The hidden role of the propensity score in matching for inference about causal effects

Ben Hansen

Department of Statistics
University of Michigan, Ann Arbor
www.stat.lsa.umich.edu/~bbh/

Midwest Biopharmaceutical Statistics Workshop
Ball State U.; May 2011

Outline

Example:Coaching for the SAT

- Powers & Rock (1999, *J. Ed. Meas.*) asked a nationally representative sample of SAT takers: Did you receive for-pay coaching for the SAT?
- Asked background questions too. Responses linked to PSAT, SAT scores; demographics; . . . .

Propensity score model:

```
> ps.mod <- glm(Coach~PSAT-M+PSAT-V+ dadsed +
               GPA+ <...>, family=binomial)
> ps.hat <- ps.mod$linear.predictors
```

(A "linear" propensity score.)

Example:Coaching for the SAT

- Powers & Rock (1999, *J. Ed. Meas.*) asked a nationally representative sample of SAT takers: Did you receive for-pay coaching for the SAT?
- Asked background questions too. Responses linked to PSAT, SAT scores; demographics; ....

Propensity score model:

```
> ps.mod <- glm(Coach~PSAT-M+PSAT-V+ dadsed +
               GPA+ <...>, family=binomial)
> ps.hat <- ps.mod$linear.predictors
```

(A "linear" propensity score.)

Example:Coaching for the SAT

- Powers & Rock (1999, *J. Ed. Meas.*) asked a nationally representative sample of SAT takers: Did you receive for-pay coaching for the SAT?
- Asked background questions too. Responses linked to PSAT, SAT scores; demographics; . . . .

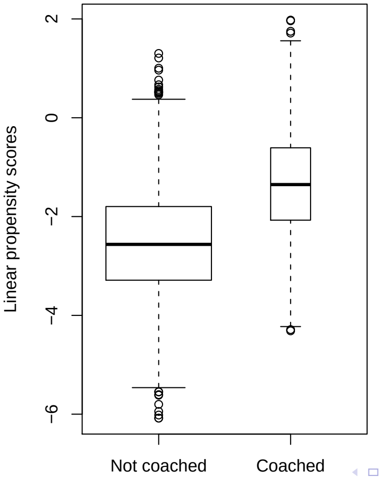
Propensity score model:

```
> ps.mod <- glm(Coach~PSAT-M+PSAT-V+ dadsed +
              GPA+ <...>, family=binomial)
> ps.hat <- ps.mod$linear.predictors
```

(A “linear” propensity score.)

The use of propensity scores to diagnose extrapolation

Comparing groups on $\hat{\varphi}(\mathbf{x})$, the linear combination of x's from logistic regression of z on $\mathbf{x}$



What is covariate balance?

...and how much of it do we need?

	No coach	Coached
Parents' income Q1	.27	.14
Parents' income Q2	.28	.2
Parents' income Q3	.15	.12
Parents' income Q4	.16	.35
Parents' income N/A	.14	.2
Dad's Ed.= high school	.43	.23
Dad's Ed.= some college	.26	.23
Dad's Ed.= grad school	.23	.42
Dad's Ed. N/A	.081	.12
Mom's Ed.= high school	.49	.29

⋮

	Not coached	Coached
Avg SAT at 1st ch. college	1059	1098
Avg SAT at 1st ch. college N/A	.36	.36
No previous score, or N/A	.32	.28
Previous score seemed fair	.22	.14
Previous score seemed unfair	.46	.58

Measuring imbalances in comparative studies

- Notation: **x**, v covariates; z= treatment indicator.
- Assume **x**, v fixed, unaffected by treatment. $Z = (Z_1', \ldots, Z_s')'$ a random vector, with $Z_s = (Z_{s1}, \ldots, Z_{sm_s})' \in \{0, 1\}^{m_s}$ indep. across matched sets s.
- Measure balance on v by:

