

❁ 医療職のための統計シリーズ

医療職のための学び直し—研究デザインから論文報告までの生物統計学の道標—
第7回 推測統計学の基礎

ウヰムラ コウヘイ
上村 鋼平*

I はじめに

第5回¹⁾と第6回²⁾の連載では、一般的な臨床論文³⁾の表1や図に掲載される数値の算出に必要となる、対象データの特徴を記述するための方法等について学んだ。第7回では、推測統計学の基礎について学ぶ。推測統計は、論文の表2や表3あたりに示される、研究仮説に直結する主要な解析結果や副次的な項目やサブグループにおける解析結果を示すために用いられる手法に該当する。例えば、あるワクチンの疾患発症予防効果を調べるための研究を行ったとする。ワクチン投与群と非投与群の患者背景の集計結果が、論文の表1に入される。これは記述統計を用いた結果である。表2では、各群の疾患の発症数とその割合、群間比較の指標として割合の比（リスク比）の点推定値とその95%信頼区間とp値といった推測統計に基づく数値が、主要な解析結果として示される。表3以降では、重症化した人数や退院までの時間等の副次的な項目による解析結果、男女別や年齢層別などのサブグループ解析結果が、表2と同様に群間比較の指標とその95%信頼区間とp値（p値は省略されることもある）といった数値とともに示されるだろう。今回の連載では、研究結果の正しい解釈や研究計画（リサーチクエスチョンの選択やサンプルサイズ設計）のベースにある、推測統計の原理を解説する。

II 結果の信頼性を意味する結果のばらつき

推測統計はなぜ必要か。研究の主要な結果の報告として、リスク比やオッズ比の値を提示するだけではなぜ駄目なのか。例えば、通信販売である商品を購入しようとした際に、カスタ

表1 カスタマー調査の際の獲得星数のばらつき

	5件	10件	50件	100件	1000件	10000件
真の星数	4.5	4.5	4.5	4.5	4.5	4.5
調査1	4.8	4.3	4.5	4.5	4.5	4.5
調査2	3.8	4.5	4.5	4.5	4.5	4.5
調査3	4.6	4.3	4.5	4.5	4.5	4.5
調査4	4.8	4.8	4.6	4.4	4.5	4.5
調査5	4.0	4.1	4.5	4.5	4.5	4.5
調査6	3.8	4.6	4.4	4.6	4.5	4.5
調査7	4.4	4.6	4.5	4.6	4.5	4.5
調査8	4.6	4.4	4.5	4.5	4.5	4.5
調査9	5.0	4.0	4.5	4.6	4.5	4.5
調査10	5.0	4.7	4.5	4.4	4.5	4.5

注 レビュー件数（調査のサンプルサイズ）の異なるカスタマー調査を各々10回繰り返し、各カスタマーが1件ずつレビューを行った際の星数の平均

マー評価を一つ星で表したものを見かけるが、獲得した星の数（レビューを行ったカスタマーにおける平均値）が星5つ中4.8だけれどもレビュー件数が5件しかない商品と、レビュー件数が1000件で星5つ中4.6だった商品と、どちらの商品を安心して購入できるか。星4.8の商品の方がより評価が高いという値になっているが、たまたまレビューを行った5人のカスタマーが高い星数を報告しただけかもしれない。もっと多くのカスタマーがレビューを行った場合もう少し低い星数になるかもしれず、レビュー件数が多い商品の星数の値の方が、星数が若干低くても信頼性が高いと直感的に思う人も多いただろう。このように、評価人数が多いほど出た目の数の信頼性は高く、評価人数が低いほど出た目の数の信頼性が低いと感じるのは自然である。従って、推測統計を用い、出た目の数の信頼性を定量的に示すことが必要となる。

ここで、出た目の数値が実際にどの程度のばらつきを伴って観察されるのか、簡単なシミュレーション実験により確認してみよう（表1）。1000万人の潜在的なカスタマーの評価を平均すると、星5つ中4.5の評価をつけるような商品があったとする。つまり、この商品の星数の真

* 東京大学大学院情報学環（生物統計情報学講座）特任講師