

経営データの処理と分析

a2p21063
光山 徹

1、度数分布表

- 階級と級(class) : 生データを大きさの順にいくつかのグループ
- 度数(frequency) : 各階級に属するデータの個数
- 級間隔(class interval) : 階級の幅、または区間と呼ぶ

度数分布表作成(Excel)その1

- Excelを使用するには、FREQUENCY関数、データの「分析ツール」、またはExcel VBAを使用する

度数分布表作成(Excel)その2

- データ数Nは変化するので、各度数を全数(度数の総和)で割った相対度数で、表した方が同種でデータ数が異なる他のデータ群と比較するのが便利



相対度数分布表

作成手順

- Excel画面上に、生データを入力する
- データの数Nを求める
- 最大値(max)と最小値(min)を見つけ、範囲(R:range)を計算する($R = \max - \min$)
- 階級数mを決める

スタージェス(Sturges)の方法

スタージェス1	データ数N	32	64	128	256	512	1024
	階級m	6	7	8	9	10	11
スタージェス2	データ数N	30~45	46~90	91~181	182~362	363~724	725~1448
	階級m	6	7	8	9	10	11

データ数N	30~50	50~100	100~250	250以上
級数m	5~7	6~10	7~12	10~20

階級数mの求め方(スタージェス)

$$m = 1 + \log_2 N = 1 + \log_2 N / \log_{10} 2 \approx 1 + 3.322 \times \log_{10} N$$

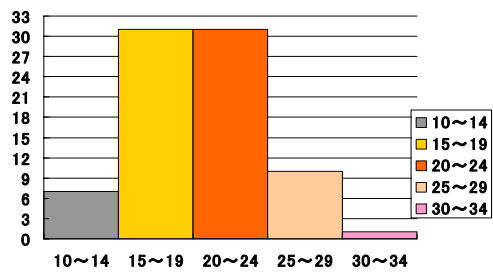
階級数mのまとめ

- 階級の数が少なすぎる場合、情報の喪失が大きくなり、一方クラスの数が多すぎると手数が増える割には精度が上がらず全体的な特徴がつかみにくい

販売データの整理と加工

30	19	14	17	15	18	12	15
18	23	19	25	23	24	24	13
19	10	19	16	28	13	16	20
18	21	28	22	17	23	20	18
20	17	21	18	24	16	19	25
20	17	25	14	11	21	18	22
23	22	15	24	24	21	15	28
12	21	26	22	13	20	27	16
15	25	17	20	19	14	23	18
20	22	22	15	22	26	22	21

度数柱状図(histogram)



- 連続量に対応するヒストグラムは、棒の間が離れていないものを指す