

古典的なノンパラメトリック検定はなにを仮定しているのか

1. 前置き

もともと工学部出身のぼくが、医学系の研究室に行っておもしろかったことはいろいろある。その中の一つが、医学系の人アーチファクト (artifact) という言葉の使い方だ。アーチファクトというのは、辞書的には人工物とか工芸品を指す。ぼく感覚からしたら「この結果はアーチファクトだよ」と言われたら褒められているのかな？ みたいな感じがする。でも違う。医学系の言葉使いではアーチファクトというのは、自然現象とか生命現象の本質じゃない人工的な混ざりもの、みたいなイメージだ。

ところで、ぼくは統計モデリングというのが好きだ。モデリングというのはデータを取った人からいろいろ話を聞いて、関数とか確率分布とか微分方程式とか差分方程式とかを組み合わせて、データを取った人のイメージとか目的とかを統計の言葉に翻訳するような作業だ。モデルを作るというのは、ぼく感覚では仮定を置くことに等しい。それはもうアーチファクトのかたまりみたいな作業だ。

そのせいかどうかは知らないが、医学系の人には検定が好きな印象がある。統計屋というと、いろんな検定を知っていて、場合に応じて正しい検定を選べる人、みたいな認識をされることもある。

なかでも、医学系の人にはノンパラメトリック検定が好きなようだ。

ノンパラメトリック検定とは、母集団に対して特定の確率分布を仮定しないで検定をする手法の総称とされる。

仮定が少ないほうが客観的な分析でえらいという感覚があるんだと思う。

そこがぼくの好みと相容れない。

でも好みじゃないのでノンパラメトリック検定に関する相談には乗りませんというのも大人げないので、これからちょっとノンパラメトリック検定について勉強していきたいと思っている。

この文書はそういったものだ。

2. ノンパラメトリック・モデル

ノンパラメトリックモデルという言葉とノンパラメトリック検定という言葉はまったく別物だと思ったほうがよさそうだ。

ノンパラメトリックモデルというのは、パラメータが多すぎるモデルのことを指す。

データのサイズにほぼ比例して、モデルのパラメータが増えるようなモデルはノンパラメトリックモデルと呼ばれる。

以下では、ノンパラメトリックモデルの話はしない。

ノンパラメトリック検定の話をする。

3. 符号検定

検定というのは帰無分布を一つ決めて、そこからのずれを測っている。

帰無分布を決めなきゃいけないのに、「特定の確率分布を仮定しない」なんてことができるのか。

どうやらできるようだ.

データを生成した分布が連続型の分布で, タイ (同順位) がないということだけを仮定しよう.