第 II 部　統計

1　度数分布

ポイント

- 記述統計とは
- 度数分布表
- ヒストグラム

1.1　はじめに

- deta は複数形です。単数形は datum です。
- 複数の datum の集合が deta です。
- deta には様々な特徴が潜んでいます。
- deta に内在する情報を抽出すると特徴を把握することができます。

1.2 記述統計とは

「集団としての特長を記述するために、観測対象となった各個体について観測し、得られたデータを整理・要約する方法である。」(出典：東京大学教養学部統計学教室 篇『統計学入門』東京大学出版会)

ここで『観測』とは、調査や実験を指し、調査や実験を通じて調査対象となった各個体に関する観測値を得る。
各個体の観測値をまとめたものを『データ』という。
個体を識別する添え字に i を採用する。

1.3 データの種類

データは、観測の特性により 3 つに分けられる。ここでは基数データ, 序数データ, カテゴリカル・データに分けて説明する。

1.3.1 基数(きすう)データ

- 数の概念の内、量的側面をあらわす値。
- 物理量または数え上げ等によって測定される値。
- 単位が定義され、量によって測られる値。
- 値の大小に意味がある。

1.3.2 序数(じょすう)データ

- 数の概念のうち、順序的側面をあらわす値。
- 何らかの基準により順序付けされた、順位データ。
- 上位・下位の概念はあり、序列の比較は可能。
- 上位・下位との量的な差は不明。

1.3.3 カテゴリカル・データ

- 対象を属性に基き、いくつかに分類するときに、分類先を示す値。
- 同質か異質かによって識別されるデータ。
- 数字を利用している場合でも大小といった概念は持っていない。

1.4 度数分布表

- 『度数分布表』とは、観測値の取りうる値をいくつかの階級に分け、それぞれの階級で観測値がいくつあるか度数を数えて表にしたものをいう。階級の幅は全ての階級において等しい場合もあるし、異なる幅を採用する場合もある。
- 『階級値』とは、階級を代表する値をいう。階級が範囲を持つ場合は階級の上限値と下限値の中間値を階級とする場合が多い。
- 『相対度数』とは、観測値の総数で各度数を割った値をいう。総数の異なるデータの比較において有効。

問題 II−1 − 1

度数分布表を作成しなさい。階級の設定は、各自で判断するものとする。

表 1　小学生の体重（単位 kg）

20.5	25.5	22.5	25.5	29.0	21.5	24.0	25.5	22.5	25.5	25.5	27.5	29.0	31.5
22.5	25.5	25.5	31.5	24.0	27.5	30.5	29.0	34.0	26.5	30.5	24.0	24.0	26.5
29.0	29.0	26.5	29.0	31.5	22.5	25.5	25.5	29.0	31.5	25.5	29.0	30.5	31.5
29.0	30.5	29.0	32.5	30.5	29.0	27.5	30.5	32.5	31.5	32.5	34.0	29.0	27.5
27.5	25.5	30.5	29.0	31.5	30.5	30.5	32.5	34.0	31.5	30.5	32.5	31.5	32.5
36.5	37.5	26.5	30.5	25.5	27.5	30.5	31.5	32.5	32.5	34.0	29.0	30.5	31.5
30.5	36.5	31.5	32.5	30.5	32.5	37.5	29.0	31.5	32.5	34.0	31.5	27.5	30.5
32.5	31.5	32.5	30.5	31.5	32.5	32.5	30.5	35.5	36.5	32.5	36.5	36.5	37.5

サンプルデータによる度数分布表

表 2 小学生の体重の度数分布表

階級	度数	相対度数
25kg 未満	10	8.9 %
25kg 以上 30kg 未満	38	33.9 %
30kg 以上 35kg 未満	55	49.1 %
35kg 以上	9	8.0 %
合計	112.0	100.0 %

相対度数は小数点処理をしているため、相対度数の合計を求めても 100%にはならない。

問題 II−1 − 2

度数分布表を作成しなさい。階級の設定は、各自で判断するものとする。

表 3 小学生の身長（単位 cm）

112.5	120.0	127.5	115.0	117.5	125.0	117.5	120.0	130.0	125.0	125.0	132.5	135.0	122.5
120.0	125.0	127.5	125.0	127.5	127.5	117.5	125.0	120.0	127.5	120.0	130.0	127.5	130.0
125.0	130.0	130.0	130.0	127.5	132.5	122.5	127.5	130.0	127.5	132.5	135.0	132.5	135.0
137.5	117.5	122.5	122.5	125.0	125.0	120.0	125.0	122.5	125.0	132.5	125.0	122.5	125.0
127.5	132.5	130.0	132.5	127.5	130.0	137.5	130.0	132.5	137.5	135.0	140.0	145.0	120.0
125.0	132.5	130.0	135.0	137.5	122.5	130.0	132.5	140.0	135.0	130.0	135.0	132.5	137.5
140.0	137.5	137.5	132.5	125.0	137.5	125.0	130.0	132.5	137.5	135.0	140.0	127.5	132.5
140.0	132.5	140.0	135.0	137.5	132.5	135.0	137.5	142.5	135.0	140.0	142.5	147.5	150.0

