

Q & A

統計処理法講座に対する質問と解答

Q 日本生理学雑誌3号の統計処理法講座を大変興味深く拝見しました。この中で、著者の栗谷先生が「その過ちを指摘されてもすぐには納得しないのが普通である」と書かれているとおり、私もよく納得できない事がありましたので、質問させていただきます。他の諸先生方からの反応や同様の質問があるなど、栗谷先生には恐らく大きなご負担をおかけすることかと存じますが、是非お教えくだされば幸いです。

質問は、1点です。同一人のデータを複数回数える誤りについて、小錦の例はよく理解できますが、例えば誤用例1-1では、たとえ同一人であっても平熱のときと、発熱したときと2回測って、2つのデータとしても、鼓膜温と直腸温の相関関係を分析するのに問題ないのではないのでしょうか。あるいは、これは「個々の測定対象患者によって回帰直線が異なる」ことを前提としていますが、個々の患者を多数回測定し、この仮定が統計的に満たされたら、上記のように同一患者の異なる2点（あるいはもっと多数）を別データとして扱ってもよいのでしょうか。

同じようなことは誤用例1-4や1-5でも言えると思うのですが、いかがでしょうか。

以上の点、お教えいただければ幸いです。

白山 武 司(京都府立医科大学 第2内科)

A 同一人のデータを複数回採用することの誤り「日本生理誌、(1996)58, 127-132」に対する質問についてお答えします。

小錦の話（小錦の体重のデータを複数回採用した集団の平均値を求める誤り）は理解できるが「誤用例1-1の相関関係の分析の場合はかまわないのではないか。すなわち、たとえ同一人であっても平熱のときと、発熱したときと2回測って、2つのデータとしても、相関関係を分析するのに問題ないので

はないか？」と言う疑問であります。その処理法がよくないとされる理由を3つ①、②、③あげてみます。

まず小錦の例は理解できたと思われた点について考え直す必要があります。小錦を例にとりあげたのは、症例によりデータの採用回数に差があると症例毎のウェイトが不均一になり、多くのデータを提供した症例に平均値が引きずられて正しい値からかけ離れると言うことを理解してもらうのが一つの目的でした。この目的は理解して戴けたようです。ところが、もひとつ大事なことがあります。

① それは、平均値を求めるために集めた標本のデータは、すべて独立データ (independent data) であることが絶対の条件であります。10個のデータの中の3個が対応のある (dependent data) であることは、対応のあるデータと対応のないデータが混在しているため平均値、SD などすべての統計量を求める素材として失格です。この素材から算出した値は数学的に平均値として使えないのです。小錦の例で、もし小錦の体重が60 kg だったとしたら、その平均値は普通の値に落ちついていたでしょう。でもこの平均値はやはり使えない値に変わりは有りません。データの提供回数に差があると、その症例の数値に引っ張られて歪むことの弊害よりも、全データの独立性の方がより重要なのです。もし小錦の体重が一般人と同じだった場合、平均値を引っ張られないが、平均値付近の値が増加しSDが小さい方へ引きずられることになります。相関係数の演算の場合と平均値±標準偏差の演算で症例数とデータ数のアンバランスは具合の悪い点で違いはありません。

② 母集団から標本を選び出す手続きのことを標本抽出法と言いますが、この抽出法に欠陥があった場合は集計は致しません。標本抽出法は目的により、多種類ありますが重複登録を許した抽出法は一つもありません。

③ 統計手法の中で、絶対に許されない手法の条件

に“統計でウソをつく法”に利用できるテクニックというのがあります。もしデータの重複登録が許されると、平均値も検定もすべての統計量が任意に調整できることになり、悪用が自由自在となります。

もし、平熱時、発熱時、の2時点の相関を調べたいのなら全症例の平熱時、発熱時の2回測定しそれぞれの相関を調べ、2つの相関係数の比較検定を行

なう必要があります。検定の結果を確認した後、平熱時と発熱時のデータを一緒にできるか、できないか、決定するのが正しい処理法です。あるデータは(平熱時+発熱時)、あるデータは1データ、あるデータは重複登録となると、もはや科学とは言えなくなっています。

栗谷典量(久留米大学小児科)