1, 1, 0, 1)	1.667	2.667	1
(0, 0, 1, 1, 1, 0)	1.334	3	1.667
(0, 1, 0, 0, 1, 1)	2	2.334	.334
(0, 1, 0, 1, 0, 1)	3.334	1	2.334
(0, 1, 0, 1, 1, 0)	3	1.334	1.667
(0, 1, 1, 0, 0, 1)	2	2.334	.334
(0, 1, 1, 0, 1, 0)	1.667	2.667	1
(0, 1, 1, 1, 0, 0)	3	1.334	1.667
(1, 0, 0, 0, 1, 1)	1.334	3	1.667
(1, 0, 0, 1, 0, 1)	2.667	1.667	1
(1, 0, 0, 1, 1, 0)	2.334	2	0.334
(1, 0, 1, 0, 0, 1)	1.334	3	1.667
(1, 0, 1, 0, 1, 0)	1	3.334	2.334
(1, 0, 1, 1, 0, 0)	2.334	2	.334
(1, 1, 0, 0, 0, 1)	3	1.334	1.667
(1, 1, 0, 0, 1, 0)	2.667	1.667	1
(1, 1, 0, 1, 0, 0)	4	.334	3.667
(1, 1, 1, 0, 0, 0)	2.667	1.667	1

In this completely randomized experiment, with $N = 6$ and $N_t = N_c = 3$ there are $\binom{6}{3} = 20$ possible assignments including the original assignment $\mathbf{T} = (1, 1, 1, 0, 0, 0)'$ with $d(\mathbf{T}, \mathbf{Y}) = 1$

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

\mathbf{T}	\bar{Y}_1	\bar{Y}_0	$d(\mathbf{T}, \mathbf{Y})$
(0, 0, 0, 1, 1, 1)	1.667	2.667	1
(0, 0, 1, 0, 1, 1)	.334	4	3.667
(0, 0, 1, 1, 0, 1)	1.667	2.667	1
(0, 0, 1, 1, 1, 0)	1.334	3	1.667
(0, 1, 0, 0, 1, 1)	2	2.334	.334
(0, 1, 0, 1, 0, 1)	3.334	1	2.334
(0, 1, 0, 1, 1, 0)	3	1.334	1.667
(0, 1, 1, 0, 0, 1)	2	2.334	.334
(0, 1, 1, 0, 1, 0)	1.667	2.667	1
(0, 1, 1, 1, 0, 0)	3	1.334	1.667
(1, 0, 0, 0, 1, 1)	1.334	3	1.667
(1, 0, 0, 1, 0, 1)	2.667	1.667	1
(1, 0, 0, 1, 1, 0)	2.334	2	0.334
(1, 0, 1, 0, 0, 1)	1.334	3	1.667
(1, 0, 1, 0, 1, 0)	1	3.334	2.334
(1, 0, 1, 1, 0, 0)	2.334	2	.334
(1, 1, 0, 0, 0, 1)	3	1.334	1.667
(1, 1, 0, 0, 1, 0)	2.667	1.667	1
(1, 1, 0, 1, 0, 0)	4	.334	3.667
(1, 1, 1, 0, 0, 0)	2.667	1.667	1

The table presents the **Randomization Distribution** of the statistic $d(\mathbf{T}, \mathbf{Y})$

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

T	\bar{Y}_1	\bar{Y}_0	$d(\mathbf{T}, \mathbf{Y})$
(0, 0, 0, 1, 1, 1)	1.667	2.667	1
(0, 0, 1, 0, 1, 1)	.334	4	3.667
(0, 0, 1, 1, 0, 1)	1.667	2.667	1
(0, 0, 1, 1, 1, 0)	1.334	3	1.667
(0, 1, 0, 0, 1, 1)	2	2.334	.334
(0, 1, 0, 1, 0, 1)	3.334	1	2.334
(0, 1, 0, 1, 1, 0)	3	1.334	1.667
(0, 1, 1, 0, 0, 1)	2	2.334	.334
(0, 1, 1, 0, 1, 0)	1.667	2.667	1
(0, 1, 1, 1, 0, 0)	3	1.334	1.667
(1, 0, 0, 0, 1, 1)	1.334	3	1.667
(1, 0, 0, 1, 0, 1)	2.667	1.667	1
(1, 0, 0, 1, 1, 0)	2.334	2	0.334
(1, 0, 1, 0, 0, 1)	1.334	3	1.667
(1, 0, 1, 0, 1, 0)	1	3.334	2.334
(1, 0, 1, 1, 0, 0)	2.334	2	.334
(1, 1, 0, 0, 0, 1)	3	1.334	1.667
(1, 1, 0, 0, 1, 0)	2.667	1.667	1
(1, 1, 0, 1, 0, 0)	4	.334	3.667
(1, 1, 1, 0, 0, 0)	2.667	1.667	1

