

大西弘高 講演「医学教育研究とデータの種類, 統計解析」

当初の目標にしていたスライドをちょっと枚数を増やしています。いくつかこれは大事だなというふう
に後から思った分です。では始めます。

スライド2 まず, 研究デザインについて, これは今 追加資料をお配りしました。事後テスト単独,
事前事後テストどちらかを選択することになります。事後テスト単独というのも結構教育のところ
では行われているのですけれど, 事前事後やると改善度が測定可能なので, その差をとってと
いうことがよく行われます。それから対照群の有無ですね。対照群, コントロール群があると当
然比較ができますからこういうのを心理学の実験の領域では準実験 (quasi-experimental
design) といっています。それからもし対象群を設けるなら, それはランダム化されているのか,
割り付けがですねされているかということでランダム化されていれば真の実験 (true
experimental design) というふうに呼ばれます。

スライド3 それらを用いで事前事後, あるいは事後のみを組み合わせると, 非ランダム化で事後の
みというのはほとんど意味がないので, この5つがよく用いられているデザインです。医学教育
学会のいろんなワークショップでは事前事後やりましょうみたいな感じで随分言っているらしい
ですが, 事前事後だけをやってコントロール群を設けないのは余り意味がありません。前実験
的 (pre-experimental design) といって非常に弱いデザインです。何故かというこの2回の測定
の間で何が起ったのかが分からないということですね。ま, 当然学生は成長するというのもあり
ますし, そのときに例えば, 医療事故で大きなニュースとかがあればですね, 何か自分も行動
改めないといけないなというふうな感じになる学生もいるでしょうし, そういふうなことでこれは非
常に弱いデザインとされています。それに比べると, できる限りコントロールグループをおいた
方がいいということになるのですが, コントロール群をおくとどうしてもお互いに釣り合ったグル
ープなのかという問題も出てきますし, 分けるということになると「オレはコントロール群に入って
教育受けられないのか」ということも問題になります。特にそれはランダム化すると問題になりや
すいので, 教育学的な研究での一番の難しいところはですね, コントロール群をおいていいの
かなどうなのかな, もしおいていいのかなどうかということをして学生に訊いたときにどういう反応を示
すのか, というみたいなことを考えないといけない。そこが臨床試験と違って一番違うところ
です。

スライド4 統計学的な解析ですけれども, 記述統計あるいは分析的な統計というふうなこと2つに
分かれます。記述統計だけでも十分モノがいえるような研究というのたくさんあります。

スライド5 ここからは, 森本先生のスライドで言っていない部分を少し入れていますが, 連続変数,
あるいは順序変数, 名義変数, そこには二元と書いていますが, すべて二区分に直しました。

心理学とか教育学のところではすべて二区分変数という名前を使っています。それから連続変数、測定のデータの多くはこの形ということですが、実際には質問紙とか使うと、加算していいかどうかということを常に問わないといけません。もし、5ポイントのライカート・スケール(Likert scale)で10問あるような、そういう質問紙を考えたときに、満点が50という形にしてもいいのですが、その50というふうな数字ははたして妥当なのか、信頼性をもっているのか、そういうふうなことを問わないといけません。だから加算していいかどうかという話はまた後で出てきます。

スライド6 連続変数は事象を鋭敏に表現できますし、統計学的な係数が高いということ一番利点ですね。しかし、欠損値が発生しやすい。多変量に不利である。あまりこれは考えたことはなかったのですが、そういわれればそうかなと思います。分布と欠損値の頻度を常にチェックする。数学的な処理で順序化、カテゴリー化するとか二区分化するというのもロジスティック(logistic)回帰分析とかでは必要だということですね。余り僕はこれ考えたことはないですけど。

スライド7 順序変数は、連続変数と二区分変数の中間で、一見連続的だけれども、例えば何かの問いに対して、全くそう思う、そう思う、あまりそう思わない、全くそう思わないというこの4つのカテゴリーは順序はありますけど、間隔が等しいかどうかは分からない。こういうのが順序変数ですね。二区分よりも鋭敏に検出できるということでアンケートなんかではよく使われているわけですが、これも欠損が発生しやすい。

スライド8 名義変数、出身校みたいなものですね。大学を国立・公立・私立などに分けることもありますが、基本的に数字変換の不可能で統計処理が困難であると、ダミーコード化しないといけませんというのがあります。

