Observational Studies

Statistics
Summer 2022

1 Randomized Experiments

1.1 Units and Treatment Assignments

Suppose there are N units divided into S strata, we write $Z_{si} = 1$ if the ith unit in stratum s receives the treatment and write $Z_{si} = 0$ if it receives the control. m_s is the number of units in stratum s, so $m_s = \sum_{i=1}^{n_s} Z_{si}$. Notice that if no covariates are used to divide the units, then there is a single stratum containing all units. $\mathbf{Z}$ is the n by 1 vector containing all the treatmente assignments.

1.2 Assigning Treatments at Random

Mathematically, every unit has a non zero chance of receiving both the treatment and the control: $0 < P(Z_{si} = 1) < 1 \, \forall s, i$. Ω_0 is the set containing all possible values of $\mathbf{Z}$

Simplest assigns treatments independently to different units, taking $P(Z_{si} = 1) = \frac{1}{2}$ for all s,i. $P(\mathbf{Z} = \mathbf{z}) = 1/2^N$ since $|\Omega_0| = 2^N$.

Note 1. If the mechanism is not independent, is it still a randomization?

Another assignment mechanism, we fix the number m_s. For instance, if n_s is required to be even and m_s is required to equal $n_s/2$ for each s, then half the units in each stratum receive the treatment and half receive the control. $S = 1$, we call it a completely randomized experiment. If $S >= 2$, we call it a randomized block experiment.

Suppose we are fixing m_s, let Ω be the set containing the $K = \prod_{i=1}^{S} \binom{n_s}{m_s}$ possible treatment assignments $\mathbf{Z}$ in which each Z_s is an n_s-tuple with m_s ones and $n_s - m_s$ zeros. Then, We call it a uniform randomized experiment. Note that if each of these K assignments is given the same probability, then $P(\mathbf{Z} = \mathbf{z}) = 1/K \; \forall \, \mathbf{z} \in \Omega$

Note 2. Rosenbaum listed the sequential assignment mechanism provided by Efron(1971).

1.3 Fisher's Sharp Null

To say that the treatment has no effect on this response is to say that each unit would exhibit the same value of the response whether assigned to treatment or control. When the treatment is without effect, the response of a unit is fixed, in the sense that this response would not change if a different treatment assignment $\mathbf{Z}$ were selected from Ω. The response of the ith unit in stratum s is written r_{si} and the N-tuple of responses for all N units is written $\mathbf{r}$.

We introduce the difference in means test statistics:

$$t(\mathbf{Z}, \mathbf{r}) = \frac{\mathbf{Z}^T \mathbf{r}}{\mathbf{Z}^T \mathbf{1}} - \frac{(1-\mathbf{Z})^T \mathbf{r}}{(1-\mathbf{Z})^T \mathbf{1}} \tag{1}$$

The significance level is defined in the common way. Suppose we have an observed value T:

$$P(t(\mathbf{Z}, \mathbf{r})) = \sum_{\mathbf{z} \in \Omega} 1\{t(\mathbf{z}, \mathbf{r}) \geq T\} P(\mathbf{Z} = \mathbf{z}) \tag{2}$$

In the case of a uniform randomized experiment, we can simplify the above expression a little bit, since $P(\mathbf{Z} = \mathbf{z}) = 1/|\Omega|$. Therefore ,

$$P(t(\mathbf{Z}, \mathbf{r}) \geq T) = \frac{|\mathbf{z} \in \Omega : t(\mathbf{Z}, \mathbf{r} \geq T)|}{K} \tag{3}$$

1.4 Some Common Randomization Tests

Here we assume the experiment is the uniform randomized experiment.

Wilcoxon's rank sum test
In the test, the responses are ranked from smallest to largest. If all N responses were different numbers, the ranks would be the numbers 1 to N. Write q_i for the rank of r_i. Note that the ranks q are a function of the responses which are fixed if the treatment has no effect, so q is also fixed. The rank sum statistics is $t(\mathbf{Z}, \mathbf{r}) = \mathbf{Z}^T \mathbf{q}$. This is equivalent to the Mann and Whitney test.

