

#	質問	回答
1	受講者から質問： 共変量が階層的（DAG）になるような場合は、回帰式を多段に作れば良いのでしょうか？	高橋先生の回答：共変量の関係が階層的になっているのであればそれでよいと思います。
2	受講者から質問： 自然科学分野での因果推論の活用場面など具体例あれば教えていただきたい。	高橋先生：ワクチンの例などは自然科学分野に該当するかと思います。（私自身は自然科学分野の研究者ではないためあまり例が出せません） 疫学の分野では因果推論が盛んに研究研究されているため、Hernan & Robinsなどを参照していただくのが良いと思います。
2	受講者から質問： 自然科学分野での因果推論の活用場面など具体例あれば教えていただきたい。	補足：Hernan & Robinsの著書「Causal Inference: What If」 https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
3	受講者から質問： 多重共線性があると、説明変数の分散共分散行列の逆行列が不安定（分母となる行列式が0近く）になり、得られる回帰式の1次係数が不安定になると思いますが、如何でしょうか。	高橋先生の回答：講義および書籍でも述べている通り、多重共線性があると不安定になって標準誤差が大きくなるというのがまさにこのことです。
4	受講者から質問： zoomに入れず、講義2の表2.2について、反実仮定の期末試験の点数の計算方法を、聴きもしました。現実の得点をもとにATTで調整するのでしょうか。Rのライブラリがあるのでしょうか。	高橋先生：昨日のQ&Aで回答済みだと思いますのでそちらをご参照ください。
4	受講者から質問： zoomに入れず、講義2の表2.2について、反実仮定の期末試験の点数の計算方法を、聴きもしました。現実の得点をもとにATTで調整するのでしょうか。Rのライブラリがあるのでしょうか。	質問：表2.2において、実際には観測されない潜在的結果変数（白抜きの値）として、どのような値を入れるかは、どうやって決めるのでしょうか？この点については、ご著書『欠測データ解析』を参照すればよいですか？ 回答：潜在的結果変数は実際には観測されないもので、どのように補っていくかをこれから2日間で学んでいきます。
5	受講者から質問： 一日目に関係しておらず大変恐縮なのですが、御著『欠測データ処理』では言及されていない、多重代入データの多変量解析(例えば因子分析や主成分分析)や、限定的な言及に留まるカテゴリカル変数の分析(二項ロジット以外の分析、例えば多項ロジットや順序ロジットなど)をRで行う際、(私の勉強不足も相まって)なかなかコードを書くのに苦労しております。これらの多重代入法のアドバンスドな話題に関して、もし講義の時間に余裕がございましたら、最近の研究の動向も踏まえて言及して頂けると幸甚です。また、英語のものでも良いので、何か参考になる書籍や研究などがあればご紹介頂きたいです。	高橋先生の回答：今回のセミナーでは多重代入法を扱う時間はありませんので、申し訳ないですが、別の機会にセミナーができればよいかなと思っています。因子分析や主成分分析に多重代入法を使うことは不可能ではありませんが、多重代入法は因果推論との相性がよく、予測や分類のためのモデルに多重代入法を使うべきかどうかはまだ議論のあるところです。

6	<p>受講者から質問：</p> <p>多重共線性の問題について、ご著書のp.109では基本的に回帰係数の推定自体は不偏に行うことが出来ており、問題は標準誤差が大きくなることだと述べられています。その一方、共変量における多重共線性の数値例として、p.112では多重共線性の影響で共変量の係数が正しく推定できていない例が示されています。そうすると、因果効果を知りたい変数X1と「完全にはないがかなり強く関連している交絡因子X2」があるとき、X2を投入すると多重共線性の影響で変数X1の係数の推定が正しくなくなる危険もあると理解しましたが、その危険を避けてX2を投入しない場合と投入した場合のX1の係数とのどちらが良いという判断基準はありますでしょうか。</p>	<p>高橋先生の回答：これは難しい判断になると思います。先ほどの多重共線性の話と同じで、多重共線性があっても点推定値の不偏性は担保されます。つまり、母集団からのサンプリングを繰り返したときに計算したβ hatの期待値が母数に一致するということです。多重共線性が発生していると、標準誤差が大きくなるため、1回のサンプリングで得られた推定値が真値から離れた値になってしまう可能性が高くなります。これはどうしようもないことで、多重共線性の問題は小サンプルの問題と考えてよいと思います。</p>
6	<p>受講者から質問：</p> <p>多重共線性の問題について、ご著書のp.109では基本的に回帰係数の推定自体は不偏に行うことが出来ており、問題は標準誤差が大きくなることだと述べられています。その一方、共変量における多重共線性の数値例として、p.112では多重共線性の影響で共変量の係数が正しく推定できていない例が示されています。そうすると、因果効果を知りたい変数X1と「完全にはないがかなり強く関連している交絡因子X2」があるとき、X2を投入すると多重共線性の影響で変数X1の係数の推定が正しくなくなる危険もあると理解しましたが、その危険を避けてX2を投入しない場合と投入した場合のX1の係数とのどちらが良いという判断基準はありますでしょうか。</p>	<p>続き：どうしても多重共線性が気になるのであれば、X_2を含むモデルと含まないモデルの結果を両方報告して丁寧に議論する手もあります。</p>
7	<p>受講者から質問：</p> <p>重回帰分析の仮定2の線形性に関して、線形でない場合はGLM（一般化線形モデル）を使用して分析すると聞いたことがあるのですが、変数変換を行い最小二乗法を適用することとの違いや使い分けがありましたらご教示いただけますと幸いです。</p>	<p>高橋先生：書籍のp96の7.2.5項に変数の線形性と母数の線形性の違いについて議論しています。昨日の内容は変数に関して非線形のケースを提示しました。一方、パラメタが非線形の場合には変数変換では対応できないため、GLMなどを用いることになります。「非線形」に2つの意味があるということをご理解いただければと思います。</p>
9	<p>傾向スコアは、状態空間モデルや、VARモデルのようや時系列のモデルでも応用可能でしょうか(よく研究されていますでしょうか?)</p>	<p>高橋先生：私はその部分はあまりよく把握していませんが、理屈の上では時系列のモデルでも応用可能だとは思いますが、そのような研究がよくなされているかは分かりません。</p>
9	<p>傾向スコアは、状態空間モデルや、VARモデルのようや時系列のモデルでも応用可能でしょうか(よく研究されていますでしょうか?)</p>	<p>横から失礼します。時系列モデル自体はわかりませんが、パネルデータ（time-series--cross-sectional data）であれば、Kosuke Imai他の"Matching Methods for Causal Inference with Time-Series Cross-Sectional Data"があります。</p>
10	<p>傾向スコアを使うには、処置群と統制群のp(X)の台が一致していないといけない認識なのですが、実データで台が一致していそうかはどうやって判別すればよいのでしょうか？</p>	<p>高橋先生の回答：傾向スコアを算出した後にオーバーラップ条件を確認する方法はあります。口頭で答えるのは難しいですが、グラフを描いて実際に調べるといったことを実データ解析で行います。</p>