スライド9 ダミーコード化というのはどのようなものかという、こんな感じですね。公立高校・私立高校・国立附属、帰国子女というのを4つのグループに分けようとするので3つのダミーコードを設けてそれぞれ000, 100とかですね、このように名義変数の種類の数-1だけのダミーコードをおけばいいということになります。

スライド10 二区分変数、イベントの評価、あるかないかですね。欠損値が少ないです。解析や結果の解釈が容易。二区分変数はテーブルに書くということになりますけど鋭敏さに欠けるということ。2変数を組み合わせてもいい、こういうのもでデータを取って統計解析にする。

スライド11 例えば平均値の差を見るという場合には、2標本か3以上の標本数があるのかということで、対応がないのかあるのかということも見ないといけません。パラメトリックを使うならこんなもの、ノンパラメトリックだったらこんなもの、というふうなことになります。ここのt検定は対応のある

ものとならないものでやり方が違います。

スライド12 なんとなくこの話題になってくると、ノンパラというのが何でもかんでもできて安全でいいんだという議論に割と医学領域ではなっていたことがあります。心理とか教育学の方では全くそういう方向には進んでいませんで、パラメトリックがいいというふうな意見が非常に多いです。ノンパラメトリックの手法は、順序とか名義変数に利用可能だということですね。パラメトリックという手法は正規分布とか等分布を仮定している、いろんな制限が多いわけですが、ノンパラメトリックの手法は分布型を問わない。しかし検出力が劣る傾向があり、計算も難しい。検出力は劣るかどうかわからない、やってみるとあまり劣らないということが多いです。計算が難しい、どちらにしてもソフトにいれば簡単です。ただ、多変量とかですね複雑なことができません。それとパワーの計算もできません。ですから、研究計画がうまく書けません。そういうふうなことで、なかなか皆さんあまり使われないようですね。それとnが30を超えるような場合、教育学の研究だと大体nが100とか1クラス全部とかやりますから、30なんか超えることが多いのですけれど、だったらパラメトリックの手法でも頑健性(robustness)というのがあって、大丈夫という話しになります。

スライド13 p値とEffect Size, p値というのは帰無仮説が成り立つ確率ということですね。これが0.5よりも小さければ、帰無仮説が却下されて、棄却されてということですが、Effect Size というのは、差が大きいかどうか平均値を見たときその差が大きいかどうかを見ていきます。基本的にはstandard deviation(SD)との比率がstandard deviation(SD)分、1コずれていますよということEffect Size は1みたいなのが大体の目安ですね。Effect Size が0.2ズレているというのはsmall, 0.5だとmedium, 0.8だとlarge, というこんなちよつとした指標も出ています。特にこの教育学の領域では、Effect Sizeを非常に皆さん重要視されまして、これって要するにNNTが大きいかわ小さいかわですね。NNTが小さいということは、Effect Sizeが大きいということですから、医学での疫学における議論と似ている部分があります。p値というのは、サンプルサイズが大きければ有意になることがありますから、Effect Sizeが0.1ぐらいでも有意差ありという判断もありうるわけです。サンプルサイズが小さくても有意に出る場合、非常にEffect Sizeが大きいということがいえま

スライド14 分散分析, ANOVA ですね。まず重要なのは要因をみた時に反復測定の有無を間違えない、反復測定というのは、繰り返しのあるデータかどうか、こういうふうな先程森本先生が言っていたデータセットがうまく組んでいるかどうか、これをみると大体その人が分析に慣れているかどうか分かります。たとえばこれですね、評価者1の測定と評価者2の測定を同じID1の学生に2回測定しているわけですが、これは反復測定をしているわけです。そういう評価者1の点と評価者2の点を比べるという分析をやろうと思うと反復測定あり、という要因になります。もう一つ男女の要因がここにありますが、男女の要因は繰り返しがありませんから、縦に積み重ね

ています。こういうふうに繰り返しのないデータは縦に並べて、繰り返しのあるデータは横に並べる、ということになります。まず4つの測定を全部評価者1と2に分けてプロットしてみるとこういうふうになるわけですがそれを男女でまとめるとこういう感じになります。これが ANOVA の一番の基本形です。交互作用なしとかありとかということが分かります。これが交互作用なしの例になります。

