CAUSAL INFERENCE

FOR

MULTIPLE TREATMENTS

VIA

SUFFICIENCY AND

RATIOS OF

GENERALIZED PROPENSITY SCORES

Yannis G. Yatracos

Cyprus U. of Technology

March 26, 2012

e-mail: yannis.yatracos@cut.ac.cy

Summary

The coarsest balancing score for multiple treatments $\mathcal{T}$ is the minimal sufficient statistic s for the covariates' distributions, $\mathcal{D}_{\mathcal{T}}$, of populations $\{\mathcal{P}_t,\ t \in \mathcal{T}\}$ receiving each treatment $t \in \mathcal{T}$. A unit in $\mathcal{P}_r$ with covariates x is a good match with respect to $\mathcal{T}$ for a unit in $\mathcal{P}_t$ with covariates y, when x and y *provide similar information* for $\mathcal{D}_{\mathcal{T}}$, i.e. when $s(x) \approx s(y)$. For finite, countably finite and often continuous treatments, $s(x)$ is shown to be equivalent to $\tilde{e}$, a vector of propensities' ratios. The units in $\{P_t,\ t \in \mathcal{T}\}$ can be divided into subpopulations where causal comparisons are simultaneously valid. Satisfactory s-matchings are obtained for simulated covariates in R^3. The use of $\tilde{e}$'s estimate rather than s's estimate allows to avoid the x-curse of dimensionality, but the available data's size in each case is critical for the final choice.

Some key words: Causal Inference, Generalized Propensity Scores, Matching, Minimal Sufficient Statistic

Running head: S-matching for multiple treatments

1 Introduction

In observational studies for causal effects of two treatments, $t = 1,\ 2$, Rosenbaum and Rubin (1983) proposed the propensity score $e(x)$, or its estimate $\hat{e}(x)$, to balance the pre-treatment covariates $x(\in R^p)$ of the n units in the treatment groups; $e(x)$ is the conditional probability of receiving, say, treatment 1 given x, and $\hat{e}(x)$ is usually obtained using the logit model. It was showed therein, among others, that $e(x)$ is the coarsest balancing score and, if the potential units' responses to treatments, $r_i(1)$ and $r_i(2),\ i = 1, \ldots, n$, and treatment assignment are conditionally independent given x, that the difference between the sample treatments' means given $e(x)$ is an unbiased estimate of the average causal effect $E\{r(2) - r(1)\}$; E denotes expectation over the whole population. According to Rubin and Thomas (1996, p. 250), R. Bahadur recognized $e(x)$ is equivalent to the likelihood ratio of the x-populations which is minimal sufficient.

For more than two treatment levels $t \in \mathcal{T}$, Joffe and Rosenbaum (1999) studied causal effects using, under special circumstances, a single-variable balancing score, and under certain models a small number of balancing linear functions of x. For multi-valued categorical treatments and T the treatment variable, Imbens (2000) introduced the generalized propensity score $P(T = t|x)$ for each level t and used it to estimate average causal effects. For general

treatment regimes, with t-values either discrete (ordered or not), or continuous, semi-continuous or multivariate, Imai and Van Dyk (2004) introduced the propensity function and, assuming for all x-values it is defined from a unique finite dimensional parameter, established causal effects.

The generalized propensity scores and the propensity function are balancing scores only under specific models and assumptions. The main contribution in this work is to provide in *general setting* the coarsest balancing score s and use it in causal inference; s is the minimal sufficient statistic of the covariates' distributions and is shown to be equivalent to $\tilde{e}$, consisting of propensities' ratios. Logit modeling can be used for both s and $\tilde{e}$ confirming the results in Joffe and Rosenbaum (1999). A unit with covariate y is a good match for a unit with covariate x with respect to $\mathcal{T}$, when they provide similar information for all the covariates' distributions, i.e. when $s(x) \approx s(y)$. These subpopulations can be used in causal comparisons, for example, to determine the "right" dose for a new drug, by examining simultaneously the expected response differences $E\{r(t_2) - r(t_1)\},\ E\{r(t_3) - r(t_2)\}, \ldots, E\{(r(t_k) - r(t_{k-1})\}$ for different doses' levels $t_1 < t_2 < \ldots < t_k$. Assuming monotonicity of $Er(t_j),\ j = 1, \ldots, k$, the "right" dose could be the smaller in the expected response difference that is closer to zero.

Strong ignorability of treatment assignment given $s(x)$ is established, and

the expected treatments' differences given $s(x)$ are shown to be unbiased for the average causal effects of any treatments' differences. These results hold also using $\tilde{e}(x)$ instead of $s(x)$. Some directions are given for s-mathing's implementation in practice, followed by s-matchings of simulated covariates from normal mixtures in R and normal densities in R^3. When s and $\tilde{e}$ are unknown, $\tilde{e}$'s estimate may be preferred to avoid the x-curse of dimensionality via propensity scores' methodology, but the available data's size in each case is critical for the final choice The proofs are in the Appendix and the Figures after the references.

2 Causal inference framework and assumptions

For random vectors $U,\ V$, use $p(u|v)$ to denote the conditional probability or conditional density $p_{U|V}(u|v)$. Let $\mathcal{T}$ denote the treatments and let T be the treatment variable with values in $\mathcal{T}$. Treatment $t(\in \mathcal{T})$ is used in selected elements of population $\mathcal{P}_t$ having balanced covariates $x \in R^p$ with respect to $\mathcal{T}$. The units in $\mathcal{P}_t$ have covariates $x \in \mathcal{C}(\mathcal{P}_t) \subset R^p$. Let $p(x|t)$ denote the x-covariates' density/probability of units in $\mathcal{P}_t$ and let $\mathcal{D}_{\mathcal{T}} = \{p(x|t),\ t \in \mathcal{T}\}$. Unless otherwise stated it is assumed that $\mathcal{C}(\mathcal{P}_t) = \mathcal{C}, t \in \mathcal{T}$. For unit $i,\ r_i(t)$ is the response for treatment t and the potential outcomes $\mathcal{R} = \{r_i(t), t \in \mathcal{T},\ \text{for } i = 1, \ldots, n\}$.

The Stable Unit Treatment Value Assumption ($SUTVA$) is presented as in Imai and Van Dyk (2004):

Assumption 1 ($SUTVA$, Rubin, 1980, 1990) The distribution of potential outcomes for one unit is assumed to be independent of potential treatment status of another unit given the observed covariates.

Assumption 2 (Strong ignorability of treatment assignment, Rosenbaum and Rubin, 1983)

(i) $\mathcal{R}$ and T are conditionally independent given $x : \mathcal{R} \perp T \,|x$, and

(ii) for every $t \in \mathcal{T}$, $0 < p(t|x)$ (or equivalently $0 < p(x|t)$).