11	傾向スコア順に並べて処置群と統制群がきれいに分かれてしまったらペアが作れなくなりますが、 そうならないようにできるだけ共変量を増やすという理解であっていますか？	高橋先生：「きれいに分かれてしまったらペアが作れなくなる」はオーバーラップ条件のことなのでたしかに問題となります。傾向スコアによる分析ではペアを作っていくことになるので、列よりは行を増やす（ペアリングの候補を増やす）ほうが重要になるかと思います。
12	等しい傾向スコアを持つ3人（例えば1人は処置群、 2人は統制群）がいた場合、 ペアではなく、 3人の観測値を利用するのでしょうか？	高橋先生の回答：これは後ほどマッチング方法のところで説明します。
13	重回帰の時と違い、傾向スコアは予測であるから、共変量の共線性などは深く考えなくてもよく、共変量はたくさんあればあるほどよい。のでしょうか？	高橋先生：一般的にはそうなります。基本的には気にしないで良いですが、極めて強い多重共線性の場合には推定自体が不安定になってしまうため予測値もうまく得られない可能性があります。また中間変数には気をつけてください。傾向スコアの推定に結果変数を入れないようにも気をつけてください。
14	傾向スコアを用いた分析では、最終的に平均値を比較する際のサンプルサイズが、実際に獲得したデータ総数よりも小さくなる、という理解で正しいでしょうか？	高橋先生：その通りです。マッチングされないデータが出現する分だけ減少します。どの程度減少するかは状況によります。
15	昨日の講義では、処置は連続変数であっても扱う（投薬量など）ことが可能と仰られていたかと思います。傾向スコアでも処置は連続変数を取りうるのでしょうか。	高橋先生：一般的な傾向スコアはロジスティック回帰モデルから算出するため二値であることが一般的ですが、処置が連続量の場合には通常の回帰モデルによって算出できるかもしれません。
16	独立性が前提条件となっているということですが、例えば昨日のテストの話にありました、入学試験の点数によって補習ありorなしが決まっている場合は傾向スコアマッチングを使った手法は適していないのでしょうか？	高橋先生の回答：Xが与えられたときに潜在的結果変数のペアが独立であればいいので問題はありません。昨日の補習の例では、入試得点ごとに割り当て確率が断続的に決まってしまうため、オーバーラップ条件が満たされないため傾向スコアを使うことができません。そういった場合には、この後で説明する回帰不連続デザインなどで対応します。
17	何かの措置を与えられた処置群の数の方が、統制群より少ないというアンバランスなことが、現実の観測データにはよくあると思います。このとき、処置群と統制群のアンバランス度合いというのは、どれくらい許容されるのでしょうか？判断の基準は何かありますでしょうか？	高橋先生：処置群のほうが少ないことで最終的にマッチするデータの数が少なくなることはあります。どの程度まで許容できるか、というよりは小標本の問題に繋がります（復元・非復元抽出の話題とも関連するかと思います）。ちなみに、傾向スコアマッチングでは基本的に少数である処置群に合わせてマッチさせるため、推定されるのはATTとなります（後ほど説明あり）。
18	欠測値を含むデータを用いて傾向スコアを算出する際には、補完して上で算出する必要があるのでしょうか？欠測値を含んだまま傾向スコアを算出できるのでしょうか？	高橋先生の回答：多重代入法で補完してください。ch.20.1で論じています。
19	P21の傾向スコアでペアにするところですが、被験者15と17のペアを、被験者17と20に設定されないのは、被験者15と17のps1の値が17と20よりも小さいからでしょうか？	高橋先生の回答：これはどちらでもよかったかもしれませんが、丸める前の傾向スコアで近いものを選んだだけかもしれません。あくまでも例として用意しただけなので、あまり気にしないでください。現実には判断の分かれるところだと思います。

20	<p>傾向スコアの関心は予測精度に帰着するため、いわゆる機械学習のワークフローに則るものと理解したのですが、相違ないでしょうか (i.e., データを train/validate/test に分割し、ハイパーパラメーターがある場合は validate データにてチューニングを行い、test データにおける精度がある一定の精度に到達するまでモデリングを試行錯誤する...)</p>	<p>高橋先生：考え方としてはそうなります。共変量の情報から傾向スコアを予測するまでのところは機械学習のフローになります。</p> <p>現在でもデフォルトではロジスティック回帰で傾向スコアを算出することが多いですが、最近では各種機械学習的手法で傾向スコアを算出する方法も色々あります。</p> <p>ただし機械学習と違い、得られた傾向スコアは因果推論のために用いられるので、傾向スコア推定後のフローは機械学習と異なります。</p>
20	<p>傾向スコアの関心は予測精度に帰着するため、いわゆる機械学習のワークフローに則るものと理解したのですが、相違ないでしょうか (i.e., データを train/validate/test に分割し、ハイパーパラメーターがある場合は validate データにてチューニングを行い、test データにおける精度がある一定の精度に到達するまでモデリングを試行錯誤する...)</p>	<p>横から恐縮です。傾向スコアでは確率の予測が重要であることから機械学習パートでは probability calibration というのも validation を用いて行われたりします。ご参考になればと思います。</p> <p>https://rmizutaa.hatenablog.com/entry/2019/03/30/224854</p>
21	<p>傾向スコアは目的変数なしの Xs 情報の縮約だと思うのですが、目的変数なしの状態では有益な Xs の抽出はどのように考えれば良いのでしょうか？</p>	<p>高橋先生の回答：傾向スコアは、共変量 X が与えられたときに処置が割り付けられる確率です。目的変数は処置となるのでロジスティック回帰で ok です。この話はこのあとします。</p>
22	<p>個々の x_1, x_2, x_3 の値が似ていなくても、傾向スコアが似ていればペアになるという理解でよいのでしょうか。また処置群と統制群の条件付き分布が同じとは、個々の x ごとに分布が同じである必要がありますか、あるいは多変量分布が同じということでしょうか。</p>	<p>高橋先生：前半についてはその通りです。</p> <p>後半に関しては、一般的には各変数の一変数分布が同じであり、かつ多変量の縮約としての傾向スコアの分布が同じであれば良い、ということになります。</p> <p>したがって、マッチごとに見たときには各変数の値は近くないかもしれませんが、全体として分布が近くなれば良い、ということになります。</p>
23	<p>バランシングスコアの算出で、データからペアをつくることからサンプル数が十分ないと標準誤差が大きくなるという理解は正しいでしょうか？</p>	<p>高橋先生：その理解で正しいです。サンプルサイズが減少することよりもマッチングで偏りを取り除くメリットのほうが大きいためにマッチングを行っています。</p>
24	<p>今回は傾向スコアが近いサンプルを人間が選びましたが、自動的に選べますか？</p> <p>また、バランスさせるために、1:3 みたいな感じでマッチングさせることはできますか？</p>	<p>高橋先生の回答：はい、そうです。マッチングは 1:1 で対応するのが分かりやすく簡単ですが、復元抽出によるマッチングでは 1:3 とか 1:多のマッチングとなります。この話はあとでします。実際の解析はパッケージを使って自動で行います。</p>
25	<p>初歩的な質問かもしれませんが、傾向スコアマッチングは処置変数が 3 値以上であっても使用できますか？</p>	<p>高橋先生：多項ロジットなどを用いれば可能です。(もともとの状況設定としては二値を想定しています)</p>
26	<p>p.27 で求まっているものは ATE ではないのでしょうか？ (ATT と比較されていますが)</p>	<p>高橋先生：傾向スコアは処置群にマッチさせるためのものなので p.27 で推定しているものは ATT という理解で良いです。</p>
27	<p>条件付き独立性のお話の中で「昨日の重回帰と同様、中間変数は入れない方が良い」と仰いましたが、傾向スコアは y に対する 1 つの x の直接効果に関心があるわけではないため、入れても入れなくてもどちらでも良いようにも思いました。この認識は間違っていますでしょうか。</p>	<p>高橋先生の回答：結果変数に対する処置変数の効果を見たいというのは傾向スコアでも同様なので、中間変数を入れるのは同じ理由で問題となります。</p>