Wilcoxon's signed rank test
In the case of S matched pairs with n_2=2 and $m_s = 1$, the test is commonly used for responses taking many values. In this test, the absolute differences in responses within paris $|r_{s1} - rs2|$ are ranked from 1 to S, with average ranks used for ties. Let d_s be the rank of $|r_{s1} - rs2|$ thus obtained. The signed rank statistics is the sum of the ranks for pairs in which the treated unit had a higher response than the control unit.

Explanation
Let $c_{s1} = 1$ if $r_{s1} > r_{s2}$ and $c_{s1} = 0$ otherwise. Similarly, $c_{s2} = 1$ if $r_{s2} > r_{s1}$ and $c_{s2} = 0$ otherwise. Then $Z_{s1}c_{s1} + Z_{s2}c_{s2}$ equals 1 if the treated unit in pairs had a higher response than the control unit, and equals zero otherwise. It follows that the signed rank statistics is $\sum_{s=1}^{S} d_s \sum_{i=1}^{2} c_{si} Z_{si}$.

Note 3. There are other test statistics in the book and also other classes of statistics except the sum statistics. Refer to P.30

1.5 SUTVA

Rubin calls it the stable unit treatment value assumption. Formally, no interference means that $r_{si\mathbf{z}}$ varies with z_{si} but not with the other coordinates of $\mathbf{z}$. In other words, the response of the ith unit in stratum s depends on the treatment assigned to this unit, but not on the treatments assigned to other units, so this unit has only two possible values of the response rather than $|\Omega|$ possible values. When the model is assumed, we write $R_{si} = Z_{si} r_{Tsi} + (1 - Z_{si}) r_{Csi}$.

1.6 Confidence Interval

Consider testing the hypothesis $H_0 = \tau = \tau_0$ in the model of an additive effect, $r_{si} = r_{Csi} + \tau Z_{si}$. If the null hypothesis was true, then $\mathbf{r_c} = \mathbf{R} - \tau_0 \mathbf{Z}$. So testing the hypothesis is the same as testing that $\mathbf{R} - \tau_0 \mathbf{Z}$ satisfies the null hypothesis of no treatment effect.

Under the model of an additive treatment effect, a $1 - \alpha$ confidence set for τ is obtained by testing each value of τ and collecting all values not rejected at level α into a set A.

1.7 Point Estimates

Randomized experiments lead to unbiased estimates of the average treatment effect.

$$E\left\{\sum \frac{Z_i R_i}{m} - \frac{(1-Z_i)R_i}{N-m}\right\} = E\left\{\sum \frac{Z_i r_{Ti}}{m} - \frac{(1-Z_i) r_{Ci}}{N-m}\right\} = \frac{1}{N}\sum (r_{Ti} - r_{Ci}) \tag{4}$$

The difference in sample means may be biased when there are two or more strata and the experimenter assigns disproportionately more subjects to the treatment in some strata than in others. We can correct it by a direct adjustment:

$$\sum_{s=1}^{S} \frac{n_s}{N} \sum_{i=1}^{n_S} \left\{ \frac{Z_i R_i}{m} - \frac{(1-Z_i)R_i}{N-m} \right\} \tag{5}$$

We can check that the above estimate is unbiased.

2 Overt Bias in Observational Studies

An overt bias is one that can be seen in the data at hand - for instance, prior to treatment, treated subjects are observed to have lower incomes than controls. Overt biases are controlled using adjustments, such as matching or stratification.

