Attributable Effects in Randomized Experiments and Observational Studies

Paul R. Rosenbaum*

Abstract

The motivation for randomization inference and sensitivity analysis is reviewed. New methods are discussed for inverting randomization tests to provide interval estimates of the magnitude of treatment effects.

1 Notation and Review

1.1 Notation

1.1.1 Strata or Matched Sets S strata defined by pretreatment covariates, $s = 1, \ldots, S$, with n_s subjects in stratum s, and $N = \sum n_s$. Write $Z_{si} = 1$ if the i^{th} subject in stratum s is treated, $Z_{si} = 0$ if control. There are $m_s = \sum_{i=1}^{n_s} Z_{si}$ treated subjects in stratum s. Matched pairs is the special case with $n_s = 2$, $m_s = 1$ for $s = 1, \ldots, S$.

1.1.2 Randomization Ω is the set of the $K = \prod_{s=1}^{S} \binom{n_s}{m_s}$ possible values $\mathbf{z}$ of the $N-$dimensional treatment assignment $\mathbf{Z} = (Z_{11}, Z_{12}, \ldots, Z_{S,n_S})^T$ with $m_s = \sum_{i=1}^{n_s} z_{si}$ for $s = 1, \ldots, S$. Randomization: $\mathbf{Z}$ picked at random from Ω, that is, $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$ for each $\mathbf{z} \in \Omega$. Difficulty in an observational study is that this may not be true, and $\Pr(\mathbf{Z} = \mathbf{z})$ is typically unknown

1.1.3 Treatment Effects Each subject has two potential responses, a response r_{Tsi} that would be seen under treatment and a response r_{Csi} that would be seen under control. (Neyman 1923, Rubin 1974), but observe only one. Vectors $\mathbf{r}_T$, $\mathbf{r}_C$. Fixed features of the finite population of N subjects. *Null hypothesis of no effect*: $H_0 : r_{Tsi} = r_{Csi}$ for all s, i, or $\mathbf{r}_T = \mathbf{r}_C$. *Additive effect*, $r_{Tsi} = r_{Csi} + \tau$ for all s, i. *Nonnegative effect* if $r_{Tsi} \geq r_{Csi}$ for all s, i. Observe $R_{si} = Z_{si} r_{Tsi} + (1 - Z_{si}) r_{Csi}$, which is a random variable.

1.2 Review

1.2.1 Randomization Test of No Effect Under the null hypothesis $H_0 : \mathbf{r}_T = \mathbf{r}_C$, the observed responses, $\mathbf{R} = \mathbf{r}_C$, are fixed and observed, so any test statistic, $T = t(\mathbf{Z}, \mathbf{R})$, is a function $t(\mathbf{Z}, \mathbf{r}_C)$ of fixed, observed constants, $\mathbf{r}_C$, and a random variable $\mathbf{Z}$ with a known distribution, $\Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{K}$ for each $\mathbf{z} \in \Omega$, so T has a known null distribution created by randomization, which forms the "reasoned basis for inference," in Fisher's phrase.

1.2.2 Randomization Inference for an Additive Effect With an additive effect, $\mathbf{r}_T = \mathbf{r}_C + \tau \mathbf{1}$, the observed responses are $\mathbf{R} = \mathbf{r}_C + \tau \mathbf{Z}$. To test $H_0 : \tau = \tau_0$, compute the (observed) adjusted responses, $\mathbf{R} - \tau_0 \mathbf{Z}$, which equal $\mathbf{r}_C$ if H_0 is true, so randomization creates the null distribution of $T = t(\mathbf{Z}, \mathbf{R} - \tau_0 \mathbf{Z}) = t(\mathbf{Z}, \mathbf{r}_C)$. The set of values τ_0 not rejected by a 0.05 level test is the 95% confidence interval, and the value $\widehat{\tau}$ that equates $t(\mathbf{Z}, \mathbf{R} - \widehat{\tau} \mathbf{Z})$ to its null expectation is the Hodges-Lehmann point estimate.