28	<p>そもそもですが、非実験研究では、最終的に検定をしたい場合でも、サンプルサイズは事前に設計しなくてもよいのでしょうか？</p> <p>傾向スコアを使うと、サンプルサイズが、集めた検体の量よりも減ってしまうため、そもそものサンプルサイズ設計に意味がなさそうなのですが。</p>	<p>非実験研究＝観察研究</p> <p>高橋先生：観察研究ではデータを解析する人がデータの収集には関与していないのが一般的だと思います。その意味で、観察研究では事前のサンプルサイズの設計の話との相性はあまり良くないです。ただしサンプルサイズが検定結果に影響するのは実験研究と同じことなので、その点には注意を払う必要があります。</p> <p>サンプルサイズが減少することよりも、バイアスが取り除かれることのメリットが大きいと判断したらサンプルサイズの減少はやむなし、ということになります。</p>
29	<p>傾向スコアが同じ人をマッチングさせるのは、共変量（傾向スコア）の処置変数への影響を排除することができるからでしょうか？</p>	<p>はい、そういうことです。</p> <p>傾向スコアで同じ人をマッチングしているというのは、事実上の同じ人とみなしているということです。</p> <p>傾向スコアは人同士ではなくグループでbalancingしてはいます。</p> <p>実験研究と同じで、集団としてみたとき同一のように扱います。</p> <p>平均値や分布がそろっているということはそういうことです。</p>
30	<p>電波が少し良くないため、私が聞き逃していたら、大変申し訳ございません。</p> <p>傾向スコア分析は、パネルデータにおける固定効果モデルや変量効果モデルと組み合わせて利用可能でしょうか？</p>	<p>横から失礼します。パネルデータ（time-series--cross-sectional data）であれば、Kosuke Imai他の"Matching Methods for Causal Inference with Time-Series Cross-Sectional Data"があります。</p>
30	<p>電波が少し良くないため、私が聞き逃していたら、大変申し訳ございません。</p> <p>傾向スコア分析は、パネルデータにおける固定効果モデルや変量効果モデルと組み合わせて利用可能でしょうか？</p>	<p>高橋先生：原理的にはパネルデータにおいても可能だと思います。</p>
31	<p>処置群と統制群の条件付き分布が同じにならないxがあった場合はどうしますか。</p>	<p>高橋先生の回答：</p> <p>後で話をしますが、モデリングのやり直しになります。</p> <p>傾向スコアを算出するために、本講義ではロジスティック回帰モデルを組み立てていました。</p> <p>その中で、Xについては線形を仮定していましたが、そこは対数を取ったり2乗を取ったりできます。</p> <p>モデルを誤って設定すると、balancingがうまくいきなことがあります。</p> <p>この辺りを設定しなくても、balancingが取れてることが多いため、柔軟であるとされていますが、やはり、balancingが取れているかを確認し、議論をする必要があります。</p>

32	素朴に、Xの代わりに傾向スコア $e(X)$ を説明変数とする 回帰分析の問題に落とし込めたりしないのでしょうか。	高橋先生：傾向スコアをマッチングではなく直接そのまま説明変数として使う手法も無くはないのですが、それが主流なのかはわかりません。 傾向スコアはATTを推定するために使用するため、処置群での効果を推定するためには良いのですが、ATE（母集団全体での効果）を推定するためには適切ではない可能性があります。ATEを推定する場合には層化解析法や重み付け法のほうが良いと思います（書籍Chap. 12）
32	素朴に、Xの代わりに傾向スコア $e(X)$ を説明変数とする 回帰分析の問題に落とし込めたりしないのでしょうか。	横から失礼いたします。おっしゃっていた分析に自分も興味があったので Rでやってみたところ、やはり真のATTを推定できる手法ではないようです。使用したRコードを添付いたしますので、ご参考になれば幸いです。 ----- rm(list = ls()) data11 <- read.csv("data11.csv") attach(data11) mean(y1[t1==1]) - mean(y0[t1==1]) #true score of ATT: 2.889 #propensity score ps_maker <- glm(t1 ~ x1 + x2 + x3 + x4 + x5 + x6, family = binomial(link = "logit")) ps <- ps_maker\$fitted.values summary(lm(y3 ~ t1 + ps)) #coefficient of t1: 3.4142
32	素朴に、Xの代わりに傾向スコア $e(X)$ を説明変数とする 回帰分析の問題に落とし込めたりしないのでしょうか。	krskさんがブログで取り上げていて、以下の記事の「アウトカムモデルによる調整」に相当すると思います。 https://www.krsk-phs.com/entry/propensity_score_comparison このブログ記事は以下に論文化されているようです。 https://www.jstage.jst.go.jp/article/jea/advpub/0/advpub_JE20210145/_article/-char/en
33	元のサンプルが無作為抽出されているとしても、傾向スコアでマンチングすると無作為抽出とはならなくなると缶絵が得て良いのでしょうか	高橋先生の回答：通常はそういう理解はしていないと思います。共変量の分布を揃えているだけです。ただ、マッチングの仕方によって代表性が損なわれることは場面によってはあり得るかもしれません。この部分が問題視されているという話はあまりないと思います。

34	<p>予測精度の良くないロジスティック・モデルで算出された傾向スコアを使ってATTを推定したとき、真のATT(があるととして)に対してどのようなバイアスが発生すると考えられますか？ロジスティック・モデルの精度はどの程度あるばよいという目安はありますか？</p>	<p>高橋先生：書籍の中では真の傾向スコアが既知の場合と予測値が使える場合については言及しています。真の傾向スコアがわかっていたとしても、予測モデルから得られる傾向スコアを用いたほうがより良く因果効果を推定できることが知られています。</p> <p>その点でも、重回帰モデルよりも傾向スコアモデルのほうがフレキシブルだと言えます。傾向スコア自体の正しさはそこまで大きな問題とはならなさそうです。</p> <p>とはいえ、あまりにかけ離れた値にはならないように、共変量のバランシングには気をつけておく必要があります（バランシングがうまく行っていない場合には傾向スコアを用いた推定はうまく行かないかもしれない）。</p>
34	<p>予測精度の良くないロジスティック・モデルで算出された傾向スコアを使ってATTを推定したとき、真のATT(があるととして)に対してどのようなバイアスが発生すると考えられますか？ロジスティック・モデルの精度はどの程度あるばよいという目安はありますか？</p>	<p>予測精度が低い場合に、モデリングは正しいが処置群と統制群の共変量分布が被っているから精度が低い場合と、モデリングがおかしいから精度が低い場合があって、その二つの区別が難しいということがあると思います。</p> <p>目的は共変量をバランスさせることなので、その目的に従って検討すべきと私は理解しています。（補足訂正があれば識者にお願いしたいです）</p>
35	<p>質問ではありませんが、自身の先ほどの質問についてFYIです。Propensity scoreは機械学習的な手法（e.g., random forestやboosting等）でも算出は可能ですが、左記モデルの分類器としての性能と確率が正確に算出ができるかは分けて考える問題であり、Calibrationとといった特殊な処置を行う必要があります（加えて実際に出てくる確率が正しいとも言えないケースもあり）。Logistic回帰（おそらくProbitも）はこのような処理は不要なため、やはりLogistic回帰がシンプルかつ強力な手法と言えそうです。※参加者の青木さんにご教示いただきました、ありがとうございます！！</p>	
36	<p>傾向スコアマッチングではATEでなくATTを求めているとのことですが、これは処置群のサンプルに対する反実仮想を統制群からマッチングによりサンプルしているから、という理解ですがあってますでしょうか。またその場合、見方を変えたと統制群に対してマッチングにより処置群をサンプルしていると見ることもできるので、ATTとも言えないような気もしますが、いかがでしょうか。</p>	<p>高橋先生の回答：はい、前半に関してはそうです。一般には、処置群の人に対して統制群の人をマッチングする方法をとります。例えば処置群10名、統制群30名といった場合に、処置群の10名に対して統制群をマッチングすると、残った統制群20名はマッチングされないことになるので、ATTとなります。</p>
37	<p>共変量の縮約をするのにロジスティック回帰による傾向スコアではなく、主成分分析のような次元圧縮手法も使えるでしょうか。</p>	<p>高橋先生：そのような研究例があるか私はよく知らないですが、原理的にはできると思います。</p> <p>ロジスティック回帰で傾向スコアを得る方法が最も基本的なのですが、様々な機械学習の手法を使うことは可能です（どの方法が最適かはわかりません：場合によると思いますが）。</p>