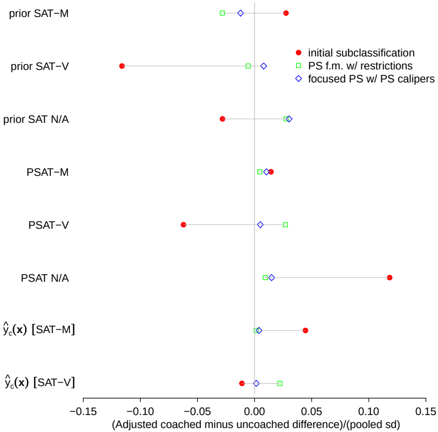
Definition

$\Delta_z[v] :=$ z-coeff. in `lm(v~z + s)`

- Without matching, $\Delta_z[\cdot] =$ diff of means. With matched pairs, $\Delta_z[\cdot] =$ mean paired difference.
- Write P for the permutation distribution[1].
- It's natural to compare $\Delta_z[v]$ to $\mathbf{Pr}_P(\Delta_Z[v] \in \cdot)$.

The use of propensity scores to reduce covariate imbalance

- Match on fitted PS. (Hansen (2011) gives details for these 2 variations.)
- Either match much improves balance on these and other variables, if it doesn't send imbalance to 0.



Measuring imbalances in comparative studies

- Notation: $\mathbf{x}$, v covariates; z= treatment indicator.
- Assume $\mathbf{x}$, v fixed, unaffected by treatment. $Z = (Z_1', \ldots, Z_S')'$ a random vector, with $Z_s = (Z_{s1}, \ldots, Z_{sm_s})' \in \{0, 1\}^{m_s}$ indep. across matched sets s.
- Measure balance on v by:

Definition

$\Delta_z[v] :=$ z-coeff. in `lm(v~z + s)`

- Without matching, $\Delta_z[\cdot] =$ diff of means. With matched pairs, $\Delta_z[\cdot] =$ mean paired difference.
- Write P for the permutation distribution[1].
- It's natural to compare $\Delta_z[v]$ to $\mathbf{Pr}_P(\Delta_Z[v] \in \cdot)$.

[1] I.e., $\mathbf{Pr}_P(Z = z) = \mathbf{Pr}_P(Z = w)$ whenever z, w are s.t. $\forall s$, z_s and w_s are the same up to re-sorting: $(z_{s[1]}, \ldots, z_{s[m_s]}) = (w_{s[1]}, \ldots, w_{s[m_s]})$. (Expressed in R/Splus, `table(z,s)` equals `table(w,s)`.)

The randomization distribution as a yardstick

	No coach	Coached	z	
Parents' income Q1	.27	.14	-6.4	***
Parents' income Q2	.28	.2	-4.1	***
Parents' income Q3	.15	.12	-1.7	.
Parents' income Q4	.16	.35	1.3	***
Parents' income N/A	.14	.2	3.5	***
Dad's Ed.= high school	.43	.23	-8.4	***
Dad's Ed.= some college	.26	.23	-1.7	.
Dad's Ed.= grad school	.23	.42	9.2	***
Dad's Ed. N/A	.081	.12	3.1	**
Mom's Ed.= high school	.49	.29	-8.5	***

⋮

	Not coached	Coached	z	
Avg SAT at 1st ch. college	1059	1098	9.5	***
Avg SAT at 1st ch. college N/A	.36	.36	.2	
No previous score, or N/A	.32	.28	-1.9	.
Previous score seemed fair	.22	.14	-3.8	***
Previous score seemed unfair	.46	.58	4.9	***

Were this an experiment, differences would be improbably large.

Strongly ignorable treatment assignment

Rosenbaum & Rubin (1983, *Biom'ka*). A.k.a. CIA, no unmeas. confounding, ...

$$(Y_t, Y_c) \perp Z | X;$$
$$0 < \mathbf{Pr}(Z = 1 | X) < 1$$

- If true, there is an experiment hiding in the observational study.
- Essential assumption for estimating intervention effects with observational data — whether or not we use propensity scores or matching.

Strongly ignorable treatment assignment

Rosenbaum & Rubin (1983, *Biom'ka*). A.k.a. CIA, no unmeas. confounding, ...

$$(Y_t, Y_c) \perp Z | X;$$

$$0 < \mathbf{Pr}(Z = 1|X) < 1$$

- If true, there is an experiment hiding in the observational study.
- Essential assumption for estimating intervention effects with observational data — whether or not we use propensity scores or matching.

