

分布の視覚化

データの特徴を調べる方法の1つに、データの分布を調べるという方法があります。では、データの分布とはどのようなものなのでしょうか？ この章では、散布図や箱ひげ図を作成することで、データの分布とはどのようなものであるのかを確認した後、データの分布を調べる手法として、度数分布表とヒストグラムや箱ひげ図、度数分布表やヒストグラムは非常に強力なツールで、データの分布を視覚化し、分析されることが多くあります。また、度数分布表やヒストグラムを用いるというのが統計解析の基本です。ここでは、度数分布表やヒストグラムを用いてデータの分布を視覚化する方法について説明しましょう。

度数分布表

幹葉図

「度数分布表とヒストグラム」の「Sheet1」には、20人分の数学テストと英語テストの点数のデータがあります。この2つのデータは、特徴としてどのような違いを持っているのでしょうか？ ここでは、この2つのデータを「分布の違い」という観点から比較してみます。データの分布を視覚化する最も基本的な方法のひとつに、**幹葉図**という図を書くというのがあります。幹葉図は、図1のように、「幹」とよばれる左側の桁（例ではテストの点数の十の位）と「葉」とよばれる右側の桁（例ではテストの点数の一の位）に分けて数字を並べて、データの分布を視覚化したものです。この例では、例えば、7の行の左から3番目のデータが、73という点数のデータに対応します（図2）。数学テストのデータの例における75などのように、データ中に同じ値が複数個存在する場合は、それに応じて幹葉図にも同じ個数だけ続けて書きます。幹葉図に書き出されたデータは、必ずデータの総数と等しくなります。