38	傾向スコアでマッチングして処置群と統制群の結果の平均の差を算出した後で、その差が有意かどうかをt検定等で確認する必要はあるのでしょうか？	高橋先生の回答：はい、あります。少なくとも標本から母集団の推測を行いたいのであれば、t検定や信頼区間等を使って評価することになります。
39	現在のx1からx3は実測数値ですけれども、データを標準化した場合で行うと、異なる概念になるのでしょうか？	高橋先生：データを標準化したあとで傾向スコアを出した場合に異なる結果が出るかについては把握していません。もし大きく異なる結果が出るのであれば、2つの結果を見比べてどちらがより意味があるのかを考える必要があるかもしれません。
40	傾向スコアのモデリングの良さは共変量がバランシングできるかでチェックすべきと理解しています（補足訂正があればご指摘いただきたいです）が、共変量のバランスの良さについて定量的な議論はございますか？ 交絡変数が全て観測できている条件は必要と思いますが、例えば「とあるバランシング尺度が δ 以下であれば、因果効果のバイアスが ε 以下にできる」などの定理があれば最高だと思うのですが。	高橋先生の回答：後で出てきます。検定等ではありませんが目安となる基準があります。ただし、そこまで明確な定理と呼べるほどのものではありません。
41	P. 60の復元抽出から出てくる「マッチングの重み」はどのように算出されるのでしょうか。簡単で結構ですのでご教示頂けますでしょうか。	高橋先生：matchit関数の参考文献には書いてあると思います。（参考文献の214, 215番をご参照ください）
41	P. 60の復元抽出から出てくる「マッチングの重み」はどのように算出されるのでしょうか。簡単で結構ですのでご教示頂けますでしょうか。	補足：Rでhelp(matchit)で表示されるヘルプのうち"How Matching Weights Are Computed"のセクションがこれに該当する情報かと思われます。
42	傾向スコアを計算する際に選ぶ共変量は、傾向スコアを出したいただけなので共線性などは気にしなくて良く、中間変数になっていないことだけを気にするという話だったと理解しています。その際、実際にはどのように選んだ変数を中間変数だと判断するのが良いのでしょうか？ドメイン知識を用いて、処置の有無によって変わってしまいそうなものは避ける、とかでしょうか？	高橋先生の回答：重回帰のときの話と同じで、DAGを描いて、共変量がどこに入るのかをドメイン知識から考えて、処置と結果変数の間にあったら中間変数としてモデルに含めない、とすればよいです。
43	傾向スコアの解析モデルで、ご説明のように $\text{lm}(y_3 \sim t_1 + X_s)$ とするのと、 $\text{lm}(y_3 \sim t_1 + PS)$ とするのでは、違いはなんなのでしょうか？後者で良いとする文献もあるようです。	横から失礼します。直感的にはPSに縮約したことで失われた情報量を補完しているイメージかな、と思いながら聞いてました。
43	傾向スコアの解析モデルで、ご説明のように $\text{lm}(y_3 \sim t_1 + X_s)$ とするのと、 $\text{lm}(y_3 \sim t_1 + PS)$ とするのでは、違いはなんなのでしょうか？後者で良いとする文献もあるようです。	高橋先生：私は後者で良いとしている文献を存じ上げないのですが、そういった文献があるならばそちらにもメリットがあるのかもしれません。 改めてX1-6を入れているのは、傾向スコアでは取り除けなかった偏りがまだある可能性を考慮してのことです。ここをPS一つだけにすると、個々の変数による偏りを十分に取り除けない可能性があります。一方で、標本サイズに対して共変量が多い場合には、すべての共変量を放り込んでしまうと自由度が減少してしまうため、それを避けるためにPS一つだけを入れるメリットはあるのかもしれません。

43	傾向スコアの解析モデルで、ご説明のように $lm(y_3 \sim t1 + Xs)$ とするのと、 $lm(y_3 \sim t1 + PS)$ とするのでは、違いはなんですか？後方で良いとする文献もあるようですが。	(もしよろしければ、後方で良いとしている文献をあげていただけると、他の受講者の参考になるかもしれません。よろしくお願いいたします。)
43	傾向スコアの解析モデルで、ご説明のように $lm(y_3 \sim t1 + Xs)$ とするのと、 $lm(y_3 \sim t1 + PS)$ とするのでは、違いはなんですか？後方で良いとする文献もあるようですが。	<p>$lm(y_3 \sim t1 + PS)$ はマッチングデータではなく、全データを使った傾向スコア直接補正法というもので使われているようです。ありがとうございました。</p> <p>傾向（プロペンシティ）スコアの各使用法の仮定・解釈の違いを比較してみた - Unboundedly https://www.krsk-phs.com/entry/propensity_score_comparison</p>
44	P.82の(因果効果推定の)重回帰において傾向スコアマッチングから出てくる重み(weights)付けすることは必須でしょうか。重み回帰しないと問題があるのでしょうか。	<p>高橋先生：復元抽出によるマッチングの場合には重みを考慮する必要があります。非復元抽出の場合はどちらも問題ありません。</p> <p>非復元抽出よりは復元抽出のほうが良いので、復元抽出で重み付けを行うのが良いと思います。</p>
45	マッチング法ではATTしか推定できないということですが、であれば、ATEを推定できる層化解析法や重み付け法を使う方がより適切、ということでしょうか？換言すれば、我々が知りたいのは本来はATEであり、ATTの方を積極的に知りたい場合、というのはそんなに多くあるのでしょうか？Kingらの批判も踏まえると、層化解析や重み付けとは別に、マッチング法を理解することの意義がよくわからなくなりました。	高橋先生の回答：それは研究のテーマに依存します。講義2で解説したように、研究テーマによってはATTに関心がある場合もあります。ATEを推定したいのであれば、傾向スコアマッチングではなく層化解析や重み付け法を用いる必要があります。
46	アンバランスな変数がある可能性もあり、共変量あり解析モデルでATT（やATE）を推定したほうが性能が良いとありましたが、アンバランスな共変量のみを解析モデルに投入するというのは避けるべきでしょうか（そういった解析をしている論文があり悩んでおりました）。	<p>高橋先生：p85の例ではバランシングはうまく行っているように見え、「アンバランスな共変量のみを解析モデルに投入する」としたらどの共変量も投入する必要はないと思われるのですが、実際には投入しない場合推定値は真値からずれた値になっています。このように、バランシングがうまく行っている場合でも念の為にしておいたほうが良いだろうと思います。</p> <p>書籍では指摘していますが、傾向スコア推定に用いなかった共変量を解析モデルに入れるのは明確にダメですが、用いた共変量に関しては入れておいたほうが良いと思います。</p>
46	アンバランスな変数がある可能性もあり、共変量あり解析モデルでATT（やATE）を推定したほうが性能が良いとありましたが、アンバランスな共変量のみを解析モデルに投入するというのは避けるべきでしょうか（そういった解析をしている論文があり悩んでおりました）。	高橋先生：PSモデルの修正については両方だと思います。変数変換などについてはドメイン知識が活かされるでしょうし、うまくバランシングができていのかは得られた結果を見て判断することになります。