Strongly ignorable treatment assignment

Rosenbaum & Rubin (1983, *Biom'ka*). A.k.a. CIA, no unmeas. confounding, ...

$$(Y_t, Y_c) \perp Z | X;$$

$$0 < \mathbf{Pr}(Z = 1 | X) < 1$$

- If true, there is an experiment hiding in the observational study.
- Essential assumption for estimating intervention effects with observational data — whether or not we use propensity scores or matching.

Strongly ignorable treatment assignment

Rosenbaum & Rubin (1983, *Biom'ka*). A.k.a. CIA, no unmeas. confounding, ...

$$(Y_t, Y_c) \perp Z | X;$$

$$0 < \mathbf{Pr}(Z = 1 | X) < 1$$

- If true, there is an experiment hiding in the observational study.
- Essential assumption for estimating intervention effects with observational data — whether or not we use propensity scores or matching.

Strongly ignorable treatment assignment

Rosenbaum & Rubin (1983, *Biom'ka*). A.k.a. CIA, no unmeas. confounding, ...

$$(Y_t, Y_c) \perp Z | X;$$
$$0 < \mathbf{Pr}(Z = 1 | X) < 1$$

- If true, there is an experiment hiding in the observational study.
- Essential assumption for estimating intervention effects with observational data — whether or not we use propensity scores or matching.

Strongly ignorable treatment assignment

Rosenbaum & Rubin (1983, *Biom'ka*). A.k.a. CIA, no unmeas. confounding, ...

$$(Y_t, Y_c) \perp Z | X;$$

$$0 < \mathbf{Pr}(Z = 1|X) < 1$$

- If true, there is an experiment hiding in the observational study.
- Essential assumption for estimating intervention effects with observational data — whether or not we use propensity scores or matching.

Outline

Asymptotic normality of $\Delta_Z[v]$ in experiments

For $v = (v_1', \ldots, v_n')'$ put
$\|v\|_2 = [n^{-1} \sum_{s=1}^{n} \sigma^2[v_s]]^{1/2} = [n^{-1} \sum_{s=1}^{n} m_s^{-1} \sum_{i=1}^{m_s} (v_{si} - \bar{v}_s)^2]^{1/2}$;
$\|v\|_\infty := \max_{s,i} |v_{si} - \bar{v}_s|$. Now suppose a sequence of v's, $\{v_{(n)}\}$, arranged in a triangular array.

Definition

The design distributes variation in v if $\|v_{(n)}\|_\infty / (n^{1/2} \|v_{(n)}\|_2) \to 0$ as $n \uparrow \infty$.

Proposition (CLT for matched experiments)

If the design distributes variation in v,

$$\frac{\Delta_Z[v_{(n)}]}{\mathbf{V}_P^{1/2}[\Delta_Z[v_{(n)}]]} \overset{P_n}{\to} N(0,1).$$

The situation after *inexact* matching on an *estimated* score

Preliminaries

Recall that for $z \in \mathcal{Z}$, $\mathbf{Pr}_P(Z = z) \propto 1$. Write φ for $\varphi(\mathbf{x})$ (the true PS).

Definition

For $z \in \mathcal{Z}$, $\mathbf{Pr}_P(Z = z) \propto 1$; $\mathbf{Pr}_Q(Z = z) \propto \exp\{z^t\varphi\}$.

In matched analysis one supposes that

no hidden bias $\Rightarrow \Delta_Z[y_c], \Delta_Z[y_t]$ have the P-distribution;

but matching arranges no more than

no hidden bias $\Rightarrow \Delta_Z[y_c], \Delta_Z[y_t]$ have the Q-distribution.

Now P and Q will be "close" if

$$\|\varphi\|_2^2 = \frac{1}{n}\sum_{s=1}^{n} \frac{\sum_{i=1}^{m_s}(\varphi_{si} - \bar{\varphi}_s)^2}{m_s} \approx 0.$$

Is it possible to tell when this is so?

