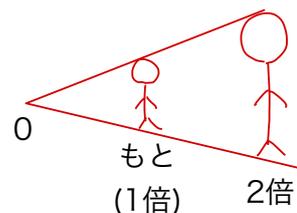


■変数の分類

	尺度	特徴				変数の例	利用できる統計量の例
		区別	順序	差の意味	比		
質的変数	名義尺度	○	×	×	×	性別, 血液型, 電話番号	度数, 最頻値
	順序尺度	○	○	×	×	好みの評価, マラソンの着順	(上に加え)中央値
量的変数	間隔尺度	○	○	○	×	温度(°C, °F), 偏差値	(上に加え)平均, 標準偏差
	比率尺度 (比例尺度)	○	○	○	○	身長, 体重	(上に加え)相乗平均

(注意1)例えば好き嫌いを答える質問回答の処理で、厳密には順序尺度であるが間隔尺度とみなして平均を考える場合もあるなど、解釈によって尺度の分類は曖昧なこともある。上表はあくまで原則として考えておくとよい。
 (注意2)比率尺度は、0が「何もないこと」を意味し、負の値はとらない(これも原則として考えておくとよい)。



量的変数は次のようにも分類できる

連続変数…ある範囲の値を連続的にとる (例)身長

離散変数…とびとびの値をとる (例)回数, 人数

※離散変数であっても、取りうる値が多い場合には連続変数と見なして分析することがよくある。

■度数分布表, ヒストグラム

量的変数の観測値の整理のために、度数分布表やヒストグラムが使われる。

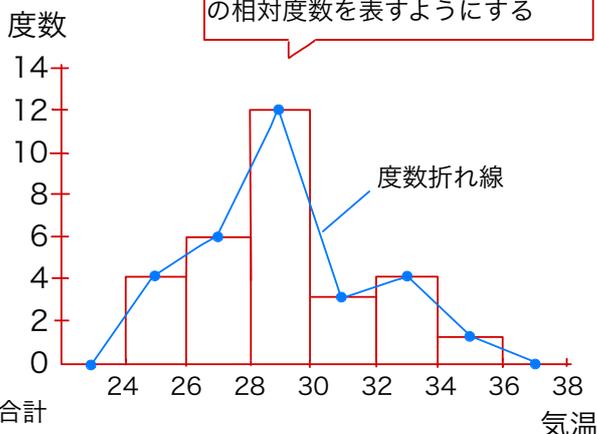
質的変数の場合はヒストグラムではなく棒グラフ(柱の隙間が空いている)を使う

(例)ある都市における9月の最高気温(°C)

35.8 33.3 33.7 30.5 29.1 28.0 29.2 28.9 28.4 27.9
 27.3 31.5 32.5 32.0 29.3 29.4 28.7 28.7 28.9 29.2
 29.8 30.3 27.1 26.4 24.5 24.3 24.2 25.3 26.2 26.8

ヒストグラムでは、全体に対する各長方形の面積の割合がその階級の相対度数を表すようにする

階級	階級値	度数	相対度数	累積相対度数
24 ≤ x < 26	25	4	4/30=0.13	4/30=0.13
26 ≤ x < 28	27	6	6/30=0.20	10/30=0.33
28 ≤ x < 30	29	12	12/30=0.40	22/30=0.73
30 ≤ x < 32	31	3	3/30=0.10	25/30=0.83
32 ≤ x < 34	33	4	4/30=0.13	29/30=0.97
34 ≤ x < 36	35	1	1/30=0.03	30/30=1.00
合計		30	30/30=1.00	



階級値とは、階級の真ん中の値のこと

その階級以前の相対度数の合計
 $4/30+6/30+12/30=22/30$

階級の数の目安として、次の公式が知られている。

観測値の総数を n としたとき、階級の数 C の目安は

$$C = 1 + \frac{\log n}{\log 2} \approx 1 + 3.3 \log_{10} n \quad (\text{スタージェスの公式})$$

ここで真ん中の式の分数内の対数は自然対数である。

スタージェスの公式による階級数の目安

観測数 n	20	40	80	150	300	500	1000
階級数 C	5	6	7	8	9	10	11

(01-1) 温度計で計測した気温を比較するとき、「30℃は20℃よりも10℃高い」と表現することはあっても、「30℃は20℃よりも1.5倍高い」という表現は通常はしない。これは、気温が(①)尺度であるからで(②)に意味がないからである。

【選択肢】 順序, 間隔, 比, 質的, 離散

(01-2) 100人の学生に対し100点満点のテストを行い、その結果を以下の度数分布表とヒストグラムにまとめた。()に適する数値を小数第3位を四捨五入して求めよ。

階級	度数	相対度数	累積相対度数
$0 \leq x < 10$	9	0.090	0.090
$10 \leq x < 20$	9	0.090	0.180
$20 \leq x < 30$	20	0.200	0.380
$30 \leq x < 40$	5	(ア)	(イ)
$40 \leq x < 50$	8	0.080	0.510
$50 \leq x < 60$	9	0.090	0.600
$60 \leq x < 70$	13	0.130	0.730
$70 \leq x < 80$	10	0.100	0.830
$80 \leq x < 90$	9	0.090	0.920
$90 \leq x \leq 100$	8	0.080	1.000
合計	100	1.000	

