Randomization Inference

Charlie Gibbons
Political Science 236

October 1, 2008

Outline

These notes are based upon Rosenbaum (2002).

Experiments

Fischer makes four important points about experiments:

1. Experiments do not require units to be homogeneous
2. Experiments do not require units to be a random sample from a population of units
3. Treatment only needs to be allocated randomly among experimental units
4. Probability ("chance") enters the analysis only through random treatment assignment, which the researcher controls

Experiments

Fischer makes four important points about experiments:

1. Experiments do not require units to be homogeneous
2. Experiments do not require units to be a random sample from a population of units
3. Treatment only needs to be allocated randomly among experimental units
4. Probability ("chance") enters the analysis only through random treatment assignment, which the researcher controls

Experiments

Fischer makes four important points about experiments:

1. Experiments do not require units to be homogeneous
2. Experiments do not require units to be a random sample from a population of units
3. Treatment only needs to be allocated randomly among experimental units
4. Probability ("chance") enters the analysis only through random treatment assignment, which the researcher controls

Experiments

Fischer makes four important points about experiments:

1. Experiments do not require units to be homogeneous
2. Experiments do not require units to be a random sample from a population of units
3. Treatment only needs to be allocated randomly among experimental units
4. Probability (“chance”) enters the analysis only through random treatment assignment, which the researcher controls

Experiments

Fischer makes four important points about experiments:

1. Experiments do not require units to be homogeneous
2. Experiments do not require units to be a random sample from a population of units
3. Treatment only needs to be allocated randomly among experimental units
4. Probability ("chance") enters the analysis only through random treatment assignment, which the researcher controls

Experiments

Fischer and Neymann had different ideas about what experiments achieved.

Experiments

Fischer and Neymann had different ideas about what experiments achieved.

Fischer believed that the experiment gave an accurate inference for those units that had to be treated (consistent with his preceding points), but not necessarily a sound inference for the entire population. Hence experiments give *internal* validity, but not *external* validity.

Experiments

Fischer and Neymann had different ideas about what experiments achieved.

Fischer believed that the experiment gave an accurate inference for those units that had to be treated (consistent with his preceding points), but not necessarily a sound inference for the entire population. Hence experiments give *internal* validity, but not *external* validity.

Neymann believed that units can be chosen randomly from the population to generate external validity.

The framework

Let there be N units divided into S strata prior to treatment, giving n_s units in strata s. Define the following

$$\begin{aligned}
Z_{si} &= \mathbb{I}\{\text{unit } i \text{ in strata } s \text{ received treatment}\} \\
Z_s &= (Z_{s1}, \ldots Z_{sn_s})' \\
Z &= (Z_1, \ldots, Z_S)' \\
m_s &= \sum_{i=1}^{n_s} Z_{si}
\end{aligned}$$

Note that Z is a random variable with a distribution determined by the researcher. The only limitation on this distribution is that $0 < \Pr(Z_{si} = 1) < 1 \; \forall \; s, i$. Let Ω be the set of all possible realizations of Z.

The framework

Examples:

- Fixed margins

$$\begin{aligned} m_s &= k_s \\ K &\equiv \#\{\Omega\} = \prod_{s=1}^{S} \binom{n_s}{k_s} \end{aligned}$$

- Binomial randomization

$$\begin{aligned} \Pr(Z_{si} = 1) &= \frac{1}{2} \\ K &\equiv \#\{\Omega\} = 2^N \end{aligned}$$

The framework

Examples:

- Fixed margins

$$\begin{aligned} m_s &= k_s \\ K &\equiv \#\{\Omega\} = \prod_{s=1}^{S} \binom{n_s}{k_s} \end{aligned}$$

- Binomial randomization

$$\begin{aligned} \Pr(Z_{si} = 1) &= \frac{1}{2} \\ K &\equiv \#\{\Omega\} = 2^N \end{aligned}$$

The framework

Examples:

- Fixed margins

$$
\begin{aligned}
m_s &= k_s \\
K &\equiv \#\{\Omega\} = \prod_{s=1}^{S} \binom{n_s}{k_s}
\end{aligned}
$$

- Binomial randomization

$$
\begin{aligned}
\Pr(Z_{si} = 1) &= \frac{1}{2} \\
K &\equiv \#\{\Omega\} = 2^N
\end{aligned}
$$

Testing for no effect

The simplest case of randomization inference is to test for the absence of an effect. That is, the null hypothesis is that there is no treatment effect. Why?

Testing for no effect

The simplest case of randomization inference is to test for the absence of an effect. That is, the null hypothesis is that there is no treatment effect. Why?

This sort of test can be performed with very limited assumptions and may provide a good starting point before generating point estimates.