The situation after *inexact* matching on an *estimated* score

Preliminaries

Recall that for $z \in \mathcal{Z}$, $\mathbf{Pr}_P(Z = z) \propto 1$. Write φ for $\varphi(\mathbf{x})$ (the true PS).

Definition

For $z \in \mathcal{Z}$, $\mathbf{Pr}_P(Z = z) \propto 1$; $\mathbf{Pr}_Q(Z = z) \propto \exp\{z^t\varphi\}$.

In matched analysis one supposes that

$$\text{no hidden bias} \Rightarrow \Delta_Z[y_c], \Delta_Z[y_t] \text{ have the } P\text{-distribution;}$$

but matching arranges no more than

$$\text{no hidden bias} \Rightarrow \Delta_Z[y_c], \Delta_Z[y_t] \text{ have the } Q\text{-distribution.}$$

Now P and Q will be "close" if

$$\|\varphi\|_2^2 = \frac{1}{n}\sum_{s=1}^{n} \frac{\sum_{i=1}^{m_s}(\varphi_{si} - \bar{\varphi}_s)^2}{m_s} \approx 0.$$

Is it possible to tell when this is so?

The situation after *inexact* matching on an *estimated* score

Preliminaries

Recall that for $z \in \mathcal{Z}$, $\mathbf{Pr}_P(Z = z) \propto 1$. Write φ for $\varphi(\mathbf{x})$ (the true PS).

Definition

For $z \in \mathcal{Z}$, $\mathbf{Pr}_P(Z = z) \propto 1$; $\mathbf{Pr}_Q(Z = z) \propto \exp\{z^t \varphi\}$.

In matched analysis one supposes that

$$\text{no hidden bias} \Rightarrow \Delta_Z[y_c], \Delta_Z[y_t] \text{ have the } P\text{-distribution;}$$

but matching arranges no more than

$$\text{no hidden bias} \Rightarrow \Delta_Z[y_c], \Delta_Z[y_t] \text{ have the } Q\text{-distribution.}$$

Now P and Q will be “close” if

$$\|\varphi\|_2^2 = \frac{1}{n} \sum_{s=1}^{n} \frac{\sum_{i=1}^{m_s} (\varphi_{si} - \bar{\varphi}_s)^2}{m_s} \approx 0.$$

Is it possible to tell when this is so?

Crude balance

Write $\rho_{v,w}$ for $\langle v, w\rangle/(\|v\|_2\|w\|_2)$, the matched correlation of v and w. Recall that (barring designs which do not distribute variation in v)

$$\frac{\Delta_Z[v]}{\mathbf{V}_P^{1/2}[\Delta_Z[v]]} \xrightarrow{P} N(0,1).$$

Proposition

Assume that matching distributes variation in both φ and x. If $\rho_{x,\varphi} \to \rho$ and $n^{1/2}\|\varphi\|_2 \to$ some limit in $[0, \infty]$ then

$$\lim_n \mathbf{E}_{Q_n}\left[\frac{\Delta_Z[x]}{\mathbf{V}_P^{1/2}[\Delta_Z[x]]}\right] = \rho \lim_n n^{1/2}\|\varphi\|_2.$$

The proposition says that in large samples, large values of $n^{1/2}\|\varphi\|_2$ are likely to manifest themselves in terms of large covariate imbalances.

Definition

Matching is *crudely balanced* if $\lim_n n^{1/2}\|\varphi\|_2 < \infty$.