2.1 Adjustments by Exact Stratification and Matching

Initially, there are M units available for study, and each has a value of an observed covariate $\mathbf{x}$, which may contain several variables. Imagine that unit j in assigned to treatment with probability $\pi_j = P(Z_j = 1)$ and to control with probability $1 - \pi_j$, with assignments for distinct units being independent and each unit has a chance to receive the treatment. The model here is equivalent to flipping biased coins, thus the model says:

$$P(Z_1 = z, \ldots, Z_M = z_m) = \prod_{j=1}^{M} \pi_j^{z_j} (1 - \pi_j)^{1 - z_j} \tag{6}$$

The problem of an observational study is that the $\pi_j s$ are unknown, so we don't have a known distribution of treatment assignments. An observational study is free of hidden bias if the π's, are

known to depend only on the observed covariates $\mathbf{x}$, so two units with the same value of $\mathbf{x}$ have the same chance π of receiving the treatment. That means there exists a function $\lambda(\cdot)$, such that $\pi_j = \lambda(\mathbf{x}_j)$. Then the joint distribution becomes:

$$P(Z_1 = z, \ldots, Z_M = z_m) = \prod_{j=1}^{M} \lambda(\mathbf{x}_j)^{z_j}(1 - \lambda(\mathbf{x}_j))^{1-z_j} \tag{7}$$

The function is called the propensity score.

2.1.1 Stratifying on x

From the M units, select $N \leq M$ units and group them into S non overlapping strata with n_s units in stratum s. Write m_s as the number of treated units in stratum s.

An exact stratification on x has strata that are homogeneous in x, so two units are in the same stratum only if they have the same value of x, that is $x_{si} = x_{sj}$ for all s,i,j. If the study is free of hidden bias, the model implies:

$$P(Z = z) = \prod_{s=1}^{S} \prod_{i=1}^{n_s} \lambda_x^{z_{si}}(1 - \lambda_s)^{1-z_{si}} \tag{8}$$

Consider the conditional distribution of Z given m, so the treatment assignment z has the unconditional probability:

$$P(Z = z) = \prod_{s=1}^{S} \lambda_s^{m_s}(1 - \lambda_s)^{n_s - m_s} \tag{9}$$

The result does not say that there is no difference between an experiment and an observational study. The difference is that in a uniform randomized experiment, the assignment probabilities are known to equal 1/K because we forced this to be true by randomizing. In an observational study, the conclusion $P(Z = z|m) = 1/K$ is deduced from the premise that the study if free of hidden bias.

2.1.2 Matching on x

One way that matching differs from stratification is that there are constraints on the number m_s of treated units and the number $n_s - m_s$ of control units in a stratum. For instance, pair matching requires $n_s = 2$ and $m_s = 1$ for each s.

3 Sensitivity to Hidden Bias

Motivating example: Whether the association between smoking and lung cancer is an effect caused by smoking or whether it is instead due to a hidden bias.

3.1 Model Expressed in Terms of Assignment Probabilities

There is hidden bias if two units with the same observed covariates x have different chances of receiving the treatments, that is, if $x_j = x_k$ but $\pi_j \neq \pi_k$.

A sensitivity analysis asks: How would inferences about treatment effects be altered by hidden biases of various magnitudes?

The odds that units j and k receive the treatment are, respectively, $\pi_j/(1-\pi_j)$ and $\pi_k/(1-\pi_k)$, and the odds ratio is the ratio of these odds. Imagine that we knew that this odds ratio for units with the same x was at most some number $\Gamma \geq 1$:

$$\frac{1}{\Gamma} \leq \frac{\pi_j(1-\pi_k)}{\pi_k(1-\pi_j)} \leq \Gamma \;\; \forall j,k \;\; and \;\; x_j = x_k \tag{10}$$

If $\Gamma = 1$, then the study would be free of hidden bias. If $\Gamma = 2$, then two units who appear similar, who have the same x, could differ in their odds of receiving the treatment by as much as a factor of 2, so one could be twice as likely as the other to receive the treatment. So Γ is a measure of the degree of departure from a study that is free of hidden bias.

