

はしがき

最近、ビッグデータ、オープンデータという言葉に象徴されるように、統計学に対する知識がますます必要とされるようになってきています。本書は、主に文系の大学1年生を対象として、統計学の基礎の部分を例を用いてわかりやすく解説した入門書です。様々なデータに対して、統計学ではどのように考えるのかを読者に体感してもらうことを目的としています。そのため、厳密性を多少犠牲にしても、複雑な数式を用いなくて、わかりやすい文章で表現することを心掛けています。

統計学は大きく分けて2つあります。1つ目は初等的な考察を主な方法とする**記述統計**といわれるものです。本書では第1章で学ぶことになります。たとえば、ある物の重さを知りたいとしましょう。100回測ると100個のデータが得られますが、100個のデータを見ているだけではよくわかりません。そこで、データがどのようになっているのかがわかるようにデータを整理します。度数分布表、ヒストグラム、箱ひげ図にまとめたり、標本平均、標本分散を求めたりします。

2つ目は母集団という概念を念頭に置き、確率論を伴った考察を主な方法とする**推測統計**といわれるものです。推測統計は第2章以降で学ぶことになります。ある物の重さ (g) を10回測ったら、

24.5, 22.8, 23.7, 21.7, 24.3, 22.1, 23.4, 21.8, 25.2, 20.8

というデータが得られたとしましょう。このデータには、この物の本当の重さ w と観測誤差が含まれています。本当の重さ w は未知の定数で、観測誤差は確率的に大きくなったり小さくなったりすると考えられます。このデータから、 w は23gぐらいであり、それが30gであるということはまずありそうにありません。それは、 $w = 30$ とするより $w = 23$ とするほうが、このようなデータの得られる確率が大きくなるからです。 $w = 30$ としても、このようなデータの得られる確率はゼロではないので、 $w \neq 30$ という判断は正しいというわけでもありません。しかし、 $w = 30$ はまずないといってもいいでしょう。 $w = 23$ であり、 $w \neq 30$ という判断は絶対に正しいとはいえませんが、まず確からしい判断です。推測統計では、このような確からしい判断、つまり、不確実性を含む判断を導き、その不確実性を確率で測ることになります。このように推測統計の内容は確率論を伴うので理解するのが容易ではないかもしれませんが、専門的な確率論を知らなくても本書では十分理解できるようにしてあります。

今後、読者のそれぞれの専門分野で統計学が必要になる場合に本書が少しでもその手助けになればと願っています。最後に、本書の原稿を読んでいただき有益なコメントをしていただいた大阪府立大学高等教育推進機構の川添充教授、電気通信大学大学院情報システム学研究科の川野秀一准教授には心よりお礼を申し上げます。また、本書の最初の構想から出版まで長い時間を費やしてしまい、その間、我慢強く待っていただいた共立出版の信沢孝一氏、三浦拓馬氏に心よりお礼を申し上げます。

2016年1月 著者一同