Recall that $b(x)$ is a balancing score if the conditional distribution of x given $b(x)$ is the same for all treatment values, i.e.

$$p(x|t, b(x)) = p(x|b(x)), \text{ for all } t \in \mathcal{T}. \tag{1}$$

From (1), thinking of t as parameter value for the distribution of x it follows that $b(x)$ is sufficient statistic for the family $\mathcal{D}_{\mathcal{T}} = \{p(x|t); t \in \mathcal{T}\}$.

3 Large-sample theory

As in Rosenbaum and Rubin (1983), the results in this section treat the minimal sufficient statistic $s(x)$ as known and are applicable for large samples.

Proposition 3.1 *Assume that there are k treatments, ordered or not, with the treatment variable taking values in $\mathcal{T} = \mathcal{T}_k = \{t_1, \ldots, t_k\}$. When the distributions of the covariates $\mathcal{D}_{\mathcal{T}_k} = \{p(x|t),\ t \in \mathcal{T}_k\}$ have common support $\mathcal{C}$,*

$$s(x) = s^{(1)}(x)^1 = \left(\frac{p(x|t_2)}{p(x|t_1)}, \frac{p(x|t_3)}{p(x|t_1)}, \ldots, \frac{p(x|t_k)}{p(x|t_1)}\right) \tag{2}$$

is a minimal sufficient statistic. Thus, $s(x)$ is the coarsest balancing score and treatment assignment and the observed covariates are conditionally independent given $s(x)$: $x \perp T|s(x)$. The same result holds for countably finite treatments and in some situations for treatments with continuous values.

The s-Matching Rule: *Match u to v when $s(u) = s(v)$.*

The next proposition shows that s-matching is not changed when

$$s^{(j)}(x) = \left(\frac{p(x|t_1)}{p(x|t_j)}, \ldots, \frac{p(x|t_{j-1})}{p(x|t_j)}, \frac{p(x|t_{j+1})}{p(x|t_j)}, \ldots, \frac{p(x|t_k)}{p(x|t_j)}\right), j \neq 1, \tag{3}$$

is used instead of $s = s^{(1)}$ in (2).

Proposition 3.2 *If $s(u) = s(v)$, then $s^{(j)}(u) = s^{(j)}(v),\ j > 1$.*

Without loss of generality $s(x) = s^{(1)}(x)$ is used in this section.

Proposition 3.3 *Consider $(k-1)$ ratios of generalized propensity scores*

$$\tilde{e}_j(x) = \frac{P(T = t_j|x)}{P(T = t_k|x)},\ j = 1, \ldots, k-1. \tag{4}$$

[1] In $s^{(1)}(x)$, (1) indicates the denominator is $p(x|t_1)$.

The ratios of propensity scores statistic

$$\tilde{e}(x) = (\tilde{e}_1(x), \ldots, \tilde{e}_{k-1}(x)), \tag{5}$$

is equivalent to s, i.e. $\tilde{e}$ is the coarsest balancing score.

From Proposition 3.3, one can use either s or $\tilde{e}$ but the results herein are obtained using s. When $k = 2$, $\tilde{e}(x)$ is equivalent to the propensity score $e(x)$.

Corollary 3.1 *Let $q_i(x) = \frac{p(x|t_i)}{p(x|t_i)+p(x|t_1)}$, $i = 2, \ldots, k$. Then,*

$$s(x) = \left(\frac{q_2(x)}{1 - q_2(x)}, \ldots, \frac{q_k(x)}{1 - q_k(x)} \right). \tag{6}$$

An expression similar to (6) can be obtained for $\tilde{e}$.

Remark 3.1 *Propositions 3.1 and 3.3 indicate that, with several treatments, the coarsest balancing score is expected to have dimension larger than one.*

Remark 3.2 *When the covariates' distributions $\mathcal{D}_{\mathcal{T}_k} = \{p(x|t),\ t \in \mathcal{T}_k\}$ do not have common support, the minimal sufficient statistic $s(x)$ has dimension that depends on x (Lehmann and Casella, 1998, p. 70, Theorem 9.1).*

We revisit an example in Rosenbaum and Rubin (1983, p. 47) when the number of treatments k is larger than 2.

Example 3.1 *Let $p(x|t)$ be a polynomial exponential family distribution,*

$$p(x|t) = h(x) \exp\{P_t(x)\},\ t = 1, \ldots, k,$$

with $P_t(x)$ a degree m polynomial. Then, the statistic

$$\left(\ln\frac{p(x|t_2)}{p(x|t_1)},\ldots,\ln\frac{p(x|t_k)}{p(x|t_1)}\right) = (P_2(x)-P_1(x),\ldots,P_k(x)-P_1(x))$$

$$= (Q_1(x),\ldots,Q_{k-1}(x))$$

is equivalent to the minimal sufficient statistic (2) with $Q_i(x)$ a degree m polynomial, $i = 1,\ldots,k-1$.

Proposition 3.4 *Strong ignorability of treatment assignment given the minimal sufficient statistic $s(x)$: Under Assumption 2, for the responses $\mathcal{R}$ and the treatment variable $T = t$ it holds*

$$p\{t,\mathcal{R}|s(x)=s\} = p\{t|s(x)=s\}\cdot p\{\mathcal{R}|s(x)=s\}. \tag{7}$$

Using $s(x)$ (or equivalently $\tilde{e}(x)$) to balance subpopulations for all treatments, causal comparisons can be drawn for any 2 or more treatments. If treatment assignment is strongly ignorable, adjustment for $s(x)$ or $\tilde{e}(x)$ is sufficient to obtain unbiased estimates of the average treatment effect(s).

Proposition 3.5 *Suppose that treatment assignment is strongly ignorable (Assumption 2) and that a value s_0 of $s(x)$ is randomly sampled from the population of units with covariates $x \in \mathcal{C}$. Units receiving treatments t_i and t_j are sampled with s-value for their covariates equal s_0, $i \neq j$. Then, the expected difference in response for the units chosen is the expected treatment effect at*

$s(x) = s_0$. The mean of such pair differences over all $s(x)$-values is unbiased for the average treatment effect $E\{r(t_i) - r(t_j)\}$ and the same holds, concurrently given $s(x)$, for any number of average treatment effects.