3.2 Model Expressed in Terms of an Unobserved Covariate

Unit j has both an observed covariate x_j and an unobserved covariate u_j. The model has two parts, a logit from linking treatment assignment Z_j to the covariates (x_j, u_j) and a constraint on u_j, namely,

$$log\left(\frac{\pi_j}{1-\pi_j}\right) = \kappa(x_j) + \gamma u_j, \quad 0 \leq u_j \leq 1 \tag{11}$$

Here u_j is the unobserved covariate, κ and γ are unknown function and parameter.

Note 4. Is it possible to use the magnitude of the error term to detect the hidden bias? For example, randomize within our samples, and control on x, comparing the two groups' error term's magnitude.

Suppose units j and k have the same observed covariate, so $x_j = x_k$, and hence $\kappa(x_j) = \kappa(x_k)$. In adjusting for x, these two units might be matched together or might be placed in the same stratum. Consider the ratio of the odds that units j and k receive the treatment:

$$\frac{\pi_j(1-\pi_k)}{\pi_k(1-\pi_j)} = exp\{\gamma(u_j - u_k)\} \tag{12}$$

In other words, two units with the same x differ in their odds of receiving the treatment by a factor that involves the parameter γ and the difference in their unobserved covariates u.

Proposition 1. *With $e^{\gamma} = \Gamma \geq 1$, there is a model of the logit form which describes the $\pi_1, \ldots, \pi_M$ if and only if the model expressed in terms of assignment probabilities hold.*

Converse can be proved by reverse engineering.

3.3 Distribution of Treatment Assignments After Stratification

The conditional distribution of the treatment assignment Z given m is no longer constant because of the hidden bias.

$$P(Z = z|m) = \frac{exp(\gamma z^T u)}{\sum_{b \in \Omega} exp(\gamma b^T u)} = \prod_{s=1}^{S} \frac{exp(\gamma z_s^T u_s)}{\sum_{b_s \in \Omega_s} exp(\gamma z_s^T u_s)} \tag{13}$$

u_j is the vector of unobserved covariates under stratum j. Ω_s contains the $\binom{n_s}{m_s}$ different n_s-tuples with m_s ones and $n_s - m_s$ zeros.

The model implies that stratification on x was useful in that it eliminated part of the uncertainty about the unknown π's, the part due to $\kappa(x)$, but stratification on x was insufficient to render all treatment assignments equally probable, as they would be in an experiment.

3.4 Matched Pairs

With S matched pairs, each stratum s has 2 units, one of which received the treatment, $1 = m_s = Z_{s1} + Z_{s2}$, so $Z_{s2} = 1 - Z_{s1}$. The model describes S independent binary trials with unequal probabilities, namely,

$$P(Z = z|m) = \prod_{s=1}^{S} \left[\frac{exp(\gamma u_{s1})}{exp(\gamma u_{s1}) + exp(\gamma u_{s2})}\right]^{z_{s1}} \left[\frac{exp(\gamma u_{s2})}{exp(\gamma u_{s1}) + exp(\gamma u_{s2})}\right]^{1-z_{s1}} \tag{14}$$

If $\gamma > 0$, then the first unit in pair s is more likely to receive the treatment than the second if $u_{s1} > u_{s2}$.

Common test statistics for matched pairs are sign-score statistics, including Wilcoxon's signed rank statistic and McNemar's statistic.

$$T = t(Z, r) = \sum_{s=1}^{S} d_s \sum_{i=1}^{2} c_{si} Z_{si} \tag{15}$$

c_{si} is binary and both $d_s \geq 0$ and $c_s i$ are functions of r, and so are fixed under the null hypothesis of no treatment effect. However, since (γ, u) is unknown here, we don't have the distribution as in the randomization inference.