1.2.3 Sensitivity Analysis Z_{si} initially independent, and two subjects i, j in the same stratum s differ in odds of treatment by at most $\Gamma \geq 1$

$$\frac{1}{\Gamma} \leq \frac{\Pr(Z_{si} = 1) \,/\, \Pr(Z_{si} = 0)}{\Pr(Z_{sj} = 1) \,/\, \Pr(Z_{sj} = 0)} \leq \Gamma \text{ for all } s, i;$$

then condition on $m_s = \sum_{i=1}^{n_s} Z_{si}$ to return distribution to Ω. With $\Gamma = e^{\gamma}$, equivalent to

(1.1)

$$\Pr(\mathbf{Z} = \mathbf{z}) = \frac{\exp(\gamma \mathbf{z}^T \mathbf{u})}{\sum_{\mathbf{b} \in \Omega} \exp(\gamma \mathbf{b}^T \mathbf{u})} \text{ for a } \mathbf{u} \text{ with } 0 \leq u_{si} \leq 1, \forall s, i,$$

(Rosenbaum 1995, 2002, §4), where $\Gamma = 1$ yields the randomization distribution. Fix Γ, then for any inference quantity (eg significance level, point estimate), find max and min subject to (1.1), repeat for a range of Γ to display sensitivity.

2 Attributable Effects

2.1 Simplest Case: Fisher' Exact Test for 2×2 Table One stratum, $S = 1$, drop s subscript, binary responses, $r_{Ti} = 1$ or 0, $r_{Ci} = 1$ or 0, dead or alive, with nonnegative effect, $\delta_i = r_{Ti} - r_{Ci}$. Number of

*Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6340 USA, rosenbaum@stat.wharton.upenn.edu. Supported by a grant from the National Science Foundation.

Table 1: 2x2 Table Adjusted for Pivot.

Response	Treated	Control
1	$\sum Z_i\, r_{Ti} - A$	$\sum (1 - Z_i)\, r_{Ci}$
0	$\sum Z_i\, (1 - r_{Ti}) + A$	$\sum (1 - Z_i)\, (1 - r_{Ci})$
Total	m	$n - m$

Table 2: Station Design and Mortality. Source: Coats and Walter (1999).

	No Pit	Pit
Dead	16	14
Alive	5	18

deaths among treated subject is $T = \sum_{i=1}^{n} Z_i\, r_{Ti}$, and, in a randomized experiment, it has the hypergeometric distribution under the null hypothesis of no effect, $H_0 : r_{Ti} = r_{Ci}$, but not when there is an effect. Then $A = \sum_{i=1}^{n} Z_i\, \delta_i = \sum_{i=1}^{n} Z_i\, (r_{Ti} - r_{Ci})$ is number of events among treated subjects caused by treatment — an unobservable random variable. Key idea: A is a discrete pivot: $T - A = \sum_{i=1}^{n} Z_i\, r_{Ti} - \sum_{i=1}^{n} Z_i\, (r_{Ti} - r_{Ci}) = \sum_{i=1}^{n} Z_i\, r_{Ci}$ has a hypergeometric distribution in a randomized experiment. Work with Table 1 with adjusted cells and row margins.

2.2 Attributable Effects on Quantiles: Displacements

2.2.1 Order Statistics Continuous, untied potential responses, (y_{Tsi}, y_{Csi}), under treatment, y_{Tsi}, and control, y_{Csi}, with $y_{Tsi} \geq y_{Csi}$. Two sets of N *fixed but unobserved* order statistics: $y_{T(1)} < \ldots < y_{T(N)}$ and $y_{C(1)} < \ldots < y_{C(N)}$. We *observe the random* $Y_{si} = Z_{si}\, y_{Tsi} + (1 - Z_{si})\, y_{Csi}$ with random order statistics $Y_{(1)} < \ldots < Y_{(N)}$. (Can allow for ties.)