4 Small sample theory

When $p(x|t)$ is not known, $\mathcal{P}_t$'s available subpopulation $\tilde{\mathcal{P}}_t$ is used to obtain its estimate $\hat{p}_t(x)$, which is either parametric or non-parametric and will play $p(x|t)$'s role, $t \in \mathcal{T} = \mathcal{T}_k$. Then, $s(x)$ in (2) is replaced by $\hat{s}(x)$ obtained using $\hat{p}_t(x),\ t \in \mathcal{T}_k$. For (6), let $\hat{p}_i(x) = \hat{p}_{t_i}(x)$,

$$\hat{q}_i(x) = \frac{\hat{p}_i(x)}{\hat{p}_i(x) + \hat{p}_1(x)},\ i = 2, \ldots, k.$$

Instead of $\hat{s}$ in (6) we can consider the equivalent statistic

$$\hat{s}^* = \left(\log \frac{\hat{q}_2(x)}{1 - \hat{q}_2(x)}, \ldots, \log \frac{\hat{q}_k(x)}{1 - \hat{q}_k(x)} \right); \tag{8}$$

using logit modeling $\hat{s}^*$ consists of $(k-1)$ linear functions of x.

Proposition 4.1 *From the available subpopulation $\tilde{\mathcal{P}}_t$, obtain $\hat{p}_t(x)$ and let $p(x|t) = \hat{p}_t(x),\ x \in \mathcal{C},\ t \in \mathcal{T}_k$. Then,*

$$p(x|T = t, \hat{s}^*) = p(x|\hat{s}^*),$$

i.e. $\hat{s}^*$ in (8) is balancing score.

Let $\mathcal{MP}_t$ denote the units to be matched from $\tilde{\mathcal{P}}_t$-subpopulation. Use $s = s^{(1)}$ in (2) (or its estimates $\hat{s}$ or $\hat{s}^*$) to match a unit in $\mathcal{MP}_t$ having covariates u with a unit from $\tilde{\mathcal{P}}_r$ having covariates $v_{m,r} \in \mathcal{C}(\tilde{\mathcal{P}}_r)$, such that

$$v_{m,r} = \arg \min_{v \in \mathcal{C}(\tilde{\mathcal{P}}_r)} ||s(u) - s(v)||^2, \ r \in T - \{t\}; \tag{9}$$

$|| \cdot ||$ is the usual Euclidean distance in R^p.

Additional matching sets for $\mathcal{MP}_t$ can be obtained using $s = s^{(j)}$ (or its estimates) in (9), $j = 2, \ldots, k$, and the decision maker can select the "best" matching set, for example, that with the nearest means to the $\mathcal{MP}_t$ covariates' means with respect to $|| \cdot ||$ or the sup-norm distance $|| \cdot ||_\infty$.

In simulations, matching sets obtained using

$$\tilde{v}_{m,r} = \arg \min_{v \in \mathcal{C}(\tilde{\mathcal{P}}_r), 1 \le j \le k} ||s^{(j)}(u) - s^{(j)}(v)||^2, \ r \in \mathcal{T} - \{t\},$$

instead of $v_{m,r}$ in (9) were not satisfactory. For optimal matching methods, the interested reader is referred to the propensity score related literature, for example, in Rosenbaum (1989).

Remark 4.1 *To avoid the x-curse of dimensionality, the results in this section can be presented using the similarly obtained $\tilde{e}$'s estimate $\hat{\tilde{e}}$ (or $\tilde{e}^*$) instead of $\hat{s}$ (or $\hat{s}^*$), using existing methodology for propensity scores' estimates. However, available data's size is critical for the choice between these estimates.*

5 s-matchings with simulated covariates

It should be mentioned in advance that the quality of s-matchings in Examples 5.1 and 5.2 improves as the size N of the available subpopulations increases.

Example 5.1 *Each of 3 treatments is to be assigned to* $n = 20$ *units with covariates, respectively, in subpopulations of populations* U, V, W. *Subpopulations of size N=200 are obtained randomly from the* U*-population* $.2\mathcal{N}(0,4) + .8\mathcal{N}(3,16)$, *the* V*-population* $.4\mathcal{N}(8,9) + .6\mathcal{N}(10,25)$ *and the* W*-population* $.6\mathcal{N}(11,16) + .4\mathcal{N}(14,36)$; $\mathcal{N}(\mu,\sigma^2)$ *denotes the normal population with mean* μ *and variance* σ^2. *The first 20 values* $u_1,\ldots,u_{20}$ *from the* U*-subpopulation are to be matched with 20 values from the* V *and the* W*-subpopulations.*

The minimal sufficient statistic $s = s^{(1)}$ *from (2) is only used with* $k = 3$; $p(x|t_1)$ *is the density of the* V*-population and* $p(x|t_2)$ *and* $p(x|t_3)$ *are the densities, respectively, of the* U*-population and the* W*-population.*

Value u_i *is matched with the value* $v_{m,i}$ *from the* V*-subpopulation, obtained with replacement, such that*

$$v_{m,i} = \arg \min_{1\le j\le 200} ||s(u_i) - s(v_j)||^2,\ i = 1,\ldots,20.$$

Matching for u_i *from the* W*-subpopulation is similarly obtained,* $i = 1,\ldots,20$.

Box-plots for subpopulations and the matched u, v *and* w *values are, respectively, in Figures 1 and 2. Densities' plots for* $u_1,\ldots,u_{20}$ *and for the matched*

v and w values are in Figure 3, obtained with the R-function "density( )". The averages of the U, V, W populations are, respectively, 2.4, 9.2 and 12.2, and the matched samples' averages are $\bar{u} = 2.251,\ \bar{v}_m = 3.018$ *and* $\bar{w}_m = 3.193$.

Example 5.2 *Repeat Example 5.1 with covariates U, V, W the 3-dimensional normal distributions with means respectively* $(3, 6, 8), (4, 2, 7)$ *and* $(5, 7, 4)$ *and covariance matrix* $12 * I$, *where I is the matrix identity. Subpopulations of size* $N = 2000$ *are obtained randomly from the U, V, W populations and* $n = 20$ *units are randomly chosen with covariates* $u_1, \ldots, u_{20}$ *in the U-subpopulation to be matched with covariates in the V and the W-subpopulations.*

The means of $u_1, \ldots, u_{20}$ *are* $\bar{\mathbf{u}} = (3.39, 7.63, 7.87)$. *Sets of matched covariates from the V and W-subpopulations are obtained using* $s = s^{(j)},\ j = 1, 2, 3$. *We select the matching set with means at minimum* $||\cdot||$*-distance from* $\bar{\mathbf{u}}$. *The means of the so-obtained v and w-matches are* $\bar{\mathbf{v}} = (3.84, 7.31, 8.03)$ *and* $\bar{\mathbf{w}} = (3.85, 7.55, 8.01)$. *Box-plots of the marginals of the U, V, W-subpopulations are in Figure 4 and those of the matched covariates in Figure 5. The corresponding densities are in Figure 6 and a 3-dimensional plot of the matched covariates from the U, V and W-subpopulations is in Figure 7.*

Acknowledgments

Many thanks are due to all the members of the Statistics Department, Harvard University, for their warm hospitality during my sabbatical visit from

N.U.S. in 2005-2006 and in particular to Professor Donald Rubin, my host, who also introduced me to matching methods in his lectures and in our discussions. Thanks are also due to Dr. K. Imai for his comments and suggestions.