Because this is a completely randomized experiment, each assignment is equally likely (with probability $1/20$)

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

\mathbf{T}	\bar{Y}_1	\bar{Y}_0	$d(\mathbf{T}, \mathbf{Y})$
(0, 0, 0, 1, 1, 1)	1.667	2.667	1
(0, 0, 1, 0, 1, 1)	.334	4	3.667
(0, 0, 1, 1, 0, 1)	1.667	2.667	1
(0, 0, 1, 1, 1, 0)	1.334	3	1.667
(0, 1, 0, 0, 1, 1)	2	2.334	.334
(0, 1, 0, 1, 0, 1)	3.334	1	2.334
(0, 1, 0, 1, 1, 0)	3	1.334	1.667
(0, 1, 1, 0, 0, 1)	2	2.334	.334
(0, 1, 1, 0, 1, 0)	1.667	2.667	1
(0, 1, 1, 1, 0, 0)	3	1.334	1.667
(1, 0, 0, 0, 1, 1)	1.334	3	1.667
(1, 0, 0, 1, 0, 1)	2.667	1.667	1
(1, 0, 0, 1, 1, 0)	2.334	2	0.334
(1, 0, 1, 0, 0, 1)	1.334	3	1.667
(1, 0, 1, 0, 1, 0)	1	3.334	2.334
(1, 0, 1, 1, 0, 0)	2.334	2	.334
(1, 1, 0, 0, 0, 1)	3	1.334	1.667
(1, 1, 0, 0, 1, 0)	2.667	1.667	1
(1, 1, 0, 1, 0, 0)	4	.334	3.667
(1, 1, 1, 0, 0, 0)	2.667	1.667	1

So, what is the probability of obtaining a value as extreme as the original $d(\mathbf{T}, \mathbf{Y}) = 1$?

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

T	\bar{Y}_1	\bar{Y}_0	$d(\mathbf{T}, \mathbf{Y})$
(0, 0, 0, 1, 1, 1)	1.667	2.667	1
(0, 0, 1, 0, 1, 1)	.334	4	3.667
(0, 0, 1, 1, 0, 1)	1.667	2.667	1
(0, 0, 1, 1, 1, 0)	1.334	3	1.667
(0, 1, 0, 0, 1, 1)	2	2.334	.334
(0, 1, 0, 1, 0, 1)	3.334	1	2.334
(0, 1, 0, 1, 1, 0)	3	1.334	1.667
(0, 1, 1, 0, 0, 1)	2	2.334	.334
(0, 1, 1, 0, 1, 0)	1.667	2.667	1
(0, 1, 1, 1, 0, 0)	3	1.334	1.667
(1, 0, 0, 0, 1, 1)	1.334	3	1.667
(1, 0, 0, 1, 0, 1)	2.667	1.667	1
(1, 0, 0, 1, 1, 0)	2.334	2	0.334
(1, 0, 1, 0, 0, 1)	1.334	3	1.667
(1, 0, 1, 0, 1, 0)	1	3.334	2.334
(1, 0, 1, 1, 0, 0)	2.334	2	.334
(1, 1, 0, 0, 0, 1)	3	1.334	1.667
(1, 1, 0, 0, 1, 0)	2.667	1.667	1
(1, 1, 0, 1, 0, 0)	4	.334	3.667
(1, 1, 1, 0, 0, 0)	2.667	1.667	1

So, what is the probability of obtaining a value as extreme as the original $d(\mathbf{T}, \mathbf{Y}) = 1$? $16/20 = 0.8$ -the sharp null hypothesis is very likely!

Fisher's Randomization Inference: the Honey Study

Paul et al. (2007).

Why does this matter?