2.2.2 Displacements Fix a k, so $y_{C(k)}$ is the (*unobserved*) k/N quantile of potential responses y_{Csi} to control. Let θ be any (*unknown*) value between $y_{C(k)}$ and $y_{C(k+1)}$, so $y_{C(k)} < \theta < y_{C(k+1)}$. Subject (s, i) has a displacement if $y_{Tsi} > \theta > y_{Csi}$. Write $r_{Tsi} = 1$ if $y_{Tsi} > \theta$, $r_{Tsi} = 0$ otherwise; $r_{Csi} = 1$ if $y_{Csi} > \theta$, $r_{Csi} = 0$ otherwise; so there is a displacement if $\delta_{si} = r_{Tsi} - r_{Csi} = 1$. Then $R_{si} = Z_{si}\, r_{Tsi} + (1 - Z_{si})\, r_{Csi}$ indicates whether $Y_{si} > \theta$. *Not quite the same as binary case, because* $y_{C(k)}$ *and hence* R_{si} *are not observed.* $A = \sum_{s,i} Z_{si}\, \delta_{si}$ is number of displacements attributable to treatment; related, but not quite the same as the control median test of Gart & Gastwirth.

2.2.3 Device That Solves the Problem

Table 3: Kidney Function of Cadmium Workers and Unexposed Controls. Beta-2-microglobulin in micrograms per gram of creatine. From Thun, et al. (1989).

Pair	Cadmium Worker	Hospital Control
1	107,143	311
2	33,679	338
3	18,836	159
4	173	110
5	389	226
6	1,144	305
7	513	222
8	211	242
9	24,288	250
10	67,632	256
11	488	135
12	700	96
13	328	142
14	98	120
15	122	376
16	2,302	173
17	10,208	178
18	892	213
19	2,803	257
20	201	81
21	148	199
22	522	114
23	941	247

LEMMA 2.1. *If* $a = \sum_{s,i} Z_{si}\, \delta_{si}$, *then*

$$Y_{(k+1-a)} > \theta > Y_{(k-a)}.$$

Proof. $N - k$ subjects have $y_{Csi} > \theta$, and since $y_{Tsi} \geq y_{Csi}$, these subjects have $Y_{si} > \theta$. Because $a = \sum_{s,i} Z_{si}\, \delta_{si}$, there are a other subjects with $Y_{si} = y_{Tsi} > \theta > y_{Csi}$. The remaining $k - a$ subjects have $\theta > Y_{si}$. So $N - k + a$ of the $Y_{si} > \theta$ and $k - a$ of the $Y_{si} < \theta$, proving the lemma.

To test a hypothesis, $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, compute $A_0 = \sum_{s,i} Z_{si}\, \delta_{0si}$, determine $Y_{(k+1-A_0)}$ and $Y_{(k-A_0)}$, which determine R_{si}, which transforms the problem into the binary case studied previously.

3 Bibliography

3.1 Papers From Which the Talk is Derived

References

[1] Rosenbaum, P. R. (2001) Effects attributable to treatment: Inference in experiments and observational studies with a discrete pivot. *Biometrika*, **88**, 219-231.

[2] Rosenbaum, P. R. (2002) Attributing effects to treatment in matched observational studies. *JASA*, **97**, 183-192.
[3] Rosenbaum, P. R. (2002) *Observational Studies* (Second Edition). New York: Springer-Verlag, §5.
[4] Rosenbaum, P. R. (2003) Exact confidence intervals for nonconstant effects by inverting the signed rank test. *American Statistician*, **57**, 132-138.