6 Appendix

Proof of Proposition 3.1: It is direct consequence of Theorem 6.12, in Lehmann and Casella, 1998, p. 37. The same holds for treatments with continuous values in some situations; recall, for example, that if $p(x|t)$ follows a normal distribution with mean t and variance (say) 1, one can determine the one-dimensional minimal sufficient statistic via the likelihood ratio for any two values t_1 and t_2, using related Theorems in Lehmann and Casella (1998).

Proof of Proposition 3.2: Since $s(u) = s(v)$, it holds

$$\frac{p(u|t_i)}{p(u|t_1)} = \frac{p(v|t_i)}{p(v|t_1)},\ i = 2, \ldots, k. \tag{10}$$

In (10), divide the i-th equality with the j-th equality, $i \neq j$, and invert the j-th equality to obtain

$$\frac{p(u|t_i)}{p(u|t_j)} = \frac{p(v|t_i)}{p(v|t_j)},\ i \neq j,\ \text{or } s^{(j)}(u) = s^{(j)}(v).$$

Proof of Proposition 3.3: From Proposition 3.2, it is enough to prove that $s^{(k)}$ is equivalent to $\tilde{e}$, i.e. that $s^{(k)}(u) = s^{(k)}(v)$ if and only if $\tilde{e}(u) = \tilde{e}(v)$.

It hods

$$s^{(k)}(u) = \left(\frac{p(u|t_1)}{p(u|t_k)}, \ldots, \frac{p(u|t_{k-1})}{p(u|t_k)} \right)$$

$$= \left(\frac{P(T=t_k)}{P(T=t_1)} \frac{P(T=t_1|u)}{P(T=t_k|u)}, \ldots, \frac{P(T=t_k)}{P(T=t_{k-1})} \frac{P(T=t_{k-1}|u)}{P(T=t_k|u)} \right), \text{ or}$$

$$s^{(k)}(u) = \left(\frac{P(T=t_k)}{P(T=t_1)} \tilde{e}_1(u), \ldots, \frac{P(T=t_k)}{P(T=t_{k-1})} \tilde{e}_{k-1}(u) \right).$$

It follows that $s^{(k)}(u) = s^{(k)}(v)$ if and only if $\tilde{e}(u) = \tilde{e}(v)$.

Proof of Proposition 3.4: The proof follows the lines in Imai and Van Dyk (2004),

$$p\{x, t, \mathcal{R}|s(x) = s\} = p\{x, t|s(x) = s\} \cdot p\{\mathcal{R}|x, t, s(x) = s\}$$

$$= p\{t|s(x) = s\} \cdot p\{x|t, s(x) = s\} \cdot p\{\mathcal{R}|x, t, s(x) = s\}$$

$$= p\{t|s(x) = s\} \cdot p\{x|s(x) = s\} \cdot p\{\mathcal{R}|x, s(x) = s\}.$$

The third equality is obtained using Proposition 3.1 and strong ignorability of treatment assignment given x (Assumption 2). It follows that

$$p\{t, x, \mathcal{R}|s(x) = s\} = p\{t|s(x) = s\} \cdot p\{x, \mathcal{R}|s(x) = s\}$$

Integrating both sides of the last equation over the x's for which $s(x) = s$, we obtain that given $s(x) = s$, $\mathcal{R}$ and T are independent.

Proof of Proposition 3.5: From Assumption 2,

$$E\{r(t_i)|s(x) = s, T = t_i\} - E\{r(t_j)|s(x) = s, T = t_j\}$$

$$= E\{r(t_i)|s(x) = s\} - E\{r(t_j)|s(x) = s\} = E\{r(t_i) - r(t_j)|s(x) = s\}$$

and it follows that

$$E_s\left[E\{r(t_i) - r(t_j)|s(x) = s\}\right] = E\{r(t_i) - r(t_j)\};$$

E_s denotes expectation with respect to all values s of $s(x),\ x \in \mathcal{C}$.

Proof of Proposition 4.1: Follows from sufficiency of $\hat{s}^*$ for the family of probabilities/densities $\{\hat{p}_t(x),\ t \in \mathcal{T}_k\}$.

References

[1] Imai, K. and Van Dyk, D. A. (2004) Causal inference with general treatment regimes: generalizing the propensity score. *JASA* **99**, 854-866.

[2] Imbens, G. W. (2000) The role of the propensity score in estimating dose-response functions. *Biometrika* **87**, 706-710.

[3] Joffe, M. M. and Rosenbaum, P. R. (1999) Propensity scores. *American Journal of Epidemiology* **150**, 327-333.

[4] Lehmann, E. L. and Casella, G. (1998) *Theory of Point Estimation.* Springer, New York.

[5] Rosenbaum, P. R. and Rubin, D. B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.

[6] Rosenbaum, P. R. (1989) Optimal Matching for Observational Studies. *JASA* **84**, 1024-1032.

[7] Rubin, D. B. and Thomas, N. (1996) Matching using estimated propensity scores: relating theory to practice. *Biometrics* **52**, 249-264.

[8] Rubin, D. B. (1980) Discussion of paper by D. Basu. *JASA* **75**, 591-593.

[9] Rubin, D. B. (1990) Comments on "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," by J. Neyman, *Stat. Science* **5**, 472-480.