1. First, we can compute any statistic under the current assignment vector $\mathbf{T} = (1, 1, 1, 0, 0, 0)'$ (this is nothing special). For example, we can compute

$$d(\mathbf{T}, \mathbf{Y}) = |\bar{Y}_1 - \bar{Y}_0| \quad (5)$$

2. Second, and critically, we can compute any statistic under any assignment vector $\mathbf{T} = (T_1, T_2, T_3, T_4, T_5, T_6)'$ **because**, under the sharp null hypothesis, we know both $Y_i(1)$ **and** $Y_i(0)$.
3. Therefore we can compute the **randomization distribution** of any statistics (such as $d(\mathbf{T}, \mathbf{Y})$)
4. In other words, we can compute the **p-value** of any test $d(\mathbf{T}, \mathbf{Y})$.
5. Because in Fisher's framework sample = population, this **p-value** is **exact** for the statistics in the population²

²In stark contrast to Neyman's framework, where p-values are normally obtained from asymptotic ($N \rightarrow \infty$) approximations

Fisher's Randomization Inference: Sharp nulls.

In a randomized experiment we know the form of the assignment mechanism. Next, we need to decide on a specific sharp null hypothesis.

- ▶ No effect: $H_0 : Y_i(1) = Y_i(0)$, for $i = 1, \dots, N$
- ▶ Non-zero effect: $H_0 : Y_i(1) = Y_i(0) + C$, for $i = 1, \dots, N$ and some $C \neq 0$

Fisher's Randomization Inference: Sharp nulls.

In the example from Paul et al. (2007) study, the sharp null could instead be $H_0 : Y_i(1) = Y_i(0) + 2$, for $i = 1, \dots, N$

	<u>Cough severity</u>		<u>Observed variables</u>	
Unit	$Y(1)$	$Y(0)$	Y_i	T_i
1	3	1	3	1
2	5	3	5	1
3	0	-2	0	1
4	4	2	2	0
5	0	-2	-2	0
6	1	-1	-1	0

$$d(\mathbf{T}, \mathbf{Y}) = \left| \frac{8}{3} - \left(-\frac{1}{3} \right) \right| = 3 \quad (6)$$

Fisher's Randomization Inference: Choice of statistic..

In principle, we can use any statistic for the basis of randomization.

- ▶ Difference in means, $d(\mathbf{T}, \mathbf{Y}) = |\bar{Y}_1 - \bar{Y}_0|$
- ▶ Difference in medians, $d(\mathbf{T}, \mathbf{Y}) = |\text{med}_1(Y_i) - \text{med}_0(Y_i)|$
- ▶ Difference in quantiles, $d(\mathbf{T}, \mathbf{Y}) = |q_{\alpha,1}(Y_i) - q_{\alpha,0}(Y_i)|$
- ▶ t-statistics, $d(\mathbf{T}, \mathbf{Y}) = \left| \frac{\bar{Y}_1 - \bar{Y}_0}{\sqrt{s_1^2/N_0 + s_0^2/N_0}} \right|$

Fisher's Randomization Inference: Choice of statistic.

Rank based statistics

The rank of unit i in the sample equals the number of units with observed Y less than or equal to³ Y_i ,

$$R_i = \sum_{j=1}^N \mathbf{1}(Y_j \leq Y_i) \quad (7)$$

Often one works with the *normalised* (zero-mean) rank

$$\tilde{R}_i = \sum_{j=1}^N \mathbf{1}(Y_j \leq Y_i) - \frac{N+1}{2} \quad (8)$$

Statistics based on ranks have the advantages of being robust to outliers.

- ▶ Difference in ranks, $d(\mathbf{T}, \mathbf{Y}) = |\bar{R}_1 - \bar{R}_0|$

³Here $\mathbf{1}(\cdot)$ equals 1 if the argument in the brackets is true (otherwise $\mathbf{1}(\cdot)=0$).

Fisher's Randomization Inference: Choice of statistic.

Kolmogorov-Smirnov statistic

This statistic compares the empirical distributions of Y in the treated and control groups, formally,

$$d(\mathbf{T}, \mathbf{Y}) = \max_{i=1, \dots, N} \left| \hat{F}_1(Y_i) - \hat{F}_0(Y_i) \right| \quad (9)$$

where

$$\hat{F}(y) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(Y_i \leq y) \quad (10)$$

Fisher's Randomization Inference: Randomization Distribution.

For any given statistics and a sample/population, the critical step in Fisher's randomization inference is computing the **randomization distribution** of the statistic in the sample/population.

Obtaining the exact randomization distribution requires computing the value of the statistic for all possible assignments. For toy-examples (such as the Honey study), this can be done by hand. For real sample sizes we need computational methods.

Specifically, even for moderate samples it might be costly to obtain all the possible assignments. For instance, in a completely randomized experiment with $N = 20$ and 10 units allocated to the treatment group, there are 184,756 possible assignments.