47	<p>PSマッチングまたはIPWを行ったあと、バランスのチェックをしますが、どうしてもバランスが取れない場合があります。PS推定のためのモデルを修正して再度試みるのですが、バランスがとれるようにするための、コツなどありますか？</p> <p>交互作用項や2乗項を入れることはよくやるのですが、なかなかバランスが改善されません。</p>	<p>高橋先生：交互作用項や二乗項を入れて修正を行うという程度がアドバイスできる範囲です。</p> <p>もともとバランシングしていないものを無理やり改善させてもうまく行かないこともあるので、改善しないならばそもそも因果推論に適していないデータなのかもしれません。</p>
48	<p>傾向スコアマッチング後の共分散分析（解析モデル）において、共分散分析で問題になっていたパラメータの非線形性などは問題にならないのでしょうか？</p> <p>十分に共変量の影響を排除している状態だから、大丈夫なののでしょうか？</p>	<p>高橋先生の回答：これは変数の非線形ですね。問題にならないと断言はできません。共分散分析の場合はモデル上影響を排除しようとしているので関数の形が分からないと正しく取り除くことはできませんが、傾向スコアマッチングでは関数形に対してよりflexibleなので、リスクはあるものの共分散分析よりは問題になりにくいです。</p>
49	<p>（聞き逃していたら申し訳ありません）</p> <p>似ているとする閾値はどうやって決めますか。あるいは閾値は設けず、処置群は一人残らずペアを作のでしょうか（非復元の場合は統制群が残っている限りペアを作るのでしょうか）。</p>	<p>高橋先生：書籍の例では閾値は設けずに解析を行いました。閾値を設ける場合にはmatchit関数に引数caliperを指定することができます。ただし閾値内にマッチ可能なデータがない場合には脱落してしまうため、データサイズが小さくなってしまいます。マッチの精度とデータサイズのトレードオフを考慮して閾値を設定する必要があります。</p>
50	<p>書籍p.170で言及されている「回帰調整」とは、資料p.82の解析モデルの実行のことでしょうか。</p>	<p>高橋先生の回答：はい、そうです。マッチング済みのデータを使って解析するときに、改めて共変量もモデルに入れるという意味です。</p>
51	<p>念のため確認させていただきますが、傾向スコア算出ためのロジスティック回帰モデル（傾向スコアモデル）においては、まず高次項や交互項を入れずに必要十分な共変量をP.81のようにシンプルに入れて傾向スコアを算出するが、その後の共変量のバランスチェックでバランスしなかった場合に、高次項や交互項などを入れてモデリングし直すというのが基本スタンスと考えて宜しいでしょうか。</p> <p>他方、（P.82のような）因果効果推定のための重回帰モデルでは、上記の傾向スコアモデルに高次項や交互項などを入れたとしても、このような複雑なモデルはする必要はなく、傾向スコアモデルで使った共変量をP.82のようにシンプルに（高次項などを入れずに）投入するだけでよいと考えて宜しいでしょうか（何故なら基本的にマッチドデータにおいて交絡は除去できているので、でも復元抽出なら重み付け回帰をする必要がある）。</p>	<p>高橋先生：（前半について）私はそのように理解しています。ただ事前にドメイン知識などをもとに「交互作用項を入れておいたほうがいい」といったことがわかっていればはじめから入れておいても良いと思います。</p> <p>（後半について）傾向スコアマッチングで大半の交絡は除去できているので、因果効果推定ではシンプルなモデルで問題ない場合が多いです。ただしこれも絶対ではないので、まだうまく行かない場合には因果効果推定のモデルにも高次項や交互作用項を入れることを考えても良いと思います。</p>
52	<p>p21で、β_{1iv}とβ_1とがほぼ同じ大きさになっているのは、「x,yの重なりのうちzが重なっていない部分」が小さいからだとして理解したのですが、β_{1iv}によるβ_1の推定の適切さは「zがx,yの重なりをどれほど説明しているか」に強く依存するのでしょうか？</p>	<p>高橋先生：その通りです。操作変数の仮定2（関連性）について、計算が識別できるという意味では$cov(Z,X)$が0でなければよいのですが、実際にはこの関係が強ければ強いほど因果効果も正しく推定できるようになります。</p>

53	操作変数法で推定される因果効果は、ベン図の面積を見る限りは本来の β よりは過少（大）推定されるようですが、因果効果としてはATEなののでしょうか、それとも傾向スコアのようにATTなののでしょうか。	高橋先生の回答：ベン図は楕円の作図の問題で正確には一致しません。本来はほぼ一致するはずですが、不偏推定量ではなく一致推定量なので、小標本の場合には偏りがある可能性があります。大標本であれば偏りは小さくなります。因果効果としてはATEを推定していると考えて結構です。
54	P.17 $X^{\{1\}}$ を「操作変数」として、は誤植でしょうか。Zが操作変数でしょうか。	高橋先生：誤植ではありません。Zが（1段階目の）操作変数ではあるのですが、Zから算出した $X^{\{1\}}$ も「（2段階目の）操作変数」と呼んでいます。 操作変数Zが複数個あるケースで言えば、 $X^{\{1\}}$ は複数の操作変数を1変数に集約した「操作変数」である、ということです。
55	操作変数法の一段階目の回帰はノンパラメトリックにやっても良いものなのでしょうか？ その際の注意点について簡単に触れていただければ幸いです。	高橋先生の回答：おそらくだめだと思います。ノンパラ回帰でやることの是非について検討したことはないのですが断言はできませんが、最小二乗法以外を使ったときの話については書籍のp.195で紹介しています。そこでは、tが二値変数のときにロジスティック回帰を使うのはよくない、という話をしています。最小二乗法を使ったときのみ相関が0になることが保証されているからです。ノンパラの場合もこれと同様にうまくいかないと考えられます。
55	操作変数法の一段階目の回帰はノンパラメトリックにやっても良いものなのでしょうか？ その際の注意点について簡単に触れていただければ幸いです。	回答メモありがとうございます。上記の「tが二値変数のとき～」のtはスライドでのx、処置変数と思います。備忘まで。
56	P60のPPについての質問です。PPは、P55の表で見ると、遵守者のみを対象となっている（CACEになっている）ように思うので、適切な対象を観察しているように思います。ATM同様に不適切となるのはなぜでしょうか？	p.52の表に示したとおり、 $T=1$ で $D=1$ の人は「遵守者」だけでなく「常に処置を受ける人」を含みます。また、 $T=0$ で $D=0$ の人は「遵守者」だけでなく「常に処置を受けない人」を含みます。したがって、PP推定量は、「遵守者」の平均処置効果ではなく、「遵守者」、「常に処置を受ける人」、「常に処置を受けない人」の平均処置効果であるため、偏りが発生すると考えられます。IV推定量を用いることで、「常に処置を受ける人」と「常に処置を受けない人」を除外制約によって解析から排除して、ITT推定量の過小推定を是正した推定ができると理解できます。岩崎先生の『統計的因果推論』のpp.162-164にも解説がありますので、そちらもあわせてご覧ください。
57	実際に受けた処置 d_1 はどのように観察することが多いのでしょうか。薬の例だと、実際に薬を飲んだかどうかは患者さんの自己申告ということになるのでしょうか。これだとバイアスがかかると思うのですが、どのように対処されているのでしょうか。	高橋先生：確かに薬の例の場合は自己申告になりますが、その自己申告には嘘はないと考えるしかないと思います。天邪鬼と同じように、いたとしてもそう多くはないと考えるか、ウソを付くことに対するインセンティブがないからウソをつかない、と仮定して分析をすすめるしかないと思います。