The randomization distribution as a yardstick

	No coach	Coached	z	
Parents' income Q1	.27	.14	-6.4	***
Parents' income Q2	.28	.2	-4.1	***
Parents' income Q3	.15	.12	-1.7	.
Parents' income Q4	.16	.35	1.3	***
Parents' income N/A	.14	.2	3.5	***
Dad's Ed.= high school	.43	.23	-8.4	***
Dad's Ed.= some college	.26	.23	-1.7	.
Dad's Ed.= grad school	.23	.42	9.2	***
Dad's Ed. N/A	.081	.12	3.1	**
Mom's Ed.= high school	.49	.29	-8.5	***

⋮

	Not coached	Coached	z	
Avg SAT at 1st ch. college	1059	1098	9.5	***
Avg SAT at 1st ch. college N/A	.36	.36	.2	
No previous score, or N/A	.32	.28	-1.9	.
Previous score seemed fair	.22	.14	-3.8	***
Previous score seemed unfair	.46	.58	4.9	***

Combining z-statistics into a χ^2 addresses multiple comparisons (Hansen & Bowers 2008, *Stat. Sci.*). Without matching, $\chi^2 = 486$ on 66 d.f.; $p \approx 0$. (After full matching, $\chi^2/\mathrm{df} = 79/69$; $p = .2$.)

Contiguity (i)

Proposition (Contiguity of P and Q,(i))

Assume:

- matching is crudely balanced, with $n^{1/2}\|\varphi\| \to \sigma_{\varphi}$;
- matching distributes variation in y_c, φ; and
- the comparison is **x**-adjustable for y_c.

If $\rho_{(n)y_c\varphi} = \langle y_c, \varphi\rangle/(\|y_c\|_2\|\varphi\|_2) \to \rho_{y_c\varphi}$, then

$$\begin{aligned} \Delta_Z[y_c]/\mathbf{V}_P^{1/2}(\Delta_Z[y_c]) &\overset{P_n}{\to} N(0,1), \text{ whereas} \\ \Delta_Z[y_c]/\mathbf{V}_P^{1/2}(\Delta_Z[y_c]) &\overset{Q_n}{\to} N(\sigma_{\varphi}\rho_{y_c\varphi}, 1). \end{aligned}$$

Outline

Consistency

Proposition

Assume:

- matching distributes variation in y_t, y_c, $y_t - y_c$, and φ;
- matching is crudely balanced, with $n^{1/2}\|\varphi\| \to \sigma_\varphi$; and
- the comparison is **x**-adjustable for y_t, y_c.

Then

$$\Delta_Z[Y_{(n)}] \overset{Q_n}{\to} \theta$$

$$=: \lim \left(\sum_s \frac{m_s - 1}{m_s} (\bar{y}_{ts} - \bar{y}_{csi}) \right) / \sum_s \frac{m_s - 1}{m_s}$$

(assuming the limit to exist).

(Here $Y_{(n)} = Z^t_{(n)} y_{(n)t} + (1 - Z_{(n)})^t y_{(n)c}$ is the vector of observed responses.)

Approximating an expt. requires matching on the PS

Tests of hypothesis that coaching doesn't effect verbal SAT scores, evaluated with P and with 3rd-order Edgeworth approximations to candidate Q's

Mahalanobis pair matching, no propensity ($\sqrt{\chi^2/\mathrm{df}} = 1.6$; $p < 10^{-10}$.)

$\Delta_Z[\mathrm{SATV}]$ under…	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	$3°$
P	0	0	22	0	$-.022$	.01	.01
$\hat{Q}_{\mathrm{mle}}$	2.9	2.2	15	$-.039$	$-.005$	.33	.34
$\hat{Q}_1$	2.8	2.0	15	$-.037$	$-.007$	.25	.25
$\hat{Q}_2$	3.0	2.0	14	$-.032$	$-.005$	.25	.25

Mahalanobis including PS + PS caliper ($\sqrt{\chi^2/\mathrm{df}} = 1.2$; $p = .02$.)

$\Delta_Z[\mathrm{SATV}]$ under…	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	$3°$
P	0	0	23	0	$-.02$	.37	.37
$\hat{Q}_{\mathrm{mle}}$	.01	.02	23	$-.002$	$-.02$	.37	.37
$\hat{Q}_1$	.14	.15	21	$-.004$	$-.02$	.36	.36
$\hat{Q}_2$	.25	.30	22	$-.013$	$-.02$	.38	.38

Approximating an expt. requires matching on the PS

Tests of hypothesis that coaching doesn't effect verbal SAT scores, evaluated with *P* and with 3rd-order Edgeworth approximations to candidate *Q*'s

Mahalanobis pair matching, no propensity ($\sqrt{\chi^2/\mathrm{df}} = 1.6$; $p < 10^{-10}$.)

