資料２

データとグラフの基礎の基礎

●試験の結果は自分の点数と平均点で見ることが多い

小学校から大学まで、色々な科目で、さまざまな目的で試験が行われます。もちろん自分の試験の結果が気になりますね。そして、他の人の点数はどうだったかしらと気になります。そこで、先生方は今回のテストの平均は 65 点だったとか、最高点は 95 点で最低は 35 点であったとか、公表します。平均点より高い点数であればほっとしますし、低い点であればがっかりすることでしょう。平均点は、その集団の学習の成果としての試験の状況をあらわす一つの目安とされているわけです。

数多くの測定値（ここでは数値データ）が集まるのは、試験のときだけではありません。たとえば、現在私たちが着ている既製服の寸法は、ＪＩＳ規格により、全国で計測された身体計測値を使って決められています。人体計測値は、椅子の座面の高さ、机の高さなどの家具の大きさや電車のシートの幅など、多くの生活場面に使われています。

このように、試験の点数、身体計測値をはじめ、私たちの身近なことがらについて、データを集めて平均値を求め、比較することで、説得力のある表現ができたり、提言ができたりすることがあります。皆さんも、演習、卒論などの時に、データ（数値データ）を集め何かを表現したり、提言したりする機会があるかもしれません。そのために、ここでは集まったデータをどのように取り扱うかについて取り上げます。

●データが集まったら

ここでのデータとは、先に書いてあるように数値データを示します。データが集まったら、データの特徴を見るために度数分布表を作ってみましょう。図１は、ある集団の身長の度数分布（ヒストグラム）です。単な

るサンプルですので、データとしての意味はないことを了解の上見てみましょう。度数分布はヒストグラムとも呼ばれます。この集団で、154.5～159cm の間の身長の人が最も多く、168cm 以上の人は少ないことがわかります。

このような分布を書いてみると、多くの場合分布の中央にデータが多数存在し、そこから遠ざかるにしたがって減少し、最大値、最小値を示すデータは極めて少ない分布、「正規分布」を示します。

特に人体に関する量的な計測値の分布の多くは正規型を示しますが、体重、皮下脂肪など正規型を示さない計測値もあります。

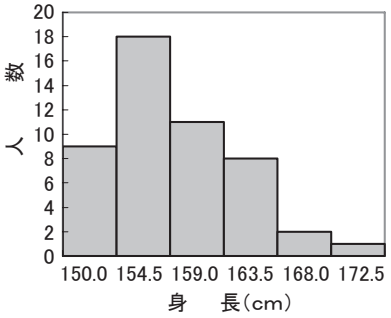

図１　ヒストグラムの一例

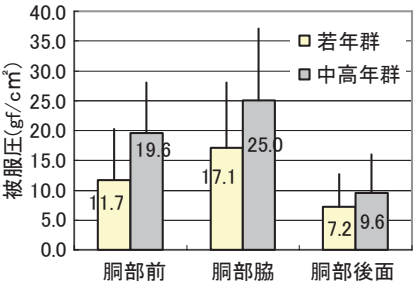

図２　日常着の衣服圧の違い

●平均値を求めよう

では、図１のデータの特徴を、数値であらわして見ましょう。データを代表する値には平均値、中央値、最頻値、最大値、最小値があります。ご存知のように平均値は、データの値すべてを加えてデータの個数で割ったものです。式であらわすと次のようになります。

平均（mean）　　$\bar{x} = \Sigma xi/n$

図１のデータにより計算してみると、この集団の身長の平均値は159.1cm、最大値は177.0cm、最小値は150.0cm でした。

●分散と標準偏差を計算してみよう

もし、ある二つの集団の身長を測ってヒストグラムを作成し、その分布を比較してみると、平均値は同じでも分布の形が異なることがあります。このことから、データを代表する値であらわす場合、平均値だけでは不十分であることがわかります。平均値以外に、データの散らばり具合を調べるために、分散や標準偏差を求めます。

分散（S^2）や標準偏差（S）は、データが平均値を中心にどのくらい散らばっているかをあらわす値で、次の式で求めることができます。

分散　（Distribution）　$S^2=(\Sigma x i^2-(\Sigma xi)^2)/n^2$

標準偏差　（Standard Deviation）　$S=\sqrt{S^2}$

計算が面倒くさそうですが、エクセルなどの表計算ソフトで簡単に求めることができます。

●標準偏差からわかること

さて、標準偏差を求めて、どのようなことがわかるのでしょうか？
標準偏差（S）を求めた時、正規分布またはそれに近い分布では、
データの 68.3%は　平均値±S　間
95.5%は　平均値±2S 間
99.7%は　平均値±3S 間 にあるという性質を持っています。

図1のデータの標準偏差を計算すると 5.2cm でした。これを上記の性質に当てはめると、この集団のほぼ7割が約 154cm から 164cm の身長であることがわかります。すなわち、平均値と標準偏差のみで、どのような分布の集団かわかるということになります。