Fisher's Randomization Inference: Randomization Distribution.

Specifically, even for moderate samples it might be costly to obtain all the possible assignments. For instance, in a completely randomized experiment with $N = 20$ and 10 units allocated to the treatment group, there are 184,756 possible assignments.

In practice, one often simulates the randomization distribution by drawing a number $B < \binom{N}{N_t}$ of assignments and computing the statistic under consideration $d(\mathbf{T}, \mathbf{Y})$ for each assignment, $d^{(b)}(\mathbf{T}, \mathbf{Y})$ for $b = 1 \dots, B$. From this collection of estimates, we compute the **approximate/simulated** exact randomization distribution as the 'histogram' of the $d^{(b)}(\mathbf{T}, \mathbf{Y})$. The associated p-value for the sharp null hypothesis corresponds to the proportion of $d^{(b)}(\mathbf{T}, \mathbf{Y})$ greater or equal than the observed statistic.

Fisher's Randomization Inference: Additional Topics.

Fisher's Randomization Inference with Covariates

Extending the analysis when k **pre-test** variables, $\mathbf{x}_i = (X_{1,i}, \dots, X_{K,i})'$ is straightforward, particularly if we believe that \mathbf{x}_i might predict the potential outcomes $Y_i(t)$.

Specifically, we could estimate a regression model

$$Y_i = Y_i(0) + \tau_{FS} \cdot T_i + \mathbf{x}_i' \beta + \varepsilon_i \quad (11)$$

via Ordinary Least Squares and use the estimated $\hat{\tau}_{FS}$ as the basis for randomization inference,

$$d(\mathbf{T}, \mathbf{Y}) = \hat{\tau}_{FS} \quad (12)$$

Fisher's Randomization Inference: Additional Topics.

Point estimation in Fisher's Randomization Inference.

The starting point in Fisher's approach is testing the sharp null hypothesis. In contrast to Neyman's framework, estimation of the treatment effect comes as a by-product of hypothesis testing.

Estimation in Fisher's setting is achieved by *inverting* the statistic whose randomization distribution we have obtained. This yields what is known as a Hodges-Lehmann estimator (Hodges and Lehmann, 1963).

What this entails is to

1. First, propose a model that details the relationship between Y_i and the treatment effect τ_{FS} . The standard model is a linear specification with $Y_i(1) = Y_i(0) + \tau_{FS}$.
2. Second, the Hodges-Lehmann point estimate of τ_{FS} is the value $\tilde{\tau}_{FS}$ for which the test of the sharp null hypothesis $H_0 : Y_i(1) = Y_i(0) + \tilde{\tau}_{FS}$ yields the largest p-value.

Concluding remarks.

1. Fisher's randomization inference limits the scope of the analysis: The sample = the population of interest and the results of the analysis should not be extrapolated to other populations. This is a strength, not a disadvantage: it obliges the researcher to moderate their claims.
2. At the core of Fisher's method is a test of a sharp null hypothesis. Sharp null hypothesis are strong statements. This is a disadvantage.
3. However, randomization inference enables testing hypothesis on the back of exact p-values. The marvel of this cannot be overstated.
4. Randomization inference allows great flexibility in the choice of statistic; choices of statistics are thus often based on anticipated properties of the statistic (such as robustness to outliers).
5. Even more interestingly, testing several hypothesis at once (multiple test) is relatively straightforward.

Concluding remarks.

6. Randomization inference relies on a minute set of assumptions. The most restrictive is unconfoundedness, but (a) this is a natural assumption in randomized experiments **and** (b) **will hold -in ways to be made explicit- in some observational studies (quasi-experimental settings).**
7. The virtually assumption free advantage of randomization inference is lost when we consider estimation: we need a specific model relating the treatment effect to the potential outcomes. However, this is in no way different to any other statistical modelling technique, so it would be far-fetched to see this as a major weakness of randomization inference.

References I

- Hodges, J.** and **Lehmann, E.** (1963). Estimates of location based on ranks. *Annals of Mathematical Statistics*, 34: 598–611
- Paul, I.M., Beiler, J., McMonagle, A., Shaffer, M.L., Duda, L., and Berlin, C.M.** (2007). Effect of honey, dextromethorphan, and no treatment on nocturnal cough and sleep quality for coughing children and their parents. *Archives of pediatrics & adolescent medicine*, 161: 1140–1146. ISSN 1538-3628. doi:10.1001/archpedi.161.12.1140