58	<p>量販店にて顧客単価の高い顧客リストを基にクーポン配布施策を行い、販促活動としてクーポンが適切だったかを検証したいとします。クーポンは顧客リスト内でランダムに配布され、結果変数として特定期間における売上を想定します。すると、実際にクーポンを使うかどうかは顧客次第なので、講義にあった非順守の問題が発生すると理解しています。</p> <p>その場合はシンプルに、遵守者・非遵守者のピボットテーブルを作成し、$\Pr[D=1 T=1]$と$\Pr[D=1 T=0]$を算出し、IV推定量を計算することのより一致推定量として算出できるという理解ですが、このアプローチは正しいのでしょうか（その場合、ivregをわざわざ使う理由がよくわからなくなっていました）</p>	<p>高橋先生の回答：どちらでやってもかまいません。点推定値を出すだけであれば計算して求めればよいです。この手計算は二段階最小二乗法をやっていることに相当します。標準誤差や信頼区間まで出したい場合は手計算では難しいのでivregを使った方がよいでしょう。</p>
59	<p>ITT推定量が保守的になるのは、直観的にどのように理解したらよいでしょうか？</p>	<p>高橋先生：どちらの群も完全に遵守者の場合、IV推定量はITT推定量に一致します。そこからTとDが乖離するほど、真の効果が小さくなるように見えている、というイメージです（書籍のp205の脚注10）。</p>
59	<p>ITT推定量が保守的になるのは、直観的にどのように理解したらよいでしょうか？</p>	<p>横から失礼します。処置群の結果変数が処置しろと言われたのに処置しなかった人のせいで低くなり、統制群の結果変数は逆に処置すると言われたのに処置した人のせいで高くなってしまいうから、と理解しました。</p>
60	<p>p. 62 の IV 推定量について信頼区間を構成できますか？</p>	<p>また IV 推定量は一致推定量ということですが、これが現時点でのベストな推定量でしょうか？</p>
60	<p>p. 62 の IV 推定量について信頼区間を構成できますか？</p>	<p>高橋先生の回答：手作業で計算したものがp.62の式で、二段階最小二乗法と同じです。Rを使えば信頼区間も計算できます。今のところは、IV推定量とITT推定量の両方を報告するのがベストとされています（IV推定量は是正しすぎている可能性もあるので）。</p>
61	<p>非遵守者をリストワイズ削除するのは、それ自体が欠測データの文脈のように何らかのバイアスになってしまうのでしょうか</p>	<p>高橋先生：試してみたことはないのですが、多分バイアスになるのだと思います。</p>
62	<p>p.73で$t1=1$かつ$d1=1$が8個で、8/20と説明されていますが、 p.62分母から見ると、$8/10 - 4/10 = 4/10$となります。結果は一緒ですが、p.62の式とは異なる説明になっているのではないのでしょうか。</p>	<p>高橋先生の回答：そうですね。正しいご指摘です。修正しておきます。</p>
63	<p>先ほどの、非遵守者の除外に関する質問に関係するのですが、非遵守者は「遵守したときの結果変数がmissing data」と見なして、他の共変量の情報などを使って補完する、という方法はあるのでしょうか。</p>	<p>高橋先生：一般論として、欠測データ解析の知見から考えると、解析対象である結果変数を共変量などから推測して補完しても意味はありません。missingの補完はモデルの右辺に関しての手続きです。</p>

63	<p>先ほどの、非遵守者の除外に関する質問に関係するのですが、非遵守者は「遵守したときの結果変数がmissing data」と見なして、他の共変量の情報などを使って補完する、という方法はあるのでしょうか。</p>	<p>高橋先生：一般論として、欠測データ解析の知見から考えると、解析対象である結果変数を共変量などから推測して補完しても意味はありません。missingの補完はモデルの右辺（説明変数の欠測）に関しての手続きです。</p>
64	<p>言葉使いの所でちょっと混乱したので確認させてください。</p> <p>以下のように言い換えて良いでしょうか。</p> <p>T1: 処置の割付け -> 何かを奨励するかどうかの割付け</p> <p>D1: 実際に処置を受けたかどうか -> 奨励されたことを実施するかどうか を表す</p> <p>先ほどのご質問に出た販促でクーポンを配る例では処置T1は「購入するよう促す（そのためにクーポンを配る）」</p> <p>ですが、D1は「購入するかどうか」であって、クーポンを実際に受け取ったかどうか、ではないという認識です。</p> <p>購入するかどうか、は結果と思われるので、何かまだ混乱していそうです。</p>	<p>高橋先生の回答：T1,D1についてはそのように言い換えて理解して大丈夫です。売り上げが増減するというのが結果変数だとすれば、購入時に受け取ったクーポンを使うかどうかD1だと思います。クーポンを配るかどうかは無作為化されているのでT1です。</p>
65	<p>講義7のスライドp.10の操作変数についてお伺いします。</p> <p>$Z \rightarrow X \rightarrow Y$と$U \rightarrow X$と$U \rightarrow Y$というDAGの矢印の向き先を見ると昨日の講義にあった中間変数が思い出されました。操作変数はUからの影響を断ち切るために、Xを中間変数として捉えてZを算出したと理解しましたがこの理解は正しいのでしょうか？</p>	<p>高橋先生：因果モデル的には、XがZより算出されているという形です。</p> <p>ちなみに重回帰モデルや傾向スコアで統制すべき共変量の中には、操作変数が含まれてはいけません。操作変数はあくまでも操作変数法の枠組みで有用な変数であるという点ご注意ください。</p>
66	<p>処置の割付け決めるための変数は基本的に連続変数でなければ回帰不連続デザインを適用することは難しい、質的変数（男 vs. 女）やその他離散変数（リッカートスケール等）は原則不可と考えてもよろしいでしょうか。</p>	<p>高橋先生の回答：はい、そうです。基本的な回帰不連続デザインでは強制変数には連続変数を考えています。二値変数は閾値の周辺というものが存在しないので根本的に不可能です。リッカートスケールも、たとえば量的変数として扱ったとしても、1～5で3が閾値、となると2～4を閾値の周辺と呼べるかは微妙でしょう。基本的には連続変数でなければいけないということになります。</p>
67	<p>特定の効果（今回でいうと降圧剤）を測るために、閾値の設定が重要なように思えた（閾値によって、効果が大きく変わりうるように思えた）のですが、その認識は合ってますでしょうか？また、その場合、閾値をどのように設定するのが良いのでしょうか？どの閾値付近の人に効果があるかを知りたいかに基づいて設定するのでしょうか？</p>	<p>高橋先生：回帰不連続デザインは観察研究なので、閾値は研究者が決めることができるものではなく、得られたデータのもとですでに決まっているもの（降圧剤の例ならば医者が決めた値）です。したがって、研究者が閾値について悩むことはないはずです。</p> <p>ですが、おっしゃるとおり閾値によって効果の大きさは相当に変わるため、回帰不連続デザインの結果は「閾値付近ではどの程度の効果があるか」という局所的な効果を指し示すものに過ぎません。</p>

67	<p>特定の効果（今回でいうと降圧剤）を測るために、閾値の設定が重要なように思えた（閾値によって、効果が大きく変わりうるように思えた）のですが、その認識は合ってますでしょうか？また、その場合、閾値をどのように設定するのが良いのでしょうか？どの閾値付近の人に効果があるかを知りたいかに基づいて設定するのでしょうか？</p>	<p>高橋先生：回帰不連続デザインは観察研究における解析手法なので、閾値は研究者が決めることができるものではなく、得られたデータのもとですでに決まっているもの（降圧剤の例ならば医者が決めた値）です。したがって、研究者が閾値について悩むことはないはずです。</p> <p>ですが、おっしゃるとおり閾値によって効果の大きさは相当に変わるため、回帰不連続デザインの結果は「閾値付近ではどの程度の効果があるか」という局所的な効果を指し示すものに過ぎません。</p>
68	<p>バンド幅に含まれるデータが小さくないか、バンド幅の境界付近のデータと閾値付近のデータを等しく扱ってよいか等が気になるのですが、閾値からの距離に依ってデータに重みをつけるような手法もあったりするのでしょうか？</p>	<p>高橋先生の回答：はい、あります。密度推定の話として、書籍のch.16のpp.232-236あたりで紹介しています。時間があれば今日話します。</p>
69	<p>バンド前後の差が無視できると仮定した場合、交差項を用いないストレートな重回帰分析でtの効果が分かる気がしたのですが、そうしない理由がありますか？</p>	<p>高橋先生：十分に局所の範囲が制限されていれば、交差項は無視できるかもしれませんが、基本的には閾値の左と右でそれぞれ推定を行い差をとっています。交差項を置かないということは閾値の左右で推定した回帰モデルの傾きが同じだと仮定していることになるため、基本的には交差項は入れておいたほうが良いと思います。</p>
69	<p>バンド前後の差が無視できると仮定した場合、交差項を用いないストレートな重回帰分析でtの効果が分かる気がしたのですが、そうしない理由がありますか？</p>	<p>高橋先生：十分に局所の範囲が制限されていれば、2つの回帰直線はほぼ平行に見えるため交差項は無視できるかもしれませんが、基本的には閾値の左と右でそれぞれ推定を行い差をとっています。交差項を置かないということは閾値の左右で推定した回帰モデルの傾きが同じだと仮定していることになるため、基本的には交差項は入れておいたほうが良いと思います。</p>
69	<p>バンド前後の差が無視できると仮定した場合、交差項を用いないストレートな重回帰分析でtの効果が分かる気がしたのですが、そうしない理由がありますか？</p>	<p>高橋先生：十分に局所の範囲が制限されていれば、2つの回帰直線はほぼ平行に見えるため交差項は無視できるかもしれませんが、交差項を置かないということは閾値の左右で推定した回帰モデルの傾きが同じだと仮定していることになるため、基本的には交差項は入れておいたほうが良いと思います。</p>
70	<p>局所線形回帰モデルは、閾値周辺のデータを使っているので原則として交絡が除かれているため、割付変数のみモデルに入れても不偏推定量となるが、交互作用項等可能な限り共変量を入れるのは標準偏差を小さくし、推計値の安定性を担保するためでしょうか。</p>	<p>高橋先生の回答：強制変数以外の交絡が除かれているというのはその通りです。交互作用項を入れているのは標準誤差を小さくするためというよりは、閾値の左側からのモデルと右側からのモデルが異なる形を取り得るためそれを反映させるために入れています。</p>
72	<p>局所的な因果効果の推定では、局所であるが故に線形モデルで十分という感じなのでしょうが？線形性について問題になるケースはございますか？</p>	<p>高橋先生の回答：一言で言えば局所だから線形で十分ということです。Gelman & Imbens (2019, https://doi.org/10.1080/07350015.2017.1366909) ではっきりそう主張されています。昔は高次の項も入れていました。考慮するとしても二次の項までで十分です。局所の範囲なので一次でも十分です（微分・積分のイメージ）。</p>