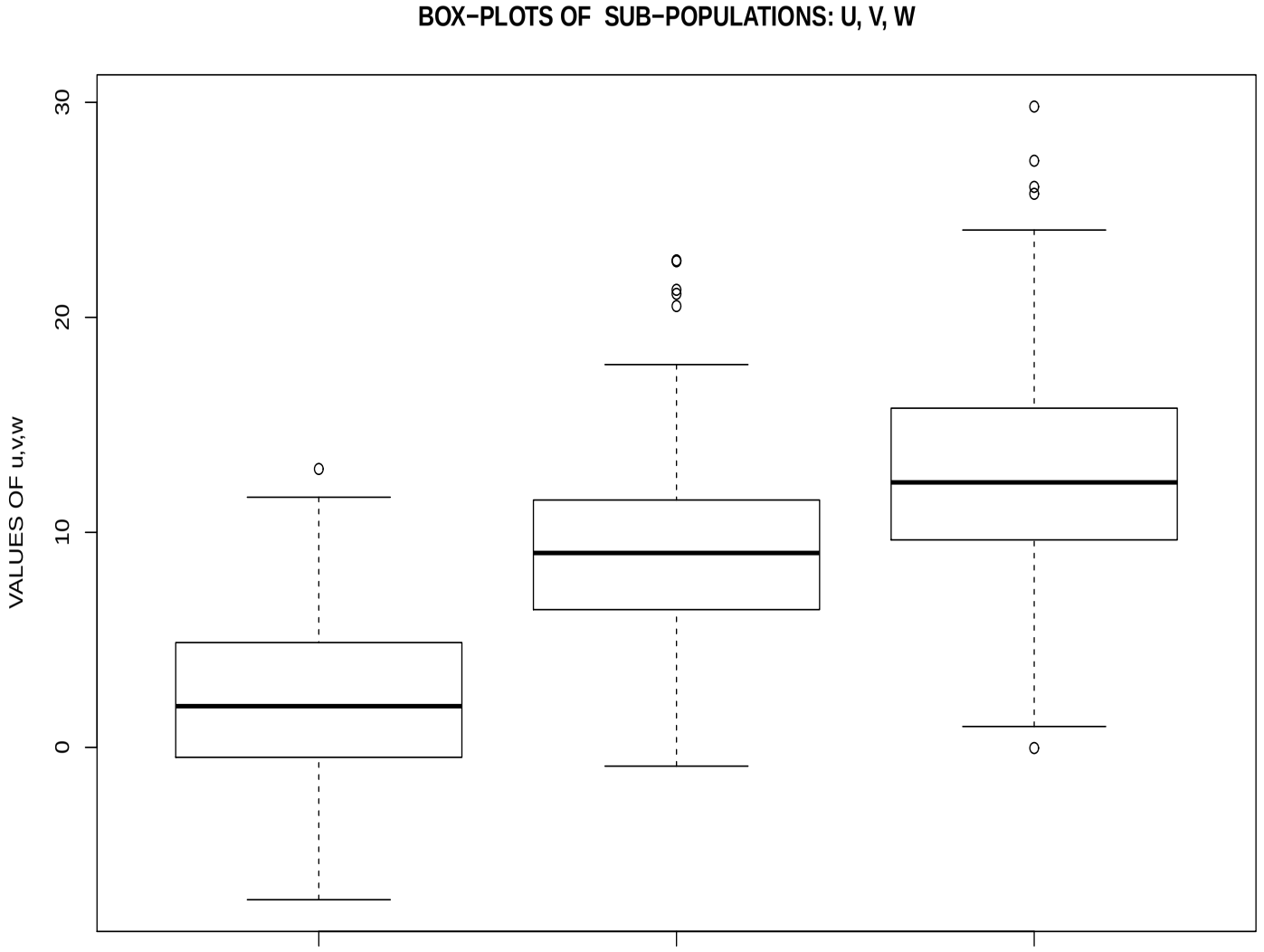

FIGURE 1
U=.2N(0,4)+.8N(3,16), V=.4N(8,9)+.6N(10,25), W=.6N(11,16)+.4N(14, 36)

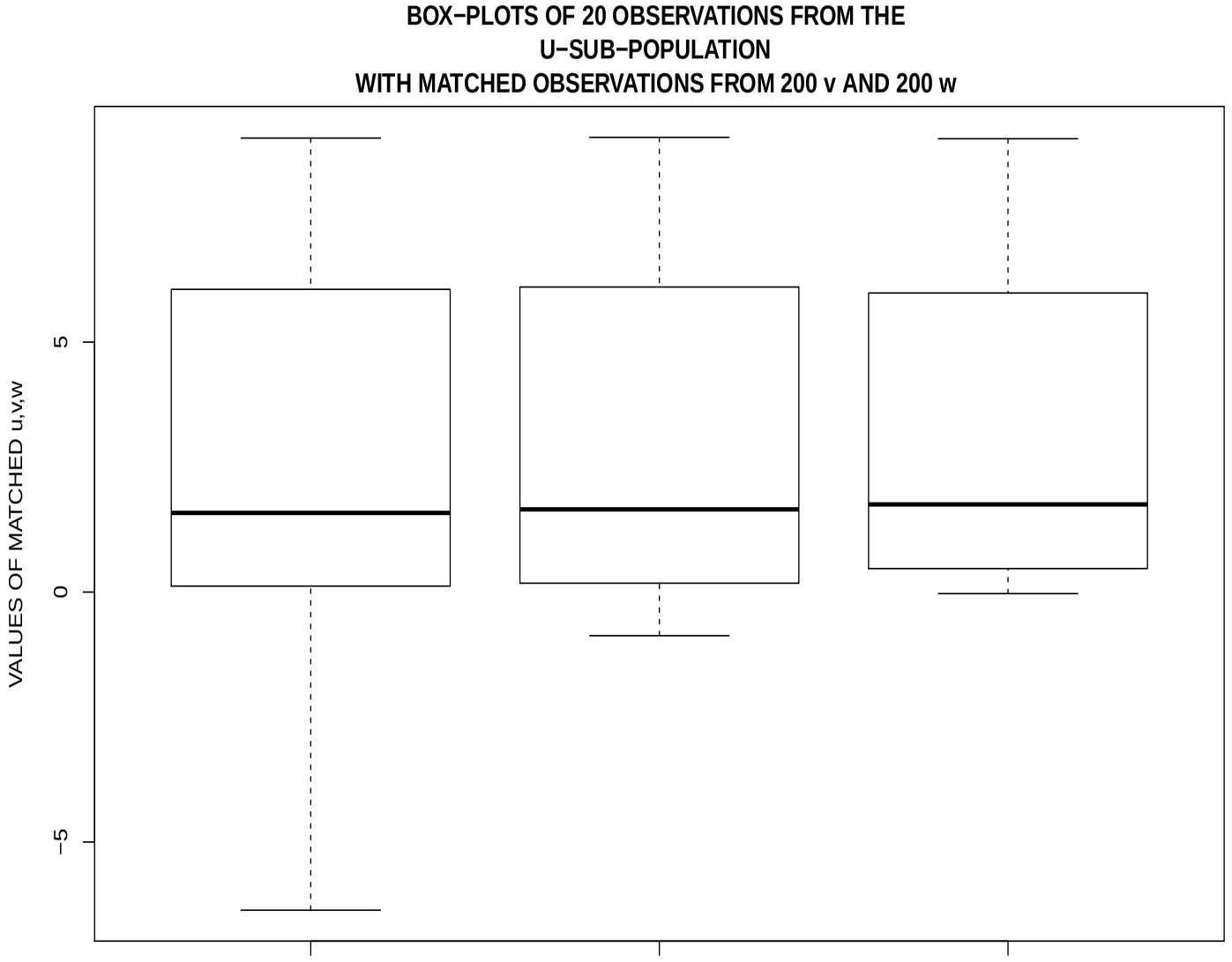
BOX-PLOTS OF 20 OBSERVATIONS FROM THE
U-SUB-POPULATION
WITH MATCHED OBSERVATIONS FROM 200 v AND 200 w
VALUES OF MATCHED u,v,w
5
0
-5