●グラフを書いてみよう

つぎに、計算した平均値と標準偏差をグラフにしてみましょう。
グラフにはさまざまな種類があるのはご存知ですね。棒グラフ、折れ線グラフ、円グラフなどです。

棒グラフは、いくつかの集団のデータを比較するときのグラフでグラフの基本型ともいえます。図2は、高齢群と若年群の日常着の衣服圧の違いについて、棒グラフで表した図です。基線と呼ばれる横軸に比較する衣服圧測定部位を示し、各年齢群の違いを棒グラフの色であらわしています。目盛線と呼ばれる横軸は0.0から40.0、単位は衣服圧の単位gf/cm²となっており、縦軸の刻みは5.0 gf/cm²としてあります。この図では、部位別にはいずれの年齢群でも胴部脇の衣服圧が最も高く、年齢群別には中高年群の衣服圧が高いことが分かります。

このようなグラフを書くためには、縦軸の単位、縦軸の目盛の刻み、横軸の比較する区分、棒グラフの色分けについて決めなくてはなりません。棒グラフを描くときには、棒の数が多すぎないか、目盛の刻み方はこれでよいか、棒の模様わけや色分けはこれでよいか、棒を並べる順序はこれでよいかなどの検討が必要です。棒グラフでは、飛びぬけて長い棒がある場合、縦軸の目盛上に二本線を描くなどにより、縦軸の目盛をカットする場合がありますが、データの読み取りに誤解を招くことがあるのでやめましょう。また、棒の高さがほぼそろっている場合は、縦軸の目盛を大きくして比較しやすいようにする場合があります。しかし、現象としてあまり差がないものを拡大してしまう目盛の設定には注意が必要です。

●折れ線グラフと円グラフ

折れ線グラフはデータ同士を線で結ぶことによって、データの変化をみることを目的として書かれる場合が多いグラフです。棒グラフにより一つ一つの集団やその時間での値を見るより、変化を見てほしい場合には折れ線グラフを書きます。結んだ線の傾斜が大きいと変化が大きいと見られ、傾斜が小さいと変化が小さいと判断されます。このため、折れ線グラフは時系列データを入れることが多いのです。図3は、昭和27年から平成15年までの全国の浴衣取扱量の推移を折れ線グラフで示した図です。浴衣の取扱量は昭和40年代後半にピークを示し、その後減少していますが、平成7年から上昇に転じています。夏のくつろぎ着として浴衣が着られなく

なり、寝巻きが浴衣からパジャマやTシャツへと転じたためです。このところの上昇は、花火大会や夏祭りなどに若者を中心に浴衣が着られるようになったいわゆる「ゆかたブーム」によるものです。

円グラフは、全体を100としたときのパーセンテージを示すときに便利です。図4は婦人下着の危害情報件数の割合を示した図です。全体を100としたとき、どのような危害情報が多いか、あるいは少ないかがひとめで分かります。なお、自分で集めたデータでないときには、データの出所がわかるように、グラフの題に書き加えておきましょう。

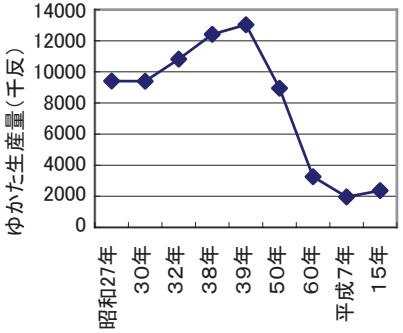

図3　全国の浴衣取扱量の推移
（関東染色工業連合会・日本ゆかた連合会調べ）

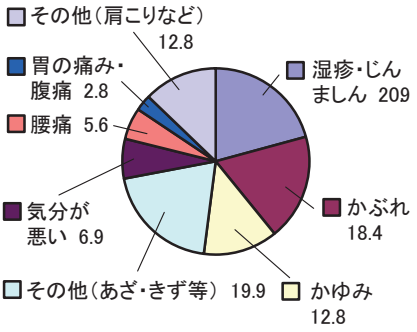

図4　婦人下着の危害情報件数の割合
（国民生活センター調べ）

さらに、このような数値データの取り扱いは、学術的には統計解析の分野となります。もっと詳しく知りたい人は、ぜひ、次の参考文献を見てみてください。

また、電卓での計算や、手書きのグラフは手間がかかります。エクセルなどの表計算ソフトでグラフ作成、計算も簡単にできますので、マニュアルを見てやってみましょう。同じデータでも、グラフを工夫することでさまざまの表現ができます。自分の言おうとしていることが最もよく表現できるグラフを書いてみましょう。

参考文献

大村平「統計のはなし－基礎・応用・娯楽－」日科技連（1969）
大村平「統計解析のはなし　データに語らせるテクニック」日科技連（1980）
石村貞夫「すぐわかる統計解析」東京図書（1993）
石村貞夫「グラフ統計のはなし」東京図書（1995）
（社）日本繊維製品消費科学会「例題を中心とした消費科学のためのデータ処理法」日本繊維製品消費科学会（1974）
上田尚一「講座情報をよむ統計学 4 統計グラフ」朝倉書店（2003）