73	lm()で推定した係数とSEのみ表示されていますが、p値は考慮するのでしょうか。もしそうなら、ごく一部のデータだけを使っているので $p < .05$ になりにくそうな気がしました。	高橋先生：標準誤差があるため、統計的推測・検定を行うことはできます（そのために標準誤差も出しています）。おっしゃるとおり、使用しているデータサイズが小さくなるため $p < .05$ にはなりにくくなります。そういった意味で、通常のデータ解析よりももともとのデータが多いほうが望ましいです（どの程度多ければ良いかは場合によります）。 今回提示した例では使用するデータの幅をかなり狭く（±2）しましたが、実際にはもう少し広くてもよいと思われ、その意味でももう少し多くのデータを用いて回帰を行うことが可能です。
74	ファジーなRDDは本質的に「局所的な操作変数法」（講義7, P.63 & 講義8a, P.53）とのことですが、ファジーなRDDにおいても、通常の操作変数法と同様に、適切なIVを見つけるのが難しいものなのでしょうか。それとも、ファジーなRDDではIVを理論的に比較的容易に見つけられるのでしょうか。	高橋先生：ファジーなRDDは無作為化奨励デザインを局所的に行ったものとイメージしていただくと良いです。したがって局所的に割り付け（血圧が閾値を超えるか超えないかはランダムな）がランダムに発生しているという感じです。したがって、ファジーなRDDでは理論的に適切なIVを見つけられる、ということになります。
74	ファジーなRDDは本質的に「局所的な操作変数法」（講義7, P.63 & 講義8a, P.53）とのことですが、ファジーなRDDにおいても、通常の操作変数法と同様に、適切なIVを見つけるのが難しいものなのでしょうか。それとも、ファジーなRDDではIVを理論的に比較的容易に見つけられるのでしょうか。	高橋先生：ファジーなRDDは無作為化奨励デザインを局所的に行ったものとイメージしていただくと良いです。したがって局所的に割り付け（血圧が閾値を超えるか超えないかはランダムな誤差によって変動する）がランダムに発生しているという感じです。したがって、ファジーなRDDでは理論的に適切なIVを見つけられる、ということになります。
74	ファジーなRDDは本質的に「局所的な操作変数法」（講義7, P.63 & 講義8a, P.53）とのことですが、ファジーなRDDにおいても、通常の操作変数法と同様に、適切なIVを見つけるのが難しいものなのでしょうか。それとも、ファジーなRDDではIVを理論的に比較的容易に見つけられるのでしょうか。	高橋先生：ファジーなRDDは無作為化奨励デザインを局所的に行ったものとイメージしていただくと良いです。したがって局所的に割り付けがランダムに発生しているという感じです（例えば降圧剤の例では、血圧が閾値を超えるか超えないかはランダムな誤差によって変動する）。したがって、ファジーなRDDでは理論的に適切なIVを見つけられる、ということになります。
75	ファジーな回帰不連続デザインは強制変数Xを条件づけたときに条件付き正値性が成り立ち、仮に条件付ける共変量が全てわかっているのであれば、この場合は傾向スコアマッチングも可能という理解でよろしいでしょうか。	高橋先生の回答：おそらくそういう考え方になるのだと思います。回帰不連続デザインの利点は、条件付けるべき共変量が全て観測できていなくても適切な因果推論ができるということなので、共変量が全て分かっているのであれば傾向スコアで対応できると思います。
76	AICによるモデル選択が因果推論に適していないという話についてなんですが、AICがベースにしている汎化誤差が外的妥当性に対応するような概念になっていて、だから内的妥当性について議論している因果推論と噛み合わないように感じたのですが、そういった雰囲気議論が統計学的にされていたりするのでしょうか？	高橋先生：そのような理解でも良いかと思います。AICは予測に関する誤差を小さくしようという目的なので、AICが選んだ「良い」モデルは、因果効果の推定における「良い」モデルとは異なる意味になります。

77	<p>今回講義していただいた考え方や分析手法が、いわゆる Rubin流の統計的因果推論なのでしょうか。他に Pearl流や Cambell流があると聞いたことがあります、全く異なるものでしょうか。(統計的因果推論の全体像を把握したいです)</p>	<p>高橋先生の回答：Cambellについては、最後に話した回帰不連続デザインはもともとCamellが提案したものであるので関連があると思います。Pearl流についてはDAGがそうです。私はRubin流に馴染みが深いのでRubin流の統計的因果推論を中心に解説しました。Pearl流については黒木先生が書かれている本などが詳しいので参考にされるとよいと思います。これらは細かいところに違いはありますが、根本的に違うものという訳ではありません。</p>
78	<p>ベイズ統計と因果推論について少しコメントいただければ幸いです。ベイズ統計では不偏性についてほとんど議論されていないようなので因果推論と相性が悪いように思いますが、いかがでしょうか？（今回のセミナーのそのものについての質問でなく恐縮です）</p>	<p>高橋先生：どのようなベ이지アンを考えているかにもよります。例えば因果推論の大御所であるRubinはベ이지アンとして認識されていますが、Rubinはベイズ以外を認めていないわけではありません。多重代入はベイズに基づく考え方ですが、そうして代入されたデータをもとに頻度論的な分析を適用することも可能であり、そういった意味で親和性も高いです。</p> <p>なお基本的に潜在的結果変数の考え方はベイズ的であり、理屈では欠測部分をベイズ的に多重代入しています。ただし個別の手法にフォーカスするとそれほどベイズ的ではないかもしれません。</p> <p>書籍でも「統計的因果推論としての多重代入」を紹介しており、これはベイズ的に因果推論しているといえます。</p>
79	<p>本日のお話を踏まえて、思考を整理させてください。</p> <p>前提として、処置が確定的に決まる場合は、観測・未観測を問わず共変量でいくら条件づけようとも条件付き正直性は担保されないため、共分散分析・傾向スコアマッチング・操作変数法のいずれもアプローチとしては不適切となります。これが、シャープな回帰不連続デザインを選択する主なモチベーションと理解しております。</p> <p>一方、ファジーな回帰不連続デザインはそもそもが処置の割付が強制変数Xにより確率的に決まる点から、Xを条件付けさえすれば、閾値周辺に限って言えば（?）、条件付き正直性が担保されるため、処置変数の係数を正確に推定するためには、共変量が全てわかっているならば傾向スコアマッチング、未観測の共変量があれば操作変数という形で分岐する理解です。</p> <p>上記の理解で間違っている箇所があればご指摘いただけますと幸いです。</p>	<p>高橋先生の回答：正しいと思います。</p>