Δ_Z[SATV] under...	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3°
P	0	0	22	0	$-.022$	.01	.01
$\hat{Q}_{\mathrm{mle}}$	2.9	2.2	15	$-.039$	$-.005$	.33	.34
$\hat{Q}_1$	2.8	2.0	15	$-.037$	$-.007$	.25	.25
$\hat{Q}_2$	3.0	2.0	14	$-.032$	$-.005$	.25	.25

Mahalanobis including PS + PS caliper ($\sqrt{\chi^2/\mathrm{df}} = 1.2$; $p = .02$.)

Δ_Z[SATV] under...	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3°
P	0	0	23	0	$-.02$	.37	.37
$\hat{Q}_{\mathrm{mle}}$	.01	.02	23	$-.002$	$-.02$	.37	.37
$\hat{Q}_1$	.14	.15	21	$-.004$	$-.02$	.36	.36
$\hat{Q}_2$	.25	.30	22	$-.013$	$-.02$	.38	.38

Approximating an expt. requires matching on the PS

Tests of hypothesis that coaching doesn't effect verbal SAT scores, evaluated with P and with 3rd-order Edgeworth approximations to candidate Q's

Mahalanobis pair matching, no propensity ($\sqrt{\chi^2/\mathrm{df}} = 1.6$; $p < 10^{-10}$.)

Δ_Z[SATV] under...	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3°
P	0	0	22	0	$-.022$	.01	.01
$\hat{Q}_{\mathrm{mle}}$	2.9	2.2	15	$-.039$	$-.005$	.33	.34
$\hat{Q}_1$	2.8	2.0	15	$-.037$	$-.007$	.25	.25
$\hat{Q}_2$	3.0	2.0	14	$-.032$	$-.005$	.25	.25

Mahalanobis including PS + PS caliper ($\sqrt{\chi^2/\mathrm{df}} = 1.2$; $p = .02$.)

Δ_Z[SATV] under...	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3°
P	0	0	23	0	$-.02$	.37	.37
$\hat{Q}_{\mathrm{mle}}$	.01	.02	23	$-.002$	$-.02$	.37	.37
$\hat{Q}_1$	.14	.15	21	$-.004$	$-.02$	.36	.36
$\hat{Q}_2$	.25	.30	22	$-.013$	$-.02$	.38	.38

Approximating an expt. requires matching on the PS

Tests of hypothesis that coaching doesn't effect verbal SAT scores, evaluated with P and with 3rd-order Edgeworth approximations to candidate Q's

Mahalanobis pair matching, no propensity ($\sqrt{\chi^2/\mathrm{df}} = 1.6$; $p < 10^{-10}$.)

Δ_Z[SATV] under...	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3°
P	0	0	22	0	$-.022$	.01	.01
$\hat{Q}_{\mathrm{mle}}$	2.9	2.2	15	$-.039$	$-.005$	.33	.34
$\hat{Q}_1$	2.8	2.0	15	$-.037$	$-.007$	.25	.25
$\hat{Q}_2$	3.0	2.0	14	$-.032$	$-.005$	.25	.25

Mahalanobis including PS + PS caliper ($\sqrt{\chi^2/\mathrm{df}} = 1.2$; $p = .02$.)

Δ_Z[SATV] under...	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3°
P	0	0	23	0	$-.02$	.37	.37
$\hat{Q}_{\mathrm{mle}}$	.01	.02	23	$-.002$	$-.02$	.37	.37
$\hat{Q}_1$	.14	.15	21	$-.004$	$-.02$	.36	.36
$\hat{Q}_2$	.25	.30	22	$-.013$	$-.02$	.38	.38

P-approximation a bit worse in tails, but workable.

Tests of hypothesis that coaching doesn't affect math SAT scores, under P and candidate Q's.

Mahalanobis including PS + PS caliper (Crudely balanced.)

