

第2章 度数分布法

統計的な考え方でデータを分析していくためには、そのばらつきの状態を把握することが基本となります。本章では、ばらつき、すなわち分布の姿を把握する最も基本的かつ重要な手法である度数分布法について焦点を当てます。

1. 統計的考え方

統計的な考え方について、整理をします。統計的な考え方は、実態として不明な母集団から、データを収集して、その分布の状態を分析し、母集団の状態を推測することです。

具体的には、

1) データを収集する目的を明確にする。

データを収集する前に、目的を明確にします。目的を明確にしないままに、データ収集を始めてしまうと、データだけが山のようになってしまう、手に負えなくなってしまう。

2) 常に母集団とサンプルの関係を考える。

収集したデータは、母集団からのサンプルになります。サンプルであることは、母集団の状態をそのまま反映できているとは限りません。サンプルから得た情報（統計量）は、必ず、母集団からのずれが存在します。したがって、常に母集団とサンプルの関係を頭に置いておく必要があります。

3) 収集したデータには、ばらつきがあることを忘れない。

収集したデータには、必ずばらつきがあります。すなわち、何らかの分布を持っています。このばらつきがあることとその分布の形態を忘れないようにします。

2. 度数分布法

統計的な考え方でデータを分析していくためには、そのばらつきの状態を把握することが基本となります。このばらつき、すなわち分布の姿を把握する最も基本的かつ重要な手法が**度数分布法**です。

(1) 度数分布の使い方

度数分布(frequency distribution)は、ばらつきを持つものの全てに活用ができます。その主な活用方法について、下記があります。

- 1) 報告用 工程の全体状況などを一目でわかるように報告する資料。
- 2) 解析用 不良の主な原因を見つけるために、原因別に層別した度数分布の作成。
- 3) 工程能力用 工場の工程能力、機械、設備の能力を把握するために作成。
- 4) 管理用 工程が管理状態に置かれているか監視するために作成。 など

(2) 度数分布表とは

度数分布とは、収集したデータが示す値（観測値）について、いくつかの等間隔の区間（級：class）に区切り、各区間に入るデータの数（度数：frequency）を分布として表すことです。このとき、始めに、この各区間に入るデータの数を表にします。この表を**度数分布表**(frequency distribution table)と言います。

関連する用語を整理すると

- ・ 級、または階級 (class) データの観測値を等間隔の範囲にわけその範囲
- ・ 度数 (frequency) その級に属する観測値の数
- ・ 代表値、または階級値 級を代表する値 通常は(上限値+下限値) / 2

(3) 度数分布表作成のポイント

度数分布表を作成するときに、最も重要なことは、級の数と級の幅の決定です。これにより、見たいデータの特徴が失われたり、あるいは、あまりにも変動が激しく、分布の分析が困難になったりします。

もし、

級の数が少ない場合 → 見たいデータの分布の特徴が隠れてしまいます。

級の数が多すぎる場合 → 級間の変動が大きく、全体の特長がわからなくなる、
あるいは、データの整理、分析が困難になります。

一般に、級の数は約 10 程度が望ましいと言われています。

なお、理論的に最適な級の数は求められていません。現在まで、平方根則（データの全数の平方根）、スタージェスの公式があります。参考までに、スタージェスの公式を以下に示します。

スタージェスの公式 望ましい級の数 = $\log_2 (n+1)$ n : データの全数

代表的な n では、

データの全数 (n)	50	100	1000	10000	100 万	1 億
望ましい級の数	7	8	11	14	21	28

(3) 度数分布の数量的表わし方

度数分布として得られた分布から、分布の特性を数量的に表すことが簡易となります。それは、その級に属するデータの値をその級の代表値として扱うことで、以下のそれぞれの値が簡易的に得られます。

- ・ 中心的傾向 ①算術平均
- ・ ばらつき ①偏差平方和
- ②分散
- ③標準偏差
- ④範囲

(4) ヒストグラム

度数分布表は、単なる数字の表です。度数分布を、視覚的に見やすくするために棒グラフにしたものを**ヒストグラム** (histogram) と呼びます。

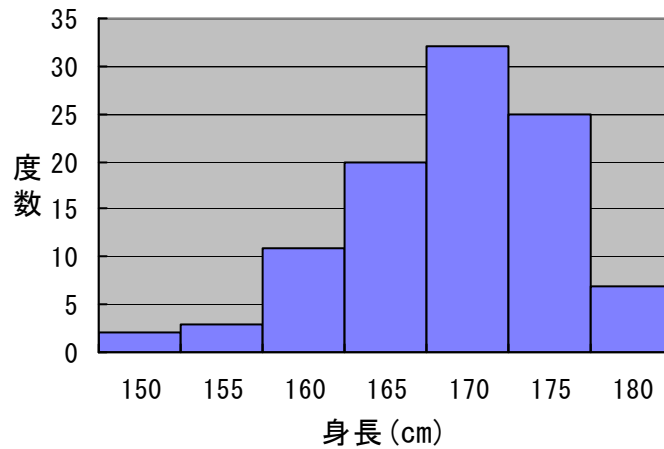


図 2-1 ヒストグラムの例

なお、ヒストグラムは長方形の面積が量（度数・比率）を表すように描かれます（長方形の底辺の長さが異なることもあります）。また、長方形の間には隙間を入れません。

(5) 度数分布の見方

度数分布表、ヒストグラムを作成したものの見方を以下に示します。

1) 規格限界に対し 製品限界が最適な位置にあるかの見方

ここで、規格限界とは、この限界を超えたものは出荷できない

製品限界とは、製造上のばらつきの限界 を言います。

- ・規格限界が製品限界に対して余裕がない場合、
→ 工程能力がない。不良品を製造する可能性が高い。
- ・規格限界が製品限界に対して狭すぎる場合、
→ 工程能力が経済的になっていない

2) 歯抜け型・くし歯型

ヒストグラムの形が歯抜け型やくし型になっている場合は、測定法、換算法にくせがないか、あるいは級分けが不適当でないかを確認する必要があります。

3) 離れ島型

ヒストグラムの形が離れ小島型の場合には、工程に異常がある可能性があります。

4) 分布が偏っている

ヒストグラムの分布が偏っている場合には、工程の中で、不良品の手直し作業、あるいは全数検査で不良品の排除等が工程内で実施されている可能性があります。

5) 山が2つ以上ある

ヒストグラムの分布に山が二つ以上ある場合には、要因として、2つ以上の要因を混在してデータを収集している可能性があります。層別して、データを再度、収集することが必要です。

(6) 度数分布使用の留意点

度数分布法も万能ではありません。また、用いる場合には留意すべき点もあります。以下に、その留意点を示します。

- ・ 時間的变化が不明
- ・ 群内変動、群間変動がつかみにくい
- ・ 多くのデータが必要