FIGURE 2

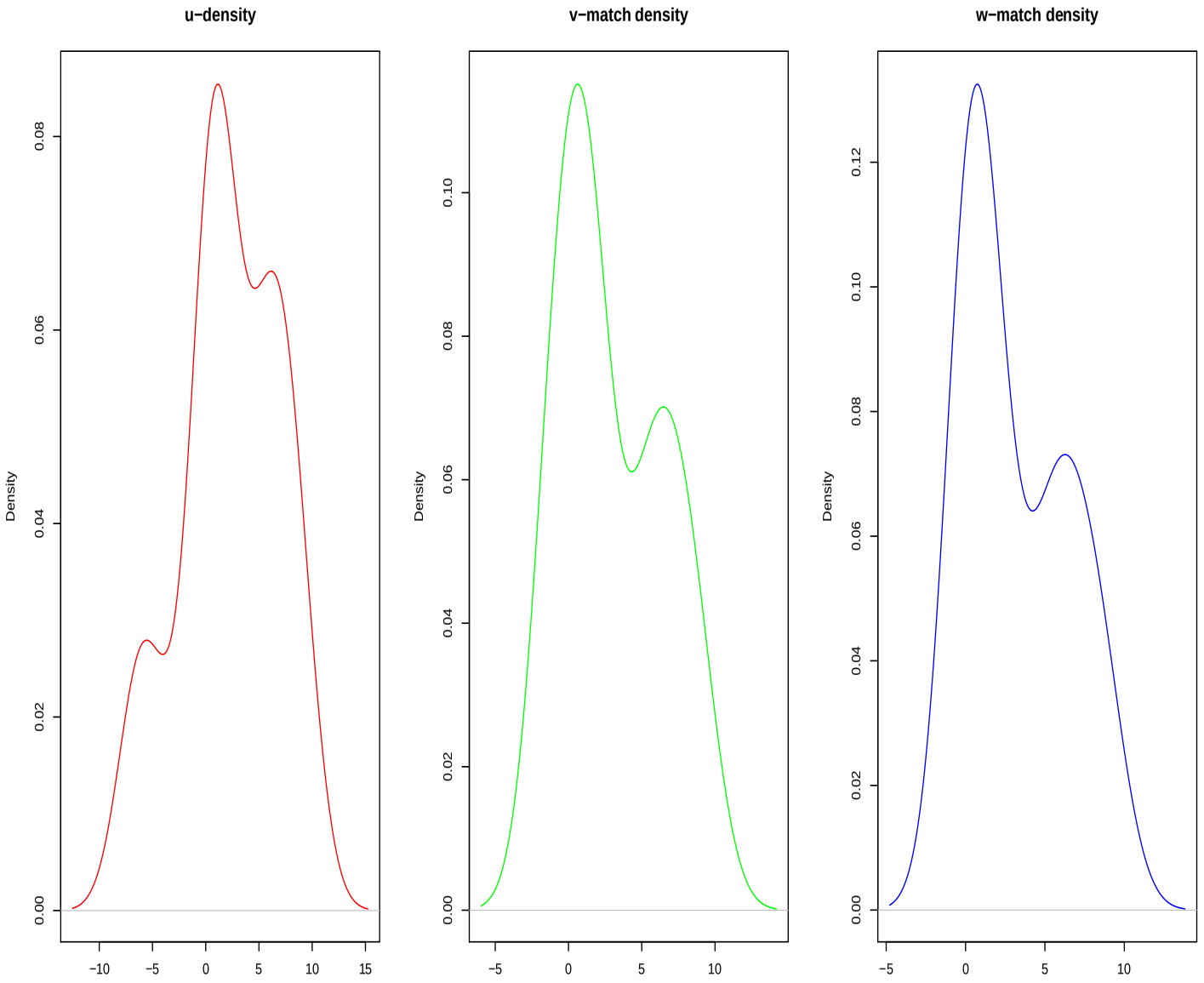
u−density
v−match density
w−match density
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
−10
−5
0
5
10
15

