

## 【データの種類と特徴】

### 1. 統計の対象となるデータの種類とは？

統計で扱うデータは大きく分けて次の2つに分類できる。

- ( ) データ…数学的な大小関係が成立するデータ。  
値の差が明確で、誰が計測しても同じになるため、比較しやすい。

(例) 身長, ( )

- ( ) データ…数学的な大小関係が成立しないデータ。  
数値で表されていても、順序を示すだけであるとか、分類のために数字を当てはめているだけのようなもの。

(例) 性別, 満足度, ( )

(注) 実務上の度合いの扱い

満足度のような度合いを統計で扱う場合には、見た目上、( 量的 ) データと同じように扱うことがある。例えば、アンケート回答項目で、

とても面白かった……5点

やや面白かった……4点

どちらともいえない……3点

ややつまらなかった……2点

とてもつまらなかった…1点

というように点数化して解釈する場合がある。

統計の対象となるデータの集団は主として次の2つである。

- ( ) …調査対象のすべてを含む集団。一般にこれは膨大なデータの個数になることが多いので、すべてを調査することは難しい。

- ( ) …調査によって得られた集団の一部のデータ。サンプルデータ。

★統計で行われるのは、( ) という方法である。これは、「データの羅列を何かの基準で整理して、意味のある情報だけを抽出する」ことをいう。その主な手法は次の2つ。

① ( ) して特徴を捉える。

② 1つの数値で特徴を ( ) させる。

## 2. 量的データの特徴をつかむ

データを分析しようというときは、そのデータはたくさん集めるほど全体のことがわかるようになる。しかし、そのたくさんのデータを数字として並べただけでは何もわからないので、量的データの特徴をつかむ必要がある。

- ( ) **分布表**…データを値のとり範囲ごとに区切って、そこに属するデータの個数を表にしたものをいう。

→区切りは ( ) といい、その代表 [真ん中] の値を ( )、区切りに属するデータの個数を ( )、全体に対するデータの個数の割合を ( ) という。それを求める式は、

$$( ) = \frac{( ) \text{ に属するデータの個数}}{( ) \text{ のデータの個数}}$$

また、その階級までの度数を合計したものを ( ) という。

**[練習 1]** 新しいスプレッドシートで次のデータの度数分布表をつくりなさい。ただし、階級は 450 円から 50 円刻みで 850 円までとし、階級・度数・相対度数・累積度数を入れる。

ある地域のラーメン店の醤油ラーメンの価格 [円]

店	価格	店	価格	店	価格	店	価格	店	価格	店	価格
A	760	G	600	M	630	S	820	Y	710	AE	660
B	480	H	670	N	780	T	650	Z	720	AF	590
C	650	I	770	O	550	U	640	AA	510	AG	610
D	530	J	660	P	530	V	720	AB	680	AH	720
E	740	K	790	Q	620	W	640	AC	580	AI	800
F	580	L	610	R	660	X	660	AD	700	AJ	650

- ( ) …度数をグラフにしたもので、いわゆる棒グラフである。ただし、横軸は ( ) した数値なので、棒と棒の間隔は ( ) こと。

**[練習 2]** スプレッドシートの同じシート上に練習 1 のヒストグラムを描かせなさい。(練習 1 と同様に**標準級間隔**は 50 円とし、**パッケージサイズ**で指定します。)

※練習 1・2 のスプレッドシートは Classroom から配付するので、それを使用しましょう。

## 【データの種類と特徴】

### 1. 統計の対象となるデータの種類とは？

統計で扱うデータは大きく分けて次の2つに分類できる。

- ( **量的** ) データ…数学的な大小関係が成立するデータ。  
値の差が明確で、誰が計測しても同じになるため、比較しやすい。  
(例) 身長, ( **体重, 年齢, 気温, 気圧, 冊数, 価格, 電圧…**など )
- ( **質的** ) データ…数学的な大小関係が成立しないデータ。  
数値で表されていても、順序を示すだけであるとか、分類のために数字を当てはめているだけのようなもの。  
(例) 性別, 満足度, ( **出身県, 英検の級, 天気, 血液型…**など )

(注) 実務上の**度合い**の扱い

満足度のような度合いを統計で扱う場合には、見た目上、( **量的** ) データと同じように扱うことがある。例えば、アンケート回答項目で、

とても面白かった……………5点  
やや面白かった……………4点  
どちらともいえない………3点  
ややつまらなかった………2点  
とてもつまらなかった…1点

というように点数化して解釈する場合がある。

統計の対象となるデータの集団は主として次の2つである。

- ( **母集団** ) …調査対象のすべてを含む集団。一般にこれは膨大なデータの個数になることが多いので、すべてを調査することは難しい。
- ( **標本** ) …調査によって得られた集団の一部のデータ。サンプルデータ。

★統計で行われるのは、( **縮約** ) という方法である。これは、「データの羅列を何かの基準で整理して、意味のある情報だけを抽出する」ことをいう。その主な手法は次の2つ。

- ① ( **グラフ化** ) して特徴を捉える。
- ② 1つの数値で特徴を ( **代表** ) させる。

## 2. 量的データの特徴をつかむ

データを分析しようというときは、そのデータはたくさん集めるほど全体のことがわかるようになる。しかし、そのたくさんのデータを数字として並べただけでは何もわからないので、量的データの特徴をつかむ必要がある。

- ( **度数** ) 分布表…データを値のとり範囲ごとに区切って、そこに属するデータの個数を表にしたものをいう。

→区切りは ( **階級** ) といい、その代表 [真ん中] の値を ( **階級値** ), 区切りに属するデータの個数を ( **度数** ), 全体に対するデータの個数の割合を ( **相対度数** ) という。それを求める式は、

$$\text{( 相対度数 )} = \frac{\text{( 各階級 ) に属するデータの個数}}{\text{( すべて ) のデータの個数}}$$

また、その階級までの度数を合計したものを ( **累積度数** ) という。

[練習 1] 新しいスプレッドシートで次のデータの度数分布表をつくりなさい。ただし、階級は 450 円から 50 円刻みで 850 円までとし、階級・度数・相対度数・累積度数を入れる。

ある地域のラーメン店の醤油ラーメンの価格 [円]

店	価格	店	価格	店	価格	店	価格	店	価格	店	価格
A	760	G	600	M	630	S	820	Y	710	AE	660
B	480	H	670	N	780	T	650	Z	720	AF	590
C	650	I	770	O	550	U	640	AA	510	AG	610
D	530	J	660	P	530	V	720	AB	680	AH	720
E	740	K	790	Q	620	W	640	AC	580	AI	800
F	580	L	610	R	660	X	660	AD	700	AJ	650

- ( **ヒストグラム** ) …度数をグラフにしたもので、いわゆる棒グラフである。  
ただし、横軸は ( **連続** ) した数値なので、棒と棒の間隔は ( **あけない** ) こと。

[練習 2] スプレッドシートの同じシート上に練習 1 のヒストグラムを描かせなさい。

(練習 1 と同様に**標準級間隔**は 50 円とし、**パッケージ**サイズで指定します。)

※練習 1・2 のスプレッドシートは Classroom から配付するので、それを使用しましょう。