

やさしい例による 統計入門 上

F. モステラー W. クラスカル R. リンク R. ピーターズ G. リージング 共編
村上正康 監訳 培風館刊

本書は、やさしい統計データを材料に初等的な手法で、あざやかにそれを料理して、統計学の面白さを読者に満喫させてくれる。使用されている例は人口統計、バスケットボールの得点と反則数の相関を示すデータ、糖尿病の遺伝に関するデータ、一家庭の子供の数と平均値の性質、大学対抗フットボールの得点の勝チームと負チームの相関表、テーブルへのすわり方、ビールの味覚テスト、トーマスベイン(200年前の人物)の犯した偏りの多いデータ・サンプリングの例、しょうじょうバエの目の大きさの遺伝といったものである。これらの実例は皆それ自体で読者の興味をそそるが、これらのデータを使い、統計データの作られ方、代表値の見方を(平均値、メジアン、モードの見方、使い方、中には間違った使い方をその間違っている理由を明らかにしながら)示している。また、縦横にグラフ化をして、グラフの中に手で適当に線を引くことによって原点を通る直線回帰とか、一般の単回帰なども、理論はまったく述べないで、唯わずかに、残差が、直線と点との距離をy軸に平行な方向で測ったものであることを、実際に手で書き込んだ直線と点の間で測らせて、その2乗和の小さいほうの直線がよりよく適合した直線だという形で説明してある。以上で本書の特色は大体ご理解いただけたかと思うので、以下に本書の内容をもう少し詳しく説明したい。

カバーする確率統計の分野 この本は在来の普通のテキストとは違って、理論や手法解説の前後関係をあまり気にせず、読者の常識と直感によってわかる範囲で説明がされている。棒グラフ・時系列データ・標本調査の初歩・相関表・散布図と原点を通る直線回帰・一般の単回帰・平均値とメジアンとモードの性質(使い方を間違えると、間違った答えになる実例)・相関表の読み方(統計の専門家でも面白い読み方が示されている)・ワークサンプリング・確率の問題・標本調査の偏り・二項分布・一樣乱数・シミュレーション・ χ^2 検定・Yets の修正・移動平均(スラッキー・ユール効果にも言及)・層別比推定などである。

以上が本書上巻で取り上げられた確率論/統計学の話題である。下巻は近く訳が完成刊行の予定である。

本書の面白さの1, 2の例 第1部第6章では大学対抗フットボールの得点が、負チームの得点と勝チームの得点の2元について度数分布として(相関表として)与えられる(1967年, 1968年)。大体 40×100 の大きな表となる。この表から勝チーム、負チームの得点の分布を導き、それから何点とったら勝つ確率はどうなるかを導く。その理論(?)と実際の観測された点とをプロットして図に示す。そしてこれらの作業を読者にも自分で行なわせるよう適切な練習問題がふんだんに(23)用意されている。

第2部第1章では盗作と確率という題で、実際に法廷で争われた記録から、その盗作の根拠として2つの小説に同じ105という番号の師団が現われる確率が、原告側の弁護士が3桁の数 $\rightarrow 1000$ ケース、これを両方で考えて $1000 \times 1000 = 10^6$ ケース、この中の1つなので 10^{-6} の確率だから、これは偶然の一致ではなく盗作だとしたものを取り上げた。そして、これにいろいろの側面から考察を加え、105師団という一致がもっと低い確率で、偶然でも起ることを示し、これに関連して25人くらいのクラスの中でも、誕生日が一致する組が存在する確率は0.5以上あることに言及している。

以上1, 2挙げたように誰でも親しみのもてる例題から統計の生きた読み方を示している。

本書の目的 編著者たちが序論に述べているように、統計学にある程度の知識を持つことがすべての人にとって必要なことは明らかであるにもかかわらず、社会人のほとんどが統計については正しい知識を持っているとは言えない状況にある。

この原因は小学校、中学校で統計的な考え方が十分に教えられてはいなかったことにある。理論的側面に片寄せた教育がされていたからであろう。これを打破する試みが本書の目的であると思われる。

本書の編集された経緯 本書は米国統計学会と米国数学教育学会とから出された委員によって構成された統計確率カリキュラム合同委員会によって編集された。このような社会的意義ある活動に手を挙げているアメリカの諸学会に敬意を表わしたい。(小林竜一)