3.2 The 'Really Modern' Idea

References

[1] Fisher, R.A. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.
[2] Neyman, J. (1923), On the application of probability theory to agricultural experiments. Essay on principles. Section 9. (In Polish) Roczniki *Nauk Roiniczych,* Tom X, pp1-51. Reprinted in English in *Statistical Science*, 1990, 5, 463-480, with discussion by T. Speed and D. Rubin. Causal effects as comparisons of potential responses under alternative treatments.
[3] Pitman, E. J. G. (1937) Significance tests which may be applied to samples from any population. *JRSS*, **4**, 119-130.
[4] Welch, B. L. (1937) On the z-test in randomized blocks and latin squares. *Biometrika*, **29**, 21-52.
[5] Kempthorne, O. (1952) *Design and Analysis of Experiments*. New York: John Wiley.
[6] Cornfield, J. and Tukey, J. (1956) Average values of mean squares in factorials. *Annals of Mathematical Statistics*, **27**, 907-949.
[7] Cochran, W. G. and Cox, G. M. (1957) *Experimental Designs*. NY: Wiley.
[8] Cochran, W. G. (1965) The planning of observational studies of human populations (with Discussion). *JRSS,* A,128, 134-155.
[9] Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**, 688-701.
[10] Rubin, D. B. (1977) Randomization on the basis of a covariate. *Journal of Educational Statistics*, **2**, 1-26.
[11] Rubin, D. B. (1984) William G. Cochran's contributions to the design, analysis and evaluation of observational studies. In: *W. G. Cochran's Impact on Statistics,* eds., P. S. R. S. Rao and J. Sedransk, New York: John Wiley, pp. 37-69.
[12] Gail, M. H., Tan, W. Y., and Piantadosi, S. (1988) Tests for no treatment effect in randomized clinical trials. *Biometrika* **75** 57-64.
[13] Raz, J. (1990) Testing for no effect when estimating a smooth function by nonparametric regression: A randomization approach. *JASA* **85** 132-138.
[14] Cox, D. R. and Reid, N. (2000) *The Theory of the Design of Experiments*. New York: CRC Press.

3.3 Other Papers Related to Attributable Effects

References

[1] Cole, P. & MacMahon, B. (1971) Attributable risk percent in case-control studies. *Brit. J. Prevent. Social Med.* **25**, 242-4.
[2] Gart, J.(1963) A median test with sequential application. *Biometrika,* **50,** 55-62.
[3] Gastwirth, J. L. (1968), The first-median test: a two-sided version of the control median test. *JASA, 63,* 692-706.
[4] Hamilton, M. A. (1979) Choosing the parameter for 2×2 and $2 \times 2 \times 2$ table analysis. *American Journal of Epidemiology* 109, 362-75.
[5] Li, G., Tiwari, R. C., & Wells, M. T. (1996) Quantile comparison functions in two-sample problems, with application to comparisons of diagnostic markers. *JASA*, **91**, 689-698.
[6] Orban, J. & Wolfe, D. A. (1982) A class of distribution-free two-sample tests based on placements. *JASA,* **77***, 666-672.*

3.4 Papers Related to Sensitivity Analysis

References

[1] Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M., and Wynder, E. (1959) Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* **22** 173-203.
[2] Copas, J. B. and Li, H.G. (1997) Inference for non-random samples (with discussion). *JRSS* **B59** 55-96.
[3] Gastwirth, J. L. (1992) Methods for assessing the sensitivity of statistical comparisons used in Title VII cases to omitted variables. *Jurimetrics* **33** 19-34.
[4] Greenhouse, S. (1982) Jerome Cornfield's contributions to epidemiology. *Biometrics* Supplement, 33-45.
[5] Rosenbaum, P. R. (1987) Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, 74, 13-26.
[6] Rosenbaum, P. R. (1995) Quantiles in nonrandom samples and observational studies. *JASA,* 90, 1424-1431.
[7] Rosenbaum, P. R., Rubin, D. B. (1983) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *JRSS*, **B45**, 212–218.

3.5 Additive Effects

References

[1] Hodges, J. L. and Lehmann, E. L. (1963) Estimates of location based on ranks. *Annals of Mathematical Statistics* **34** 598-611.

[2] Lehmann, E. L. (1963) Nonparametric confidence intervals for a shift parameter. *Annals of Mathematical Statistics* **34** 1507-1512.
[3] Lehmann, E. L. (1999) *Nonparametrics: Statistical Methods Based on Ranks.* Upper Saddle River, NJ: Prentice Hall.
[4] Moses, L. E. (1965) Confidence limits from rank tests. *Technometrics* **7** 257-260.

3.6 The Examples

References

[1] Coats, T. J. and Walter, D. P. (1999) Effect of station design on death in the London Underground: Observational study. *British Medical Journal*, **319**, 957.
[2] Thun, M. (1993) Kidney dysfunction in cadmium workers. In: *Case Studies in Occupational Epidemiology*, ed., K. Steenland, New York: John Wiley, pp105-126.
[3] Thun, M., Osorio, A., Schober, S., et. al. (1989) Nephropathy in cadmium workers: assessment of risk from airborne occupational exposure to cadmium. *British Journal of Industrial Medicine*, **46**, 689-697.