Δ_Z[SATM] under...	Bias/ $\approx$	$V_P^{1/2}$ $=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3° E-wth
P	0	0	24	0	−.018	5.7e-09	1.0e-09
$\hat{Q}_{\text{mle}}$	.5	.6	24	−.009	−.018	6.3e-08	1.7e-08
$\hat{Q}_1$	1.1	1.2	22	−.014	−.016	5.6e-07	2.3e-07
$\hat{Q}_2$	.5	.6	23	−.008	−.017	2.4e-08	5.6e-09

P-approximation a bit worse in tails, but workable.

Tests of hypothesis that coaching doesn't affect math SAT scores, under P and candidate Q's.

Mahalanobis including PS + PS caliper (Crudely balanced.)

$\Delta_Z[\mathrm{SATM}]$ under…	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3° E-wth
P	0	0	24	0	$-.018$	5.7e-09	1.0e-09
$\hat{Q}_{\mathrm{mle}}$	.5	.6	24	$-.009$	$-.018$	6.3e-08	1.7e-08
$\hat{Q}_1$	1.1	1.2	22	$-.014$	$-.016$	5.6e-07	2.3e-07
$\hat{Q}_2$	.5	.6	23	$-.008$	$-.017$	2.4e-08	5.6e-09

Conclusions

- Even without unmeasured confounders or mistakes in matching, the propensity matched observational study and the analogous matched experiment are *not* precisely the same. But the differences admit a neat characterization.
- The global χ^2 test of covariate balance distinguishes designs which do and don't attain crude balance.
- Simple & reliable robustness check for studies combining PS matching with model-based covariance adjustment.
- The contiguity approximation also sheds light on advantages of better-than-crude balance; but that's a topic for another day.

Hansen, B. B. (2008), "The essential role of balance tests in propensity-matched observational studies: Comments on "A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003" by Peter Austin, Statistics in Medicine." *Statist. Med.*, 27, 2050–2054.

— (2009), "Propensity score matching to recover latent experiments: diagnostics and asymptotics," Tech. Rep. 486, Statistics Department, University of Michigan.

— (2011), "Propensity score matching to extract latent experiments from nonexperimental data: A case study," in *Looking Back: Proceedings of a Conference in Honor of Paul W. Holland*, eds. Dorans, N. and Sinharay, S., Springer, to appear.

Hansen, B. B. and Bowers, J. (2008), "Covariate balance in simple, stratified and clustered

Outline

Outline

P-approximation a bit worse in tails, but workable.

Tests of hypothesis that coaching doesn't affect math SAT scores, under P and candidate Q's.

Mahalanobis including PS + PS caliper (Crudely balanced.)

$\Delta_Z[\mathrm{SATM}]$ under…	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3° E-wth
P	0	0	24	0	−.018	5.7e-09	1.0e-09
$\hat{Q}_{\mathrm{mle}}$	.5	.6	24	−.009	−.018	6.3e-08	1.7e-08
$\hat{Q}_1$	1.1	1.2	22	−.014	−.016	5.6e-07	2.3e-07
$\hat{Q}_2$	.5	.6	23	−.008	−.017	2.4e-08	5.6e-09

Mahalanobis with PS+ cov. adjustment as in Rosenbaum (2002 *Stat. Sci.*)

$\Delta_Z[e_{\mathrm{SATM}}]$ under…	$\mathrm{Bias}/V_P^{1/2}$ $\approx$	$=$	$V_{(\cdot)}$ $=$	Skew. $=$	Kurt. $=$	2-sided p CLT	3° E-wth
P	0	0	15	0	−.016	.015	.015
$\hat{Q}_{\mathrm{mle}}$	.01	.02	14	.001	−.015	.014	.014
$\hat{Q}_1$	−.01	−.02	13	.005	−.014	.011	.011
$\hat{Q}_2$	.01	.05	14	−.001	−.015	.012	.012