FIGURE 3

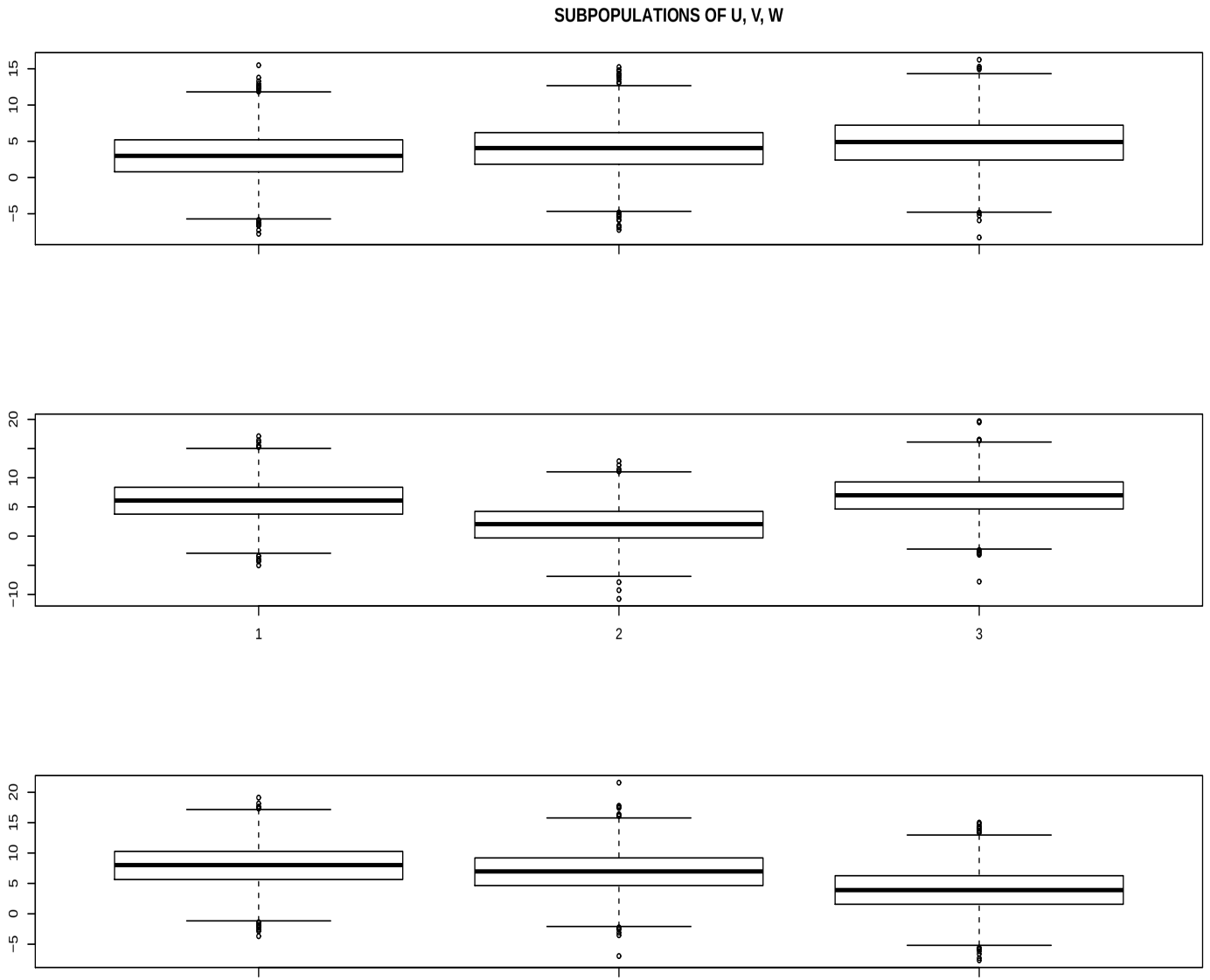

FIGURE 4

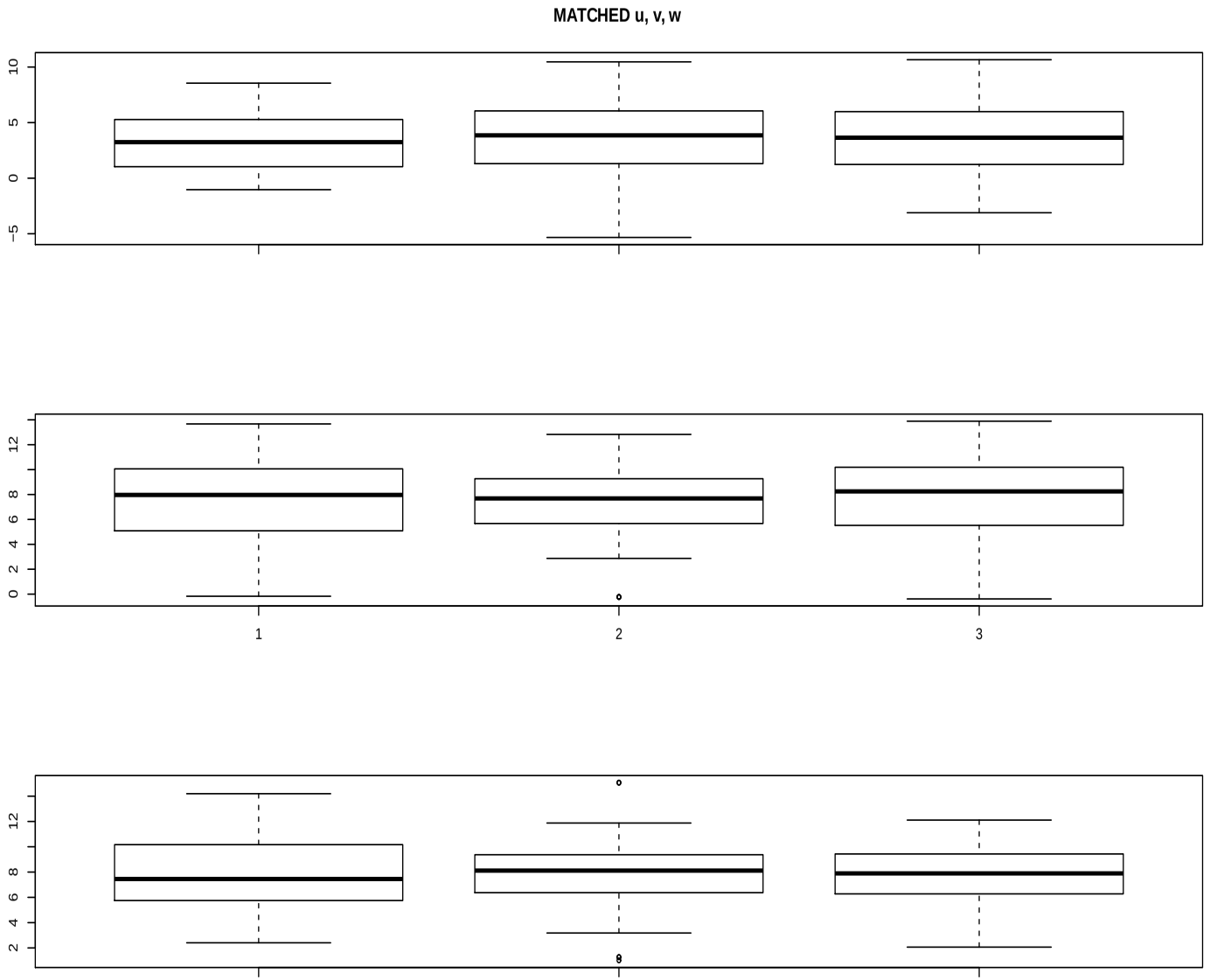