80	<p>確認なのですが、処置変数と傾向スコアで結果変数を説明する場合、あくまでも注目すべきは処置変数の係数であって傾向スコアの係数は評価・議論しないという理解でよろしいでしょうか。また、縮約された傾向スコアに名称をつけることは通常しないのでしょうか。</p>	<p>高橋先生：前半はその通りです。傾向スコアの値がどの程度増えたら結果変数がどの程度増加するか、といった因果効果に関心がある場合は解釈しても良いかもしれませんが、傾向スコア自体が複数の共変量を縮約したもので、それ自体の解釈は難しいと思います。傾向スコアはあくまでも交絡を調整するためのものだということです。なお傾向スコアに名称をつけることはふつつありません。</p>
81	<p>大変わかりやすいご講義ありがとうございます。先生のご提案されたMIRDDについて可能な範囲でご紹介いただくことは可能でしょうか。</p>	<p>高橋先生の回答：時間がありましたら後で簡単に紹介します。</p>
82	<p>充実した講義をありがとうございました。処置の割り当てが他の質的な要因とほぼ同義の場合は、処置の効果を（他の要因と分離して）推定する手法はないように思えますが、そのような認識で正しいでしょうか？例えば、付属校からの推薦入試合格者（他の質的な要因）は自動的に全員受講する入学前教育（処置）が、入学後の学業成績等（結果変数）に与える影響などです。</p>	<p>高橋先生：その通りだと思います。統計的因果推論の仮定で述べたSUTVA（この例では隠れた処置＝付属校での教育）や「操作なくして因果なし」についてよく考えるべきだと思います。</p> <p>何の因果効果を推定しようとしているのか、何が操作されているのか、をきちんと考えた上でモデルを組み立てて推測を行う必要があります。</p>
83	<p>今回のレクチャーのトピックとは少しずれてしまうかもしれませんが、逆に閾値を求めたいとき（例えばヘビーユーザーとミドルユーザーの線引きをしたいとき、どの変数のどこの数値が閾値になっているかなど）に使用できる統計的手法などありましたらご教示いただけますと幸いです。</p>	<p>高橋先生の回答：これは分類の問題になるので、判別分析とか、いわゆる多変量解析系のものになると思います。</p>
84	<p>IV推定量は、ITT推定量を補正したものということでした。</p> <p>間違った理解の可能性はありますが、IV推定量は、天邪鬼がもたらすデータ（外れ値）の対策しているという印象を持っています。</p> <p>そもそも、共変量や結果変数の外れ値対策、または外れ値補正の研究などは行われていたりするのでしょうか？</p> <p>もしご存知でしたらご教示いただけると幸甚です。</p>	<p>高橋先生：（書籍ではページが足りず載せられていないですが）外れ値があることで回帰係数は大きく影響を受けるため、きちんと考慮した上で解析を行う必要があります。</p> <p>「欠測データ解析」p104では外れ値の検証をすべき、といった話をしています。</p> <p>ちなみに重回帰モデルの場合はXの影響力とYを両方考慮することで各データの外れ値度合いを測ることができます。ちなみに私はCookの距離をよく使っています。外れ値をどう扱うかについては、例えば除外したときにどの程度結果が変化するかを検証するといったことが考えられます。</p>

85	<p>ご講義によって、傾向スコアを用いた推定の方が単純な重回帰分析よりも適切な推定ができる可能性が高い、ということは理解できたように思います。他方で、アンケート調査などを使って分析する場合、リッカート尺度で測った5段階や7段階の変数であることが前提である場合が多く、原則として処置変数を2値変数として扱わなければならない傾向スコアの分析を一体どう使ったらいいのか、よくわからなくなりました。5段階の回答を2段階に縮減して当てはめたらいいだけなのかもしれませんが、それではせっかく得られている情報を捨てるようでもったいない気がします。処置変数を連続変数にしてもよい、傾向スコアを用いた分析方法についても、どこかで是非解説してもらいたいです。</p>	<p>高橋先生の回答：5段階の処置変数に対しては、理論上は順序ロジットとか順序プロビットで予測値を出して傾向スコアとして活用することを考えることはできると思います。私も詳しくないので、もっとよい方法があるかもしれません。</p>
87	<p>アカデミア以外で、因果推論が企業や公共機関で日常的に使われている例があれば教えてください (海外の事例でも結構です)</p>	<p>横から失礼します。Cyber agent等の企業では自前の経済研究チームを有しており、この辺りの研究成果を実務に応用しているという話を聞いたことがあります。特に広告業界では実験計画をしやすいため、いくつか事例があるようです。</p>
87	<p>アカデミア以外で、因果推論が企業や公共機関で日常的に使われている例があれば教えてください (海外の事例でも結構です)</p>	<p>高橋先生：みなさんご存知であれば教えてください。 最近では、回帰不連続デザインが厚労省のワーキングペーパーで使われたことが話題になりました。役人が因果推論の枠組みを学んで「やってみよう」と思ってくれたことは素晴らしいと思いました。これを出発点に、公的機関でももっと因果推論が活発になると良いと思いました。</p>
88	<p>講義6のP. 83の傾向スコアによるマッチドデータを使って重回帰モデルで推定した因果効果推定値についての「クラスタリング標準誤差」について一つ質問があります。P. 83では「マッチングの重み」をクラスターにして標準誤差を算出されていますが、これとは違うもの(例えば、企業や年)でクラスターとして修正した標準誤差を使っても問題ないのでしょうか。通常、実証分析の実務では、本日のような傾向スコアによる分析の他に、昨日講義されたような重回帰分析(共分散分析)も並行して行うのが普通と考えます。重回帰分析をする際、特にパネルデータを使っている場合、最初からクラスタリングされた標準誤差(例えば、企業や年をクラスターにしたtwo-way clusteringなど)を使うのが一般的になっているかと思います。こうした流れから、本日のような傾向スコアによる分析の際の標準誤差の計算においては、重回帰分析(共分散分析)と同じ方法で標準誤差を計算した方が整合的なように感じました。</p>	<p>高橋先生の回答：はっきりとは分かりませんが、どちらを優先してクラスタリングの標準誤差を使うのがよいかはひょっとすると今後の検討事項であるように感じます。傾向スコアモデリングでどういう標準誤差を使うのかというのは未だに論争がありますが、単に出力された結果を使うというのはまずいというのが現代のコンセンサスかと思いますので、書籍の執筆時点では、パッケージで簡単に求められる点なども考慮して、この方法を紹介しました。</p>

89	<p>先生の教科書の次ステップとして薦められるトピック・論文・本について、いくつか挙げていただけると有難いです。</p>	<p>高橋先生：</p> <ul style="list-style-type: none"> ・ローゼンバウム 統計的因果推論入門—観察研究とランダム化実験— （ローゼンバウム：岩崎訳）傾向スコアに関して詳しい https://www.kyoritsu-pub.co.jp/bookdetail/9784320114470 ・構造的因果モデルの基礎（黒木） Pearl流の解説の和書 https://www.kyoritsu-pub.co.jp/bookdetail/9784320113176 ・統計的因果推論（岩崎）高橋先生が学んだ本 https://www.amazon.co.jp/%E7%B5%B1%E8%A8%88%E7%9A%84%E5%9B%A0%E6%9E%9C%E6%8E%A8%E8%AB%96-%E7%B5%B1%E8%A8%88%E8%A7%A3%E6%9E%90%E3%82%B9%E3%82%BF%E3%83%B3%E3%83%80%E3%83%BC%E3%83%89-%E5%B2%A9%E5%B4%8E-%E5%AD%A6/dp/4254128576 <p>岩崎先生は最近放送大学でも講義を行っているので、「統計的因果推論の考え方と技術」を見てみるとよいかも</p>
----	--	---