

「政策立案と統計」主な質問への回答

2020年2月18日

山澤成康

講義での主な質問に回答しました。参考にしてください。

さらに質問などがあれば、yamasawa@atomi.ac.jpまでお願いします。

経済分析

主成分分析

主成分分析について詳細を知りたいです。わかりやすい説明は先生のリンクに貼られていますでしょうか。

主成分分析についてもう少し詳しく説明していただきたいです。

主成分分析は学んだことがないので、わかりやすい入門書等あればご紹介いただきたいです。

「主成分分析」、「多変量解析」で検索すると、さまざまな文献があります。

私は、有馬哲・石村貞夫『多変量解析のはなし』を読んでいました。

また、金明哲「Rによるデータサイエンス(第2版)」などを参考にしますが、専門的かもしれません。

<https://www.morikita.co.jp/books/book/3063>

・統計アドインのない Excel のみの環境下で重回帰分析を行う場合、主成分分析等ができないため変数が膨大で、詳しい統計結果を具体的に政策まで落とし込むことが困難かと思われれます。どのように対処すればよろしいでしょうか。

→エクセルだと、変数の入れ替え、サンプルサイズの変更などが簡単にできません。実務でかなり最小二乗法を使う場合は、EViewsなどの統計分析ソフトを使う方がよいと思います。

主成分分析について、括り方には何か方法はあるのでしょうか。

単純に合理的であれば、分析する人の主観で括ってしまってもよいのでしょうか。

→機械的に分析するものなので、主観的に似たような指標を選んで合成していると思います。

・静岡県の事例の中で、要因分解についてもう少し詳しく説明していただきたいです。

→静岡県の資料を参考にいただければよいと思います。

https://www.pref.shizuoka.jp/kousei/ko-130/documents/00rashinban_c2.pdf

これまでも e-stat から統計データを取得し、重回帰分析等によりデータ分析を行ってきましたが、統計データの加工に時間を要します。効率的に加工するコツがあればお伺いしたいです。

→元になる統計について詳しくなることが近道だと思います。

因果関係が階層的な場合における主成分分析と重回帰分析の組み合わせについて、なぜ組み合わせる必要があるのか、もう少し詳しく説明が欲しかったです。

→主成分分析で指標を一つにまとめないと、多重共線性の問題が生じることが大きいと思います。

回帰分析

重回帰分析を利用して評価を行う際、多重共線性の確認以外に気を付けることはあるでしょうか（気軽に Excel の機能を利用してよいものか心配です）。

回帰分析の表の他の項目についてはそこまで重要視する必要はないという理解でよいのでしょうか。（自由度、有意 F、下限 95% など）

→それぞれ意味がありますが、最初は気にしなくても大丈夫だと思います。深く調べたい場合は「計量経済学」の教科書を読むことをお勧めします。

重回帰の際の VIF は単回帰ではなく説明変数への他の説明変数の重相関係数に基づき計算するものだったように思うのですが、単相関で一つ一つを見ていっても同じ結果になるのでしょうか。

→説明変数どうしに相関関係があるかどうかを総当たりで見るということです。多重共線性の問題を避けるという意味では、被説明変数との相関係数はとらなくてもいいです。

因子間の関係性が非線形の場合は重回帰分析が適用できるか否か疑問に感じた。変数の次元に関係なく適用できるのか、もしくは、関係性を事前に線形と判断はする数式などがあるのか疑問に感じた。

→変数自体を対数にしたり二乗したりして対応できる場合があります。

重回帰分析や相関係数を見るときのポイント（この部分の数値がいくつ以上なら有意等）をもう一度教えてください。細かいことは説明を聞いてもわかりませんが、そこさえ分かれば、一応実務で使えると思います。

「t 値」等の定義や妥当な範囲について覚える必要があるのか判断がつかなかったため、覚えるべきものは教えていただけるとありがたいです。

t 値は 2 以上、決定係数は 0.8 以上（諸説あり）くらいでしょうか。

分析全般

葉山町の例のように%で表す場合と、不法投棄件数（%にせず実際の件数のまま）で表す場合とがあると思うが、どちらの表現でも問題はないのでしょうか？（見せ方の違いだけで、統計学的にはどちらでも問題ないのでしょうか）

→量が大事な場合と比率が大事な場合があり、ケースバイケースだと思います。

政策立案

政策効果を金額で示す手法は確立されていますか。

→環境への負荷などを金額で表す手法などさまざまなアプローチがあると思います。

https://www.jica.go.jp/jica-ri/IFIC_and_JBICI-Studies/jica-ri/publication/archives/jica/kyakuin/pdf/200503_eva_02.pdf