FIGURE 5

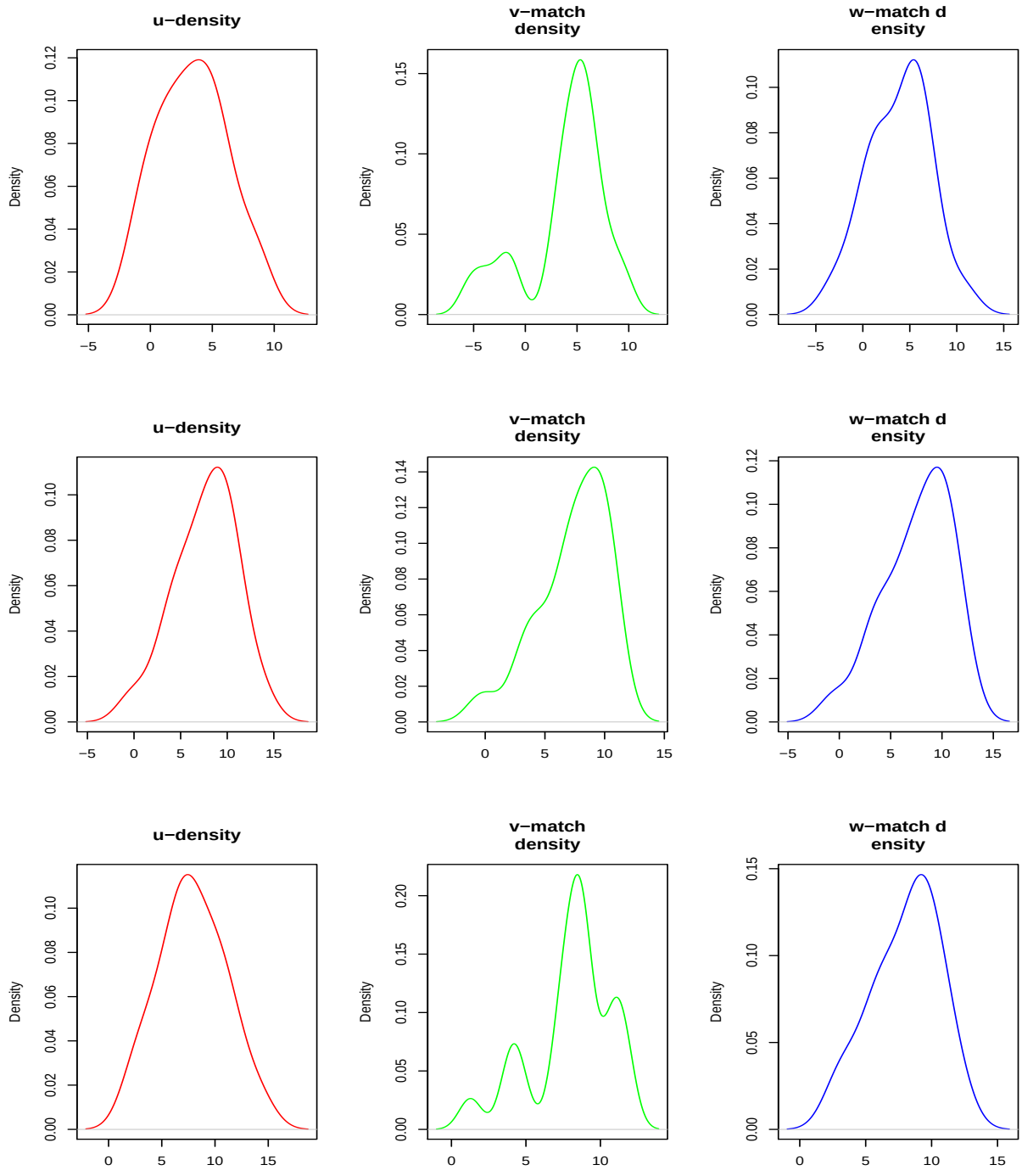

FIGURE 6

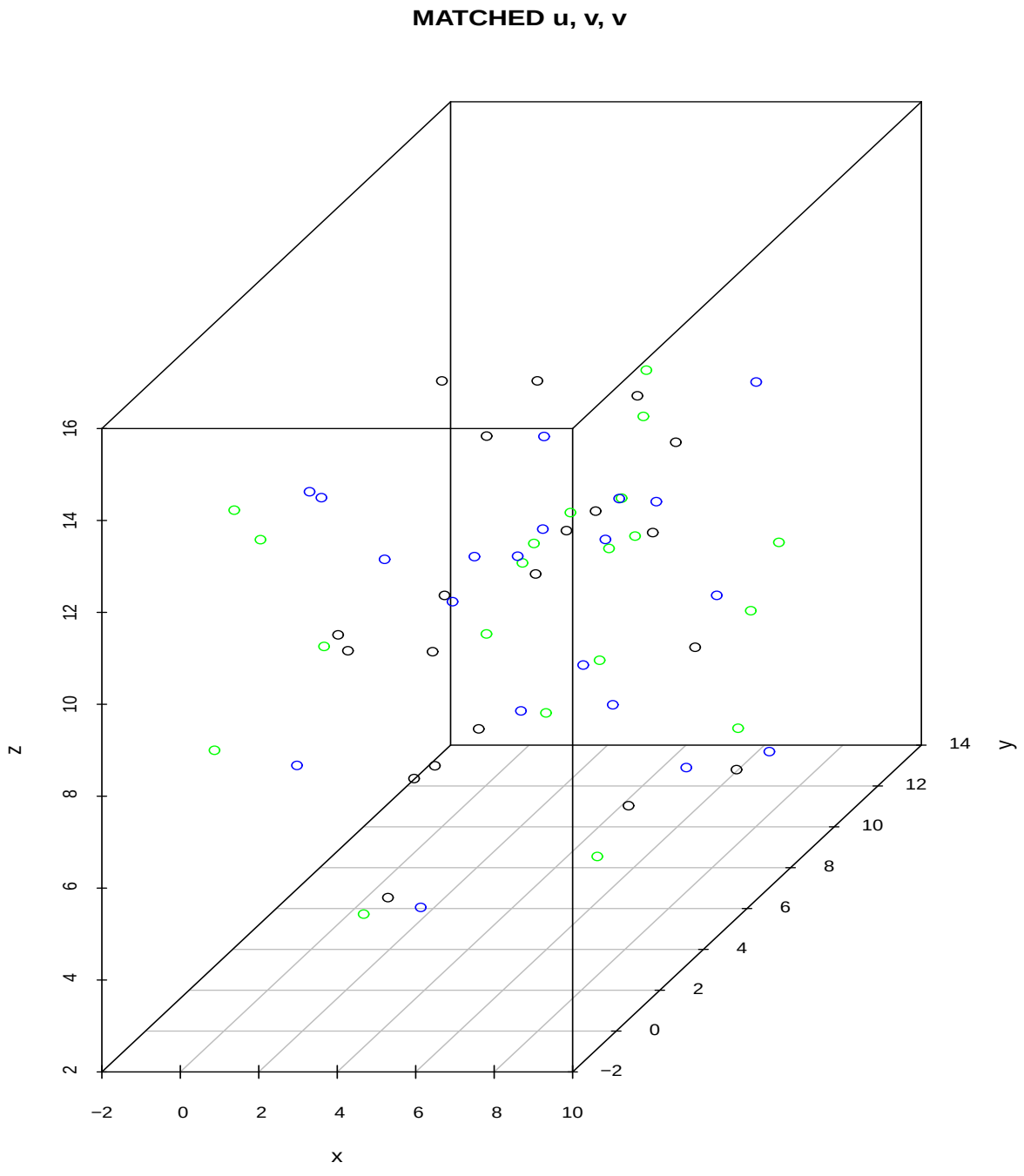

FIGURE 7