スライド15 こういうふうに 2way、3way の ANOVA をやりますと同時に2つ以上の要因の比較ができます。繰り返しはあってもなくても組み合わせることは可能です。ただし、繰り返しのある場合は、球面性(sphericity)ということが問題になります。この辺の議論に興味のある人いますか。ああ訳分かんないですか、はい いいですね。二元配置、三元配置の場合はこういう交互作用というのを確認する必要があります。評価者1が測定したのと、評価者2が測定したのが平行になっていけば交互作用なし、全然逆の測定をしちゃうというのが交互作用あり、といういい方になります。あんまり、医学研究とかで ANOVA やっている、ANOVA が得意だという人はあんまり聞いたことがないのでこれ議論すると皆さん大体ああ〜という顔をしています。

スライド16 研究計画法というのが ANOVA の複雑なデザインを組むときに非常に重要になります。3要因で要因1があるかないか要因2があるかないか要因3があるかないかこれを組み合わせるとグループ8つに分けないとうまくいかない。グループ8つというのが、もし 80 人学生がいたら 10 人ずつのグループに分けるとのことですよね。これをランダムに分けるとなるとなかなか大変ですが、こういうふうにとると、要因1と2と1と3、3、..こういう交互作用なんかもすべて求めますが、3つの交互作用というのは何に意味しているのかはよく分からなくなってきました。この場合はいずれも繰り返しがなしのパターンですね。

スライド17 カテゴリー間、二変数間の関係:ここまで ANOVA とかそういう差をみる検定をやってきましたけど、ここからは少し相関とか回帰の話になります。二区分変数がたくさん出てきてその変数間の関係をみようというときには、まず、名義変数と同様に表に書いて、これをクロステーブルといいますけれども、お互いの関係を見るということが大事です。独立性は χ^2 検定で分かりますし、要因による差というのは、odds ratio あるいは relative risk みたいな形で出てきます。順序変数、連続変数だったら、まず、散布図を書くと、ほとんどそのクロステーブルと同じような意味合いですけれども、散布図を書いてこちらがX軸、こちらがY軸でプロットしていったら、大体こう右上がりだとか右下がりだとかをみることで相関しているかどうか分かります。一応順序変数だとピアマンで、連続変数だとピアソンということになりますが、これをパラメトリック、ノンパラメトリックの議論なのでどっちでもいい、ということにもなります。3つ以上の変数があっても全部関係見たい場合は多変量解析に行くことになります。

スライド18 多変量、どうやるんでしょうね。という質問になってくるとこれ非常になんとなく哲学的な

問いになることもあるんですが、一番大事なのはまずこのたくさん変数があるって、探索的に何かと何かの間に関係がないかな…みたいなですね、そういうものをやるのか、あるいは自分であくまでも仮説をもって、これ証明したいんだというふうに行くのか、まずそこからですね。ときどきこれの区別がつかずに闇雲に有意になるのがありませんかと持ってこられる、最近はないと思いますけど、昔そういうのがありました。この辺の議論をするためには従属変数は何か、独立変数は何かみたいなことが明確になっているかどうかですね、ただときどきこれに対してはこれが独立変数だ、これに対してはこれが従属変数だと、思っているけどまだ次のステップにいくと、関係が変わることは十分あると思います。これらによって、関連性と因果関係との違いみたいな話になってきます。何が原因で何が結果なのか、このときにデータの種類によっていろいろ解析の仕方が変わってきます。ちなみに解析の細かい話はしません。

スライド19 それから内的一貫性の話ですね、信頼性というか、質問紙を作ってみたときに、例えば1番～5番というか、a～eの項目が同じような内容を質問しているかどうか、ということが信頼性とか内的一貫性の話です。よくアンケートみると同じ様な項目訊いているなあ、でもちよつと違うんだけどなあ、というふうなものも並んでいる場合があります。それは信頼性を上げるためにわざと似たような質問をいっぱい並べているのです。ときどきこういうdみたいにまったく逆の質問を入れておいて、答えが5—5—5—5—5と並ばないように5—5—5—1—5になるように工夫されてることが多いです。alpha 係数、先程 0.8 位がいいという話がありましたが、0.9 を超えてくるとあまりにも同じような質問ばかり並んでいてトートロジー(tautology)というかそれはやりすぎだよということになってきます。こういうふうにdをマイナスの方にもって行って一つの方向にまとめあげる。こういうふうにやると、たとえば5ポイントのライカート・スケールで5つの問題に対して 25 点が満点で全部足し合わせてもいいというふうなことを alpha 係数を求めて、0.7 とか 8とかの結果が出たら、よし出してやろうということになります。