統計データを集めるために正しい知識とある程度の労力があることを理解した。結局、政策の立案（要素の洗い出し）は人間が行って、それらの係数決定のために統計を利用しているかのような印象を受けたが、実際のところどうなのであろうか。

→政策立案自体は人間がするもので、ロジックモデルなどはそれを実現するためのツールだと思います。

・直接アウトカムの目標を実現するためのアウトプットの目標はどのように設定したのでしょうか。そこも回帰分析で係数を計算したのでしょうか。

→今回は、目標値も実現値も仮想値で、計算で求めたものではありません。

統計改革推進会議や行革本部など EBPM を扱う国の機関がいくつかあるように思えますが、現状として EBPM の所管はどの機関（所属）なのでしょう
か？

→EBPM推進委員会の事務局である、内閣官房行政改革推進本部事務局ではないでしょうか。

・求めようとしている効果について、外部からの影響を考慮して測定する手法はあるのか。

→外部からの影響という意味がよくわかりません。

ロジックモデルとロジックツリーの違いがわからない。

→ロジックツリーは因果関係の繋がりを表すもので、ロジックモデルはロジックツリーを使って、インプットからアウトカムまでの一連の流れを示したものだと思います。

ロジックツリーは課題の抽出、ロジックモデルは目標達成のための手順のよう
で裏返してるだけに感じられた。

→ロジックツリーを使ってロジックモデルが作られていると解釈しています。

市でデータ分析する場合、「マンション開発が見込まれるから」など、幹部からの指示で恋意的な見立て（仮定）を取らざるを得ない場合が多い。

ふじのくに少子化突破戦略事業においては、主成分分析と重回帰分析が上手く併用されているように感じたが、手順としてどちらを先に実施したほうが良いとかあるのか？

基本は回帰分析だと思います。

数値が出来ない施策があった場合はどうなるのか？

なんとか数字で表す努力をするということでしょうか。

係数と目標値と切片の関係をもう少し詳細に学びたかったです。

目標設定が、なぜ「各係数×目標値の和+切片」となるのかすぐに繋がらな
かった。「係数」や「切片」の意味が分かっていないように思うので、もう少し考えてみたい。

→テキストの51頁の囲みである部分最下段の式を参考にしてください。

重回帰分析のような分析手法を具体的な政策立案に落とし込む際に留意すべきことは何でしょうか。

→適切なデータを選ぶこと、サンプルの数を確保することなどでしょうか。

ロジックモデルについてもう少し詳細に知りたいです。(ロジカルシンキングと考え方は似ている?)ロジックモデルを作成するにあたりLINKに載せていただいているもの以外にも参考になるものがあればお伺いしたいです。

→LINKにある論文の参考文献をたどることで、さまざまな文献が見つかると思います。

中間アウトカムの目標値の分析の仕方で、最終的アウトカムの目案・演習目標値についても計算を行えるという理解でよろしいでしょうか。

→手法は同じだと思います。

特に健康関連の施策などは効果の発現・検証までのスパンが長く、また個人の生活習慣や環境など施策以外の変数が多いため因果関係の見極めが極めて難しいと考えているのですが、こういった分野でEBPMによる政策立案と評価を実践するコツや確立した手法などないでしょうか。

例えばこの因果関係の有無を確認するには、実際に検証するしか方法はないのでしょうか。

→詳しくはわかりませんが、あると思います。

目標達成かどうかを判断する際に、未婚率であれば、同じ数式に入れた際の現状の推計値と実際の値をどう使うべきでしょうか。あくまで回帰分析から求められる数字同士の比較となるのでしょうかが実際には、実績値で見ることが多く、2つ数字が存在することが誤解を生むことが多く、悩ましいです。

→実際の値優先だと思います。政策効果は、推計値どうしを比べて、実際の値に加えればよいと思います。

ランダム比較実験

ランダム化比較実験は医療以外（行政など）で何か例はあるのでしょうか。
・ランダム化比較試験が最もエビデンスが高いことは理解できますが、行政において比較的容易に導入する方法はありますか。

