

## 表計算 (Excel) とデータ処理 (10)

## まとめ

## 1. 尺度水準：計算方法からのデータの区別

名義尺度：等しい／等しくないだけが見えるデータ（名前，車のナンバーなど）

順序尺度：上記に加えて，順序（大小前後など）が見えるデータ（名簿，競争の順位など）

間隔尺度：上記に加えて，加減しても意味のある（差が意味のある）データ（摂氏温度など）

比例尺度：上記に加えて，乗除しても意味のある（割合が意味のある）データ（身長，絶対温度など）

## 2. ヒストグラム（度数分布）

データの集まりを，カテゴリや部分範囲で分けしたとき，その範囲のデータ数を整理したもの。

階級：それぞれの分けのこと。

階級値：階級を示す値のこと。階級に範囲がある場合は，その範囲の真ん中を階級値とすることが多い。

〔課題〕 各尺度について，ヒストグラムの例を挙げなさい。

## 3. データの要約（要約統計量）

## 3.1 間隔尺度以上の要約統計量

平均 [            ]

算術平均：データ値の合計をデータ数で割ったもの。

分散、標準偏差 [            ]

分散：平均からの差（偏差）の二乗和（二乗した値の合計）をデータ数で割ったもの。

データの二乗和をデータ数で割ったものから平均の二乗を引いても求められる。

標準偏差：分散の正の平方根。

歪度 [            ]（平均偏差の3乗の平均を標準偏差の3乗で割った値）

度数分布で左右対称の違いを表す。

歪度が 0 の時は左右対称の場合が多く、正の場合は裾が正方向に広がり、負の場合は負方向に広がっていることを表す。「正規分布」の場合、歪度は 0。

課題：歪度が、負、0、正のグラフを描きなさい。

尖度 [ ] (平均偏差の4乗の平均を標準偏差の4乗で割った値から3を引いた値)

度数分布で山の尖り(裾の広がり)の程度を表す

尖度が大きいほど尖りが急で裾野が広がっていることを表す

「正規分布」の場合、尖度は 0 になる。

課題：尖度が大きいグラフと小さいグラフを描きなさい。

### 3.2 順序尺度以上の要約統計量

中央値(メジアン、メディアン (median)) [ ]

データを並べたときに、真ん中の順序になるデータの値。第2四分位でもある。

第1四分位・第3四分位 [ ]

データを小さい順に並べたときに、1/4, 3/4 の順位にあるデータの値。

最小値・最大値 [ ]

データを並べたときの、最大、最小の値。

### 3.3 名義尺度以上の要約統計量

最頻値(モード (mode)) [ ]

度数分布で最も高い度数を示す階級の階級値。

[課題]

上記の統計量について、データ群の代表としての統計量(代表値)と、データ群の広がりを示す統計量(散布度)に区別し [ ] に書き入れなさい。

推測統計（付録）

大きな集団（母集団）から「無作為」に抽出した一部のデータから、集団の特性などを推測する統計的手法。

「点推定」「区間推定」「仮説検定」は、推測統計の一部としてあり、医学・医療の分野では、疫学調査（因果関係の推測や臨床試験など）の基礎となっている。

1) 点推定とは

一部のデータの平均を、母集団の推定平均とする、など。

一部のデータの「不偏分散」の平方根を母集団の推定標準偏差とする、など。

母集団の分布が歪んでいることが予測される場合は、中央値、最頻値などを使う。

2) 区間推定とは

推定した値がばらつく範囲を示すこと。

3) 仮説検定とは

母集団が特定の分布に従っているかどうかを検証すること。

たとえば、「集団A,Bは同じ母集団に含まれている」という仮説（帰無仮説）を否定することで、「集団A,Bは別の母集団に含まれる」可能性が高いことを示す。

2 正規分布