Specifically, for each possible (γ, u), the test statistic is the sum of S independent random variables, where the sth variable equals d_s with probability

$$p_s = \frac{c_{s1}exp(\gamma u_{s1}) + c_{s2}exp(\gamma u_{s2})}{exp(\gamma u_{s1}) + exp(\gamma u_{s2})} \tag{16}$$

and equals 0 with probability $1 - p_s$. A pair is said to be concordant if $c_{s1} = c_{s2}$. If $c_{s1} = c_{s2} = 1$, then $p_s = 1$, while if $c_{s1} = c_{s2} = 0$ then $p_s = 0$, so concordant pairs contribute a fixed quantity to $t(Z, r)$ for all possible (γ, u)

Though the null distribution of $t(Z, r)$ is unknown, for each fixed γ, the null distribution is bounded by two known distributions. With $\Gamma = exp(\gamma)$, define p_s^+ and p_s^- in the following way:

$$p_s^+ = \begin{cases} 0 & c_{s1} = c_{s2} = 0 \\ 1 & c_{s1} = c_{s2} = 1 \\ \frac{\Gamma}{1+\Gamma} & c_{s1} \neq c_{s2} \end{cases} \tag{17a}$$

$$p_s^- = \begin{cases} 0 & c_{s1} = c_{s2} = 0 \\ 1 & c_{s1} = c_{s2} = 1 \\ \frac{1}{1+\Gamma} & c_{s1} \neq c_{s2} \end{cases} \tag{17b}$$

It follows that $p_s^- \leq p_s \leq p_s^+$, for all s. Define T^+ to be the sum of S independent random variables, where the sth takes the value d_s with probability p_s^+ and takes the value 0 with probability $1 - p_s^+$. Define T^- similarly with in p_s^- place of p_s^+. The following proposition says that, for all $u \in U$, the unknown null distribution of the test statistic $T = t(Z, r)$ is bounded by the distributions of T^+ and T^-.

Proposition 2. *If the treatment has no effect, then for each fixed γ, $P(T^+ \geq a) \geq P(T \geq a|m) \geq P(T^- \geq a)$ for all a and $u \in U$.*

For each γ, the proposition places bounds on the significance level that would have been appropriate had **u** been observed. The sensitivity analysis for a significance level involves calculating these bounds for several values of γ.

The bounds in the proposition are attained for two values of u. The upper bound $P(T^+ \geq a)$ is the distribution of $t(Z, r)$ when $u_{si} = c_{si}$ and the lower bound $P(T^- \geq a)$ is the distribution of $t(Z, r)$ when $u_{si} = 1 - c_{si}$. The first consequence is that bounds in the proposition cannot be improved unless additional information is given about the value of u. Second, the bounds are attained at values of u which perfectly predict the signs c_{si}.

The bounding distributions have easily calculated moments.

$$E(T^+) = \sum_{s=1}^{S} d_s p_s^+ \ \ and \ \ Var(T^+) = \sum_{s=1}^{S} d_s^2 p_s^+ (1 - p_s^+) \tag{18}$$

3.5 The Power of Sensitivity Analysis and Its Limit

This section is adapted from the Design of Observational Studies by Rosenbaum. First, we define some notations, let F denote the array of quantities $\{(r_{Tij}, r_{Cij}, x_{ij}, u_{ij}, i = 1 \cdots I, j = 1, 2)\}$ Let L denote the set of the 2^I possible values of Z. The goal of this section is understanding some technical results about sensitivity analysis.

3.5.1 Computing Power in a randomized experiment

Here, we use Wilcoxon's signed rank statistic T as an example. Power is the probability of rejecting the null hypothesis with a P-value less than or equal to 0.05 when the null hypothesis is false.