Asymptotic conditional matched adjustability

Assume $\mathbf{x}$-adjustability for y_c. Assume crude balance; distributed variation in $\mathbf{x}$, y_c and their linear combinations. Let $(Y, \mathbf{X}, \Phi) = \mathrm{plim}(n^{1/2}\Delta_Z[y_c], n^{1/2}\Delta_Z[\mathbf{x}], Z^t\varphi - \mathbf{E}_P Z^t\varphi)$. The ▸ MV analogue of the previous proposition says that:

$$
\begin{aligned}
Y|\mathbf{X} &\overset{P}{\sim} N(\Sigma_{y,\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}\mathbf{X}^t, \sigma_y^2 - \Sigma_{y,\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}\Sigma_{\mathbf{x},y});\ \text{whereas} \\
Y|\mathbf{X} &\overset{Q}{\sim} N(\tilde{\sigma}_{y,\varphi} + \Sigma_{y,\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}(\mathbf{X} - \tilde{\Sigma}_{\varphi,\mathbf{x}})^t, \sigma_y^2 - \Sigma_{y,\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}\Sigma_{\mathbf{x},y}) \\
&= N\left(\begin{array}{c} \underbrace{\tilde{\sigma}_{y,\varphi} - \Sigma_{y,\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}\tilde{\Sigma}_{\mathbf{x},\varphi}}_{*} + \\ \Sigma_{y,\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}\mathbf{X}^t, \quad \sigma_y^2 - \Sigma_{y,\mathbf{x}}\Sigma_{\mathbf{x}}^{-1}\Sigma_{\mathbf{x},y} \end{array} \right).
\end{aligned}
$$

But (*) can be represented as $\mathbf{Cov}_P(Y - \mathbf{E}_P(Y|\mathbf{X}), \Phi)$, the limit of $\langle y_c - \mathrm{Pj}(y_c|\mathbf{x}), n^{1/2}\varphi\rangle$. If $\varphi \in \mathrm{lin.span}(\mathbf{x})$, then (*)=0 and

$$\mathbf{Pr}_Q(Y \in \cdot|\mathbf{X}) = \mathbf{Pr}_P(Y \in \cdot|\mathbf{X}).$$

Contiguity (ii)

Proposition (Contiguity of P and Q,(ii))

Assume **x**-adjustability for **v**; matching distributes variation in φ and $\mathbf{v}\beta$, all β; matching is crudely balanced, with $n^{1/2}\|\varphi\| \to \sigma_\varphi$. Further,

$$\begin{aligned} \Sigma_{(n)\mathbf{v}} &= [\langle v^{(i)}_{(n)}, v^{(j)}_{(n)}\rangle : i,j \leq k] \to \Sigma_{\mathbf{v}} \text{ and} \\ \tilde{\Sigma}_{(n)\varphi\mathbf{v}} &= [\langle n^{1/2}\varphi, v^{(k)}_{(n)}\rangle : k] \to \tilde{\Sigma}_{\varphi\mathbf{v}} \end{aligned}$$

$\Sigma_{\mathbf{v}}$ a nonnegative-definite $k \times k$ matrix. Then

$$\begin{aligned} \left[\frac{\Delta_Z[\mathbf{v}]}{n^{1/2}}, Z^t\varphi - \mathbf{E}_{P_n} Z^t\varphi\right] &\stackrel{P_n}{\to} N\left([\mathbf{0}, 0], \begin{pmatrix} \Sigma_{\mathbf{v}} & \tilde{\Sigma}^t_{\varphi\mathbf{v}} \\ \tilde{\Sigma}_{\varphi\mathbf{v}} & \sigma^2_\varphi \end{pmatrix}\right) \\ \left[\frac{\Delta_Z[\mathbf{v}]}{n^{1/2}}, Z^t\varphi - \mathbf{E}_{P_n} Z^t\varphi\right] &\stackrel{Q_n}{\to} N\left(\left[\tilde{\Sigma}_{\varphi\mathbf{v}}, \sigma^2_\varphi\right], \begin{pmatrix} \Sigma_{\mathbf{v}} & \tilde{\Sigma}^t_{\varphi\mathbf{v}} \\ \tilde{\Sigma}_{\varphi\mathbf{v}} & \sigma^2_\varphi \end{pmatrix}\right). \end{aligned}$$

(▸ Return to "Asymptotic conditional matched adjustability.")