スライド20 こういうことに興味持っておられるかどうかわかりませんが、今教育とか心理の中で一番注目されているのはこの共分散構造分析というか structural equation models というもので、SEM という言葉が出てきます。確認的因子分析といって、昔は因子分析というと非常に探索的なものというふうなイメージだったのですが、もう因子の数が決まっている。それでうまく適合するかどうかというふうなことを確認する分析が今非常に流行っています。この辺僕も書いてきただけで分かりません。ほとんどやったことはありません。因果関係がはっきり議論できるというようなことが一番の特徴だといっていいと思います。パス(path)解析も因果関係で議論しますがSEMには潜在変数というのを置くといわれていますが、こんなものらしいですね。

スライド21 小学校の国語算数、中学校の国語算数英語の点数とそれから小学校の学力中学校の学力、動機づけみたいなものを全部纏めたモデルを作るとこういうふうな潜在変数が3つくらい出てきて、それぞれは最終的に中学校の実力に結びついていくと、動機づけはここで重要

な役割を果たしている、そういうスタディ法を組んだみたいです。

スライド22 それから先程から信頼性、妥当性の話がいっぱい出てきていました。この次ですね、small group discussion でたとえば質問紙使うというグループ、質問紙を実際に書いて欲しいと思うのですが、評価者間の一致度とか一致率、 κ 値、ICC (intra class correlation) こういったもので一応表すことになっています。ただ一致率と κ 値あるいは一致率と ICC のどっちか2つの組み合わせでなければ一致率の判断は難しいということになっています。とくにその教育学のなかで使われる評価 OSCE なんかであれば、“挨拶ができる”みたいな項目は非常に一致率が高そうに見えるんですけど、ほとんどの人は挨拶しますから単に評価者の一致率だけみたら99%一致しています。ただ一人がずれると κ 値0ということもありえるんですね。そういうふうなことで κ 値が使いにくい状況があるというふうな論文も私と一緒に研究している人が出したことがあります。それから今注目されているのは一般化可能性理論を用いたこういう係数ですね。信頼度指数というのは、それとほとんど似たようなものですけど前者は相対評価、後者は絶対評価に利用されることになっています。統計学的には一般化可能性理論が分かっていたらクロンバック α とか ICC というのはその延長上にあるというかその中に包含されるような関係です。

スライド23 回帰分析、ほとんど触れませんが単回帰、あるいは重回帰というような形で独立変数によって従属変数の予測式をつくるのが回帰分析のやり方です。それぞれが意味のある独立変数なのかどうなのかというふうなことですね。間隔変数が従属変数なら線形回帰分析ですし、従属変数が二区分だったらロジスティック回帰ということになります。ここがカテゴリーだと悩ましいのですが、間隔変数とみなせれば線形回帰にすることもあるし、むりやり二区分にしてロジスティック回帰とかあるいはカテゴリー変数の非常にややこしいロジスティック回帰を用いるプロビット(probit)分析を用いるという方法もあるようです。あんまり詳しいことは訊かないで下さい。

スライド24 SPSS では、こういうふうな Analyze の種類がざーっと出てきます。明日統計解析をするときには一応、この辺に SPSS を使えるコンピュータを置いておいて、どうしても使いたい人が利用できるようにしておきたいと思います。

スライド25, 26 便利な Web をあげてみました。僕が使うのはこの辺ということでもっともっとほとんど統計の勉強というのは Web の上でできちゃいますから、そういうのを是非利用して頂けたらと思います。

あまり時間がないので、ここでは質問を取らずにいきたいと思いますが、次ですね 40 分間くらいのところ、もう少し、具体的な測定の内容とかあるいは実際に質問票を使うのであればそれを実際書いてみて、それがどのくらい難しいのかとかあるいは信頼性のあるものにするというのが、どういう意味なのかそういうふうなことを是非知っていただけたらと思います。課題についての質問は

ありますか？スモールグループの、大体やることはこんな感じかなということ見えていますか？それぞれタスクフォースが行って説明するということになると思います。それでは宜しく願い致します。
6時 20 分までです。