

①統計量
statistic

②平均
average :

この平均は、相加平均あるいは算術平均という。ほかに、相乗平均、調和平均などがある。

③中央値 (メジアン)
median :

極端に小さな値や大きな値があった場合、平均に比べその影響を受けにくい。

4 分析

収集・調査した情報を分析する方法を学ぼう。

数値の分析

二次データやアンケート調査などから得た数値データについては、集計し分析を行う必要がある。

A組とB組の通学時間 (分) の特徴の違いを比べるために各個人の通学時間のデータを得た。A組とB組のデータから、どのようなことがわかるだろうか。

A組の通学時間 (分)

12 42 50 10 70 15 55 62 22 50 32 40 45 65 45 75 13 20 7 25 26 5
52 30 3 35 5 45 38 60 38 20 30 25 15 40 45 60 55 42

B組の通学時間 (分)

35 35 55 48 25 45 45 30 18 15 68 48 35 38 48 20 35 45 40 48 30 28
55 25 47 38 15 25 45 38 55 50 23 10 33 25 35 45 35 27

④階級
データの範囲を分割したときの区間。

⑤度数
frequency

⑥度数分布図
histogram :
ヒストグラムともいう。

階級幅と階級値

階級の上限と下限の差を階級幅という。多くの場合、階級幅は一定にする。

また、階級の中央の値を階級値という。コンピュータを利用しないで平均値などを計算するときは、階級値を利用することもある。

階級と度数分布表

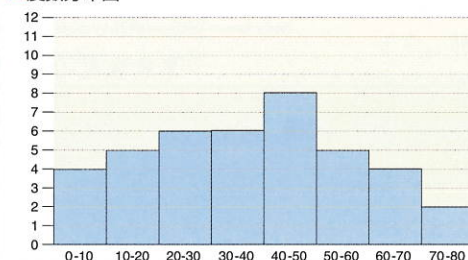
データの範囲をいくつかの階級^④という区間に分割して、各階級の度数^⑤ (区間に含まれるデータの個数) を度数分布表という表にまとめたり、度数分布図^⑥にしたりするとわかりやすくなる。

A組

■度数分布表

階級	0分以上 10分未満	10分以上 20分未満	20分以上 30分未満	30分以上 40分未満	40分以上 50分未満	50分以上 60分未満	60分以上 70分未満	70分以上 80分未満
度数	4	5	6	6	8	5	4	2

■度数分布図

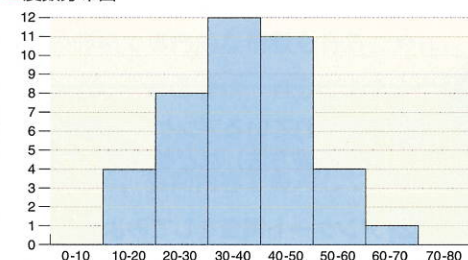


B組

■度数分布表

階級	0分以上 10分未満	10分以上 20分未満	20分以上 30分未満	30分以上 40分未満	40分以上 50分未満	50分以上 60分未満	60分以上 70分未満	70分以上 80分未満
度数	0	4	8	12	11	4	1	0

■度数分布図



・最頻値 (モード) ^⑦ : 度数分布表で最も多い度数の階級値のこと。

A組とB組のデータから、統計量は次の通りになる。

組	平均	範囲	中央値	最頻値
A	35.6	3から75	38	45
B	36.5	10から68	35	35

統計量や度数分布図から、B組よりもA組のほうが通学時間の散らばり具合が大きいことがわかる。

分散と標準偏差

散らばり具合は数値としても求めることができ、これを分散^⑧という。データ $x_1, x_2, \dots, x_n$ の分散は、次の式で求めることができる。

$$\text{分散} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

※ $\bar{x}$ は $x_1, x_2, \dots, x_n$ の平均値である。

すべてのデータが平均値と同じであれば、分散は0である。いっぽう、平均値から離れると右辺の分子はその差の二乗だけ増え、分散はどんどんと大きくなる。したがって、平均値から離れたデータが多ければ多いほど分散の値も大きくなる。

また、分散の正の平方根を標準偏差^⑨という。A組とB組の分散と標準偏差は次の通りである。

組	平均	分散	標準偏差
A	35.6	369.99	19.24
B	36.5	161.15	12.69

分散と標準偏差からも、B組よりA組のほうが、散らばり具合が大きいことがわかる。

⑦最頻値 (モード)
mode

⑧分散
variance

⑨標準偏差
standard deviation :
データの散らばり具合をあらわす数値の一つ。

表計算ソフトウェアの利用

データが多いときには表計算ソフトウェアを利用すると便利である。表計算ソフトウェアの関数を用いて値を求めることもできる。A1~A40に40個のデータが入っている場合、
平均 = AVERAGE(A1:A40)
最大値 = MAX(A1:A40)
最小値 = MIN(A1:A40)
中央値 = MEDIAN(A1:A40)
分散 = VARP(A1:A40)
標準偏差 = STDEV(P(A1:A40))

実習
42

あるスポーツ競技の大会が終了した。下の表は、大会に出場した2チームの各選手 (背番号1~10番) が、大会中に獲得した得点の一覧表である。

背番号	1	2	3	4	5	6	7	8	9	10
Aチーム	39	32	25	50	44	15	26	21	33	23
Bチーム	58	48	66	60	2	43	19	47	7	12

1 各チームの、選手の得点の平均値を計算してみよう。

2 各チームの得点の分散を求め、比較して何がわかるか答えよう。