The first step is determining the critical value assuming the null hypothesis is true. If the null hypothesis of no effect is true, then the expectation and variance of T are given by:

$$E(T|F, L) = \frac{I(I+1)}{4} \quad and \quad Var(T|F, L) = \frac{I(I+1)(2I+1)}{24} \tag{19}$$

The distribution of T is well approximated by standard Normal using CLT:

$$\frac{T - E(T|F, L)}{\sqrt{var(T|F, L)}} \to N(0, 1) \ \ as \ \ I \to \infty \tag{20}$$

With $I = 100$ pairs, compute $E(T|F, L) = 100(100 + 1)/4 = 2525$ and $Var(T|F, L) = 100(100 + 1)(2 \times 100 + 1)/24 = 84587.5$. We use the normal approximation to find the probability that $T \geq 3005$, given that $(3005 - 2525)/\sqrt{84587.5} = 1.65$, so $P(T \geq 3005|F, L) = 1 - \Phi(1.65) = 0.05$. More generally, under the null hypothesis:

$$P(T \geq \xi|F, L) = 1 - \Phi\left(\frac{\zeta - E(T|F, L)}{\sqrt{var(T|F, L)}}\right) \tag{21}$$

For power calculations, we do not want the realized P-value, instead, we want the quantile ζ_α of the distribution of T that corresponds with a P-value less than or equal to α. Rearranging the above formula:

$$\zeta_\alpha = E(T|F, L) + \Phi^{-1}(1 - \alpha)\sqrt{var(T|F, L)} \tag{22}$$

The second step is determining the power assuming the null hypothesis is false. One way to proceed would be to specify in detail what is true, that is to specify F, in particular to specify the responses, (r_{Tij}, r_{Cij}),each subject ij would exhibit under treatment and under control. This specification would permit the calculation of the conditional power given F, more precisely, $P(T \geq \zeta_\alpha|F, L)$ for some specific F for which there is a nonzero treatment effect.

A simple case: constant effect with random errors

This definition of power is quite flexible. In the simplest case, the treated-minus-control differences in observed responses, $Y_i = (Z_{i1} - Z_{i2})(R_{i1} - R_{i2})$ are iid observations drawn from a distribution $F(\cdot)$. For instance, a simple model to generate F has differences in responses to control, $r_{Ci1} - r_{Ci2}$, that are iid observations from a Normal distribution with expectation 0 and variance ω^2, that is, $r_{Ci1} - r_{Ci2} \sim N(o, \omega^2)$, and an additive constant treatment effect, $r_{Tij} - r_{Cij} = \tau$ for all ij.

Then, in a randomized experiment, the treated minus control difference in observed responses, $Y_i = (Z_{i1} - Z_{i2})(R_{i1} - R_{i2})$, is $Y_i = \tau + (Z_{i1} - Z_{i2})(r_{Ci1} - r_{Ci2})$, where given F and L, the quantity $Z_{i2} - Z_{i2}$ is 1 or -1 with equal probabilities, so $Y_i \sim N(\tau, \omega^2)$. The power in this case is the chance that $T \geq \zeta_\alpha$ when Wilcoxon's signed rank statistic is computed from I independent observations drawn from $N(\tau, \omega^2)$.

When computing the power for differences Y_i that are sampled independently from a continuous distribution. Lehmann gives an approximation to the power of Wilcoxon's signed rank statistic, T. Theapproximation appeals to the central limit theorem twice as the sample size I increases, once to obtain the critical value ζ_α in step 1 as has already been done, and the second time to approximate the behavior of T when the null hypothesis is false. For $Y_i \sim F(\cdot)$, define $p = P(Y_i > 0), p_1^{'} = P(Y_i + Y_{i'} > 0)$ and $p_2^{'} = P(Y_i + Y_{i'} > 0 \; and \; Y_i + Y_{i''} > 0)$ with $i < i^{'} < i^{''}$. Lehmann shows that the expectation and the variance of Wilcoxon's T when the null hypothesis is false and the Y_i are sampled independently from $F(\cdot)$ is:

$$\mu_F = \frac{I(I-1)p_1^{'}}{2} + Ip \tag{23}$$

$$\sigma_F^2 = I(I-1)(I-2)(p_2^{'} - p_1^{'2}) + \frac{I(I-1)}{2}\left(2(p - p_1^{'})^2 + 3p_1^{'}(1 - p_1^{'})\right) + Ip(1-p) \tag{24}$$

and that the central limit theorem yields approximate power as $P(T \geq \zeta_\alpha | F) = 1 - \Phi\{(\zeta_\alpha - \mu_F)/\sigma_F\}$

