

# 度数分布表

## 度数分布表とは

度数分布表 (frequency table) とは、データを大きさの順に並び変えいくつか区分にまとめた表で、データ全体の分布 (distribution) を把握するために用います。

## 度数分布表の作成手順

1. 階級数または階級幅を決める
2. 階級値および有効桁数を考え、階級の境界 (どこで区切るか) を決める
3. 各階級の度数を数える
4. 階級と度数を元に、度数分布表を作成し、必要に応じて分布の形を見たりする

| 階級      | 階級値   | 度数  | 累積度数 | 相対度数  | 累積相対度数 |
|---------|-------|-----|------|-------|--------|
| 90～99   | 94.5  | 2   | 2    | 1.0   | 1.0    |
| 100～109 | 104.5 | 6   | 8    | 3.0   | 4.0    |
| 110～119 | 114.5 | 16  | 24   | 8.0   | 12.0   |
| 120～129 | 124.5 | 22  | 46   | 11.0  | 23.0   |
| 130～139 | 134.5 | 24  | 70   | 12.0  | 35.0   |
| 140～149 | 144.5 | 30  | 100  | 15.0  | 50.0   |
| 150～159 | 154.5 | 44  | 144  | 22.0  | 72.0   |
| 160～169 | 164.5 | 26  | 170  | 13.0  | 85.0   |
| 170～179 | 174.5 | 15  | 185  | 7.5   | 92.5   |
| 180～189 | 184.5 | 11  | 196  | 5.5   | 98.0   |
| 190～199 | 194.5 | 2   | 198  | 1.0   | 99.0   |
| 200～209 | 204.5 | 1   | 199  | 0.5   | 99.5   |
| 210～219 | 214.5 | 1   | 200  | 0.5   | 100.0  |
| 計       |       | 200 |      | 100.0 |        |

## 用語

- 階級 (class)
  - データを区切るときの区間
  - 各階級でとりうる最大の値を上限、最小の値を下限という
- 階級数
  - いくつの区間に区切るかの区間の数
  - 一般には10前後が目安 (多すぎても少なすぎてもいけない)
  - どのように区切るかは、キリのいい値や経験などを元にする場合があるが、客観的な区切り方を検討する場合はスタージスの方法 (後述) を用いる
- 階級幅 (class interval)
  - データをいくつ刻みに区切るかの区切る値の幅
- 階級値 (class mark)
  - 各階級の中央の値

$$\text{階級値} = \{ (\text{階級の上限}) + (\text{階級の下限}) \} / 2$$

- 度数 (frequency)
  - 各階級に含まれるデータの数
- 累積度数 (cumulative frequency)
  - 階級値の小さい(または大きい)ほうから、ある階級までの度数を合計した値
  - 最後の階級での値は度数の合計
- 相対度数 (relative frequency)
  - 度数の合計に対する各階級の度数の比 (全体を1=100%または100とする)

$$\text{相対度数} = (\text{その階級の度数}) / (\text{全体の度数}) \times 100 (\%)$$

- 累積相対度数 (relative cumulative frequency)
  - 階級値の小さい(または大きい)ほうから、ある階級までの相対度数を合計した値
  - つまり、ある階級までの度数の合計が全体の何%かを示す
  - 最後の階級での値は100%または100

$$\text{累積相対度数} = (\text{各階級の累積度数}) / (\text{全体の度数}) \times 100 (\%)$$

## 階級数の算出

階級数を算出するときに、経験やキリのいい数字ではなく、何らかの客観的な理由が必要となる場合に、使用される方法がいくつかある。

- スタージス (Starges) の方法
  - データの数が  $n$  個のとき、常用対数  $\log_{10}$  を用いて算出

$$1 + (3.322) \log_{10} n = 1 + \frac{\log_{10} n}{\log_{10} 2}$$

- シャリエ (Charier) の方法: 標準偏差の1/3
- フィッシャー (Fisher) の方法: 標準偏差の1/4

## Excelで度数分布表を作成

表計算ソフトの「Microsoft Excel」を使って、度数分布表を作成する場合、関数を使わなくても、四則演算 (+ - \* /) だけでも作成できます。

しかし、データ数が多い場合に度数を求めたり、度数などの合計を求めるときには、関数を使えばデータを処理しやすくなります。

### 度数分布表の作成で使用する関数

- 合計は、SUM関数を利用します。

#### SUM(合計を計算する)

- 書式: SUM(数値1, 数値2, ...)
- 引数: 数値1, 数値2, ... : 平均を計算するセルの範囲
- 例: B1 ~ B10セルまでのセルの数値の平均値を計算する

$$=SUM(B1:B10)$$

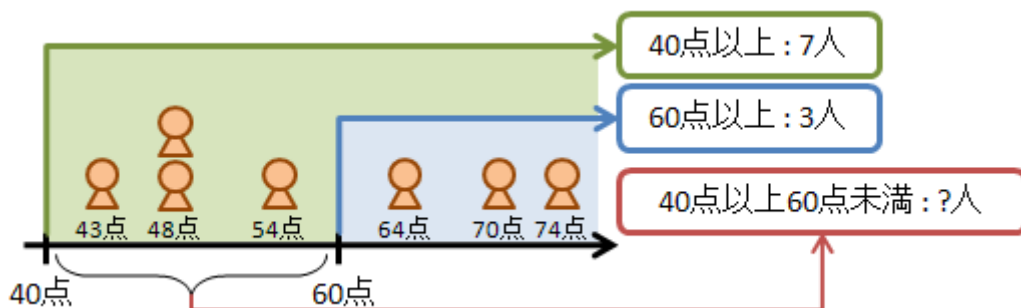
- 「値が60以上のセル」のように、特定の条件を満たすセルの個数を数えるには、**COUNTIF**関数を使います。

**COUNTIF(指定された範囲のセルのうち、検索条件に一致するセルの個数を返す)**

- 書式 : COUNTIF(範囲, 検索条件)
- 引数 : 範囲 : 個数を求めるセルの範囲
- 引数 : 検索条件 : 個数を求めるセルの検索条件
- 例 : W1 ~ W10セルまでで値が「80以上」のセルの個数を数える

`=COUNTIF(W1:W10, ">=80")`

COUNTIF関数を利用して数式をつくると、「40以上60未満の値があるセルの数」を求めることができます。ただし、COUNTIF 関数には条件は1つしか設定できないため、ひとつの数式でCOUNTIF 関数を2つ使います。



たとえば、C1～C7セルに上の図のように点数の値が入力されているとします。そのときに、「40以上60未満の値があるセルの数」を求める場合は、次のような数式になります

`=COUNTIF(C1:C7, ">=40")-COUNTIF(C1:C7, ">=60")`

つまり、「40以上の値があるセルの数」と「60以上の値があるセルの数」の差を求めればよいわけです。