→米国などの例がさまざまな文献に載っています。

伊藤光一朗 「データ分析の力 因果関係に迫る思考法」 光文社新書

・経済産業省資源エネルギー庁が主導した社会システム実証実験

中室牧子・津川友介「「原因と結果」の経済学データから真実を見抜く思考法」
ダイヤモンド社

田中隆一「計量経済学の第一歩ー実証分析のススメ」 有斐閣ストゥディア

森田果「実証分析入門ーデータから「因果関係」を読み解く作法」 日本評論社

アングリスト・ピスケ「ほとんど無害な計量経済学」

・職業訓練プログラムがその後の賃金に与える効果

実務上、最初に悩むのは無数の統計指標から何を用いるべきかということですが、先行研究に当たるといのが最も効率的な方法になるのでしょうか。

→先行研究に当たるといのは、良い方法だと思います。あとは統計書を眺めることでしょうか。

DIDのご説明をいただく前に、RCT等の説明(76. 77)を先に頂戴するとより分かりやすかったかもしれません。これを前提にベースラインを揃える必要があるが、それを解決する手段としてのDIDだと思いますので。

→検討してみます。

テキスト 79 ページの棒グラフ「看板の効果の持続力」の描画の仕方は正しいのか。縦軸のとり方を工夫して、効果を大きく見せているように感じる。

不法投棄の1割減が目標であったならば確かに効果的であるのかもしれないが…。

また、チラシと看板はどちらが効果的かという点に関して、比較・検討があったのかが疑問である。

→さらに詳しい資料があるので、こちらをご覧ください。

<https://www.town.hayama.lg.jp/soshiki/seisaku/osirase/6870.html>

演習

各演習が何を学ばせるためのものなのかガイドが欲しかったと感じます。演習の課題内容を、レジュメにのせておく等してもらわないと、何をしたらいいのか分からなくなる。

演習に最終目標があるなら、その流れを最初に示して欲しい。細切れに順番にやっていくと、今やっている作業の意味が分からない。

統計関連

調査票データとはマイクロデータのことでしょうか。

→そうです。

都道府県別の出し方が、課題では合計特殊出生率以外のやり方がわかりませんでした。

→追加で共有フォルダーに上げたものを見てください。

e-Statで統計データを閲覧した際にDB化されていないCSVのみの表も多く、こういった点も経団連から指摘（提言）を受けているのでしょうか？

→経団連は指摘してないかもしれませんが、一般的に機械判読できないPDFなどを統計に使うことは良くないとされています。

質問：都道府県別のところは合っているが、合計のところ为正解データと合わない。

→今回の演習では合計は使ってないです。

人口動態統計4-21表について、年次が2017年か2016年のものしかなかった。何を間違えたのか知りたい。

<https://www.mhlw.go.jp/toukei/list/81-1a.html>

上のURLから探せると思います。

統計数値を調べたいときどこに載っているかを概ね把握するための覚え方のコツはあるのでしょうか？ありましたら、アドバイスをいただきたいです。

→年鑑などの一覧性のあるものを見ると効果があると思います。

現状を把握できるデータが統計等でない場合、新たに調査等でデータを入力する際の注意点はありますか。

→統計作成と同様だと思います。標本の無作為抽出が最も大事だと思います。

エクセル

質問:SmartArtの階層レベルの上下の仕方がよく分からない。

→「テキストウインドウ」を表示するとわかりやすいと思います。

今後の課題

こういう分析の時は、この手法を用いるといったルールなどがありますか？（もしくはわかりやすく紹介されているサイトなどがありますか）政策評価のロジックモデルの作り方で悩むので、その時の考え方をいくつかの政策を例に話していただけると実務的だと思いました。

→今後、さまざまな例を紹介できるようにしたいと思います。

総務省の表彰制度「地方公共団体における統計データ利活用表彰」例がいくつか載っています。

https://www.stat.go.jp/info/guide/rikatsuyou/index_2019.html

感想など

→受講者によって難しいと感じる人と易しいと感じる人がおり、両者に満足してもらえるような授業運営が課題だと思っています。

後の演習部分が駆け足になったところもあり、難しく感じられた。
ロジックモデルにおける、論理的な課題抽出・目標設定を行いそれをどう統計上の数字に結び付けるかがやっと繋がった。
机上の空論な感じを受けている部分も多く、実例をいくつか見てみたい。
また、e-STATのDBの使い方をご紹介頂いたのは良かったです。
e-statの使い方より演習を優先してほしかった。
先に做った事項を演習の形で復習できた。特にe-statなどの資料探しの例が参考になった。
最後の目標値の設定について付いていくことができませんでした。重回帰分析の仕組み時間をいただけると嬉しいです。
エクセルの操作説明については、丁寧でわかりやすい解説であったがもう少し早く進行いただいてもよかったのかなと思った。
質問票は良い方法だと思うが、それを読むために、聴講生が放置されているような状態になるなら本末転倒では。