Testing for no effect

Let r_{si} be the response of unit i in strata s and r be a vector of all responses corresponding to Z.

Testing for no effect

Let r_{si} be the response of unit i in strata s and r be a vector of all responses corresponding to Z.

What is random *under the null hypothesis*?

Testing for no effect

Let r_{si} be the response of unit i in strata s and r be a vector of all responses corresponding to Z.

What is random *under the null hypothesis*?

- The allocation of treatment Z is random and under the control of the researcher
- The response vector r *is not random*—under the null hypothesis, there is no response to treatment and thus *responses would not change if a different set of treatments Z' was applied.*

Testing for no effect

Let r_{si} be the response of unit i in strata s and r be a vector of all responses corresponding to Z.

What is random *under the null hypothesis*?

- The allocation of treatment Z is random and under the control of the researcher
- The response vector r *is not random*—under the null hypothesis, there is no response to treatment and thus *responses would not change if a different set of treatments* Z' *was applied.*

Testing for no effect

Let r_{si} be the response of unit i in strata s and r be a vector of all responses corresponding to Z.

What is random *under the null hypothesis*?

- The allocation of treatment Z is random and under the control of the researcher
- The response vector r *is not random*—under the null hypothesis, there is no response to treatment and thus *responses would not change if a different set of treatments Z' was applied.*

Hence, when doing inference, we will take r as given (*i.e.*, , condition on it) and compare it to other possible realizations of treatment.

Test procedure

We have the following procedure for test statistic $t(z, r)$ for fixed set of responses and treatment assignment z:

1. The null hypothesis is assumed, thereby fixing the value of r.
2. Treatment assignment z is selected at random from Ω.
3. Find the test statistic T of the actual experiment.
4. Calculate the probability of generating a test statistic greater than or equal to the one from the actual experiment using all possible treatment assignments in Ω.

Test procedure

We have the following procedure for test statistic $t(z, r)$ for fixed set of responses and treatment assignment z:

1. The null hypothesis is assumed, thereby fixing the value of r.
2. Treatment assignment z is selected at random from Ω.
3. Find the test statistic T of the actual experiment.
4. Calculate the probability of generating a test statistic greater than or equal to the one from the actual experiment using all possible treatment assignments in Ω.

Test procedure

We have the following procedure for test statistic $t(z, r)$ for fixed set of responses and treatment assignment z:

1. The null hypothesis is assumed, thereby fixing the value of r.
2. Treatment assignment z is selected at random from Ω.
3. Find the test statistic T of the actual experiment.
4. Calculate the probability of generating a test statistic greater than or equal to the one from the actual experiment using all possible treatment assignments in Ω.

Test procedure

We have the following procedure for test statistic $t(z, r)$ for fixed set of responses and treatment assignment z:

1. The null hypothesis is assumed, thereby fixing the value of r.
2. Treatment assignment z is selected at random from Ω.
3. Find the test statistic T of the actual experiment.
4. Calculate the probability of generating a test statistic greater than or equal to the one from the actual experiment using all possible treatment assignments in Ω.

Test procedure

We have the following procedure for test statistic $t(z, r)$ for fixed set of responses and treatment assignment z:

1. The null hypothesis is assumed, thereby fixing the value of r.
2. Treatment assignment z is selected at random from Ω.
3. Find the test statistic T of the actual experiment.
4. Calculate the probability of generating a test statistic greater than or equal to the one from the actual experiment using all possible treatment assignments in Ω.

This yields a significance level of

$$\Pr(t(Z, r)) = \sum_{z \in \Omega} \mathbb{I}\{t(z, r) \geq T\} \dot{\Pr}(Z = z)$$

Test statistics

In The Lady Tasting Tea, the test statistic that we use is the number of cups correctly guessed:

$$
\begin{aligned}
T &= \underbrace{Z'r}_{\text{Treated correctly identified}} + \underbrace{\left(1 - Z'\right)(1 - r)}_{\text{Control correctly identified}} \\
&= 2Z'r
\end{aligned}
$$

Test statistics

In The Lady Tasting Tea, the test statistic that we use is the number of cups correctly guessed:

$$\begin{aligned} T &= \underbrace{Z'r}_{\text{Treated correctly identified}} + \underbrace{\left(1 - Z'\right)(1 - r)}_{\text{Control correctly identified}} \\ &= 2Z'r \end{aligned}$$

When we are calculating the significance of this result against the null, we ask, "if treatment were assigned differently, but his responses do not change, what's the probability that his slate of guesses would have performed at least as well as it did in this case?"

For example, "if we had ordered the cups differently, what's the probability that his guesses would have gotten at least as many cups correct as he did here?"