3.5.2 Computing Power in an Observational Study

If Wilcoxon's signed rank test is used to test the null hypothesis of no treatment effect in a randomized experiment, it yields a single P-value because we know the distribution of treatment assignments that was created by the random assignment of treatments. However, randomization is not used to assign treatments in an observational study, so the single P-value lacks a justification. Recall that for any given magnitude $\Gamma \geq 1$, an interval $[P_{min}, P_{max}]$ of possible P-values may be determined.

Define the "favorable situation" as a treatment that actually causes meaningful effects in an observational study that is free of unmeasured biases. In testing the null hypothesis of no treatment effect at level α, conventionally equals to 0.05, rejection of the null hypothesis is insensitive to a bias of magnitude Γ if $P_{max} \leq \alpha$. For the favorable situaion, this is the power of a test of no effect in a randomized experiment. Using the formula for quantile above:

$$\zeta_{\Gamma,\alpha} = \frac{\Gamma}{1+\Gamma}\frac{I(I+1)}{2} + \Phi^{-1}(1-\alpha)\sqrt{\frac{\Gamma}{(1+\Gamma)^2}\frac{I(I+1)(2I+1)}{6}} \tag{25}$$

Similarly, the second step is computing the power with the new critical value. It is generally performed by simulating many sets of data F from the model, determining the proportion with $T \geq \zeta_{\Gamma,\alpha}$, and in particular cases the power of the sensitivity analysis may be determined analytically.

3.5.3 Design Sensitivity

The power of a sensitivity analysis tends to 1 as the sample size I increases if the analysis is performed with a sufficiently small value of Γ. Notice that, as the sample size I increases, the power viewed as a function of Γ is tending to a step function with a single step down from power 1 to power 0, with the step located at the design sensitivity, $\tilde{\Gamma}$.

4 Detecting Hidden bias

After reading the two books written by Rosenbaum, I still can not figure out the most suitable topic for me. However, before reading the Sensitivity Analysis part, I realized there are probably some ways to detect hidden bias. Now I should restrict myself to this topic and see what I will get.

4.1 Known Effects

In addition to the response R of primary interest, the study includes another outcome, say Y, in an effort to detect hidden bias. Suppose a test statistic $T = t(Z, R)$ is computed from the responses R of primary interest, and then T is compared to the uniform or random distribution of Z on Ω under the null hypothesis that R=r is unaffected by the treatment. If the significance level is small, it may be evidence that the null hypothesis of no treatment effect is untrue and the treatment does affect R, or it may be evidence that there is a hidden bias and the distribution of treatment assignments Z is not random.

Assume y is unaffected by the treatment, we can calculate a statistic $T^* = t^*(Z, y)$ and its significance level $|z \in \Omega : t^*(z, y) \geq T^*|/K$. If the significance level is small, then there is evidence of hidden bias, that is, evidence that Z is not uniformly or randomly distributed on Ω.

Let 'a' be a number such that:

$$\alpha = \frac{1}{K} \sum_{z \in \Omega} 1(t^*(z, y) \geq a) = \frac{|\{z \in \Omega : t^*(z, y) \geq a\}|}{K} \tag{26}$$

So a test for hidden bias which rejects when $t^*(Z, y) \geq \alpha$ has level α. The power of this test under the model in matched pair is:

$$\beta(y, u) = \sum_{z \in \Omega} 1(t^*(z, y) \geq a) \frac{exp(\gamma z^T u)}{\sum_{b \in \Omega} exp(\gamma b^T u)} \tag{27}$$