サンプルデータによる度数分布表

表 4 小学生の身長の度数分布表

階級	度数	相対度数
120cm 未満	6	5.4 %
120cm 以上 130cm 未満	42	37.5 %
130cm 以上 140cm 未満	52	46.4 %
140cm 以上 150cm 未満	11	9.8 %
150cm 以上	1	0.9 %
合計	112	100.0 %

相対度数は小数点処理をしているため、相対度数の合計を求めても 100%にはならない。

1.5 ヒストグラム

度数分布表をもとに、横軸に階級値をとり、縦軸に度数または相対度数をとった棒グラフを『ヒストグラム』という。全ての階級値の幅がほぼ等しい場合は棒を隣接するように作図されることが多い。

問題 II−1 − 3

表**??**を用いて、ヒストグラムを作成しなさい。階級の設定は、問題 II−1 − 1 と同じものを採用するものとする。

問題 II−1 − 4

表**??**を用いて、ヒストグラムを作成しなさい。階級の設定は、問題 II−1 − 2 と同じものを採用するものとする。

図 1　ヒストグラム例 1

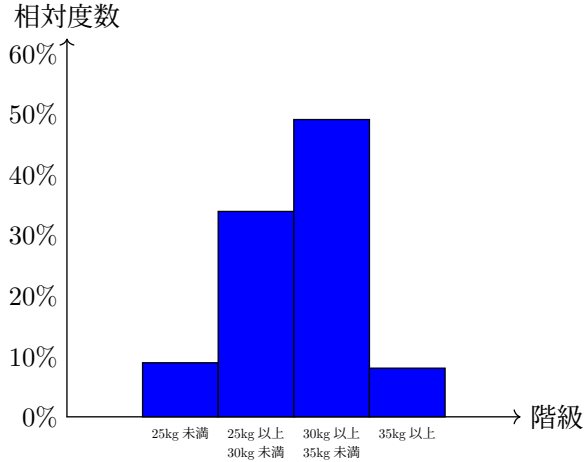
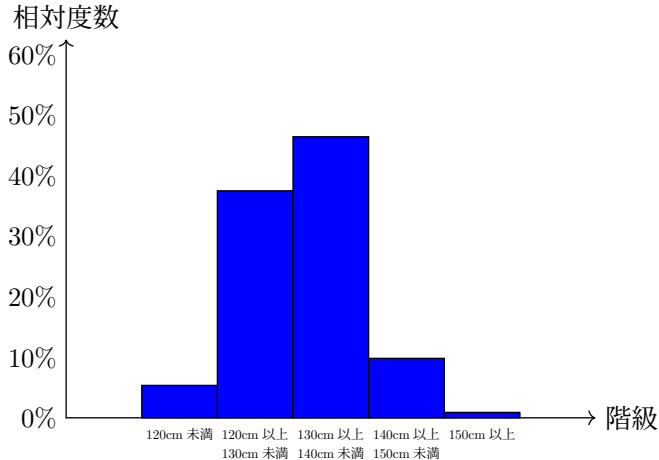

相対度数
60%
50%
40%
30%
20%
10%
0%
階級
25kg 未満
25kg 以上
30kg 未満
30kg 以上
35kg 未満
35kg 以上

図 2　ヒストグラム例 2


相対度数
60%
50%
40%
30%
20%
10%
0%
階級
120cm 未満
120cm 以上 130cm 未満
130cm 以上 140cm 未満
140cm 以上 150cm 未満
150cm 以上

1.6 まとめ

- データを、序数データ、基数データ、カテゴリカル・データと分類数する。
- データの整理要約する方法を記述統計という。
- データの取りうる値ごとに集計した表を度数分布表という。
- 階級は相当に恣意的に決めている。
- 度数分布を相対的に示した表を相対度数分布表という。
- 度数分布表を棒グラフにしたものをヒストグラムという。
- ヒストグラムには相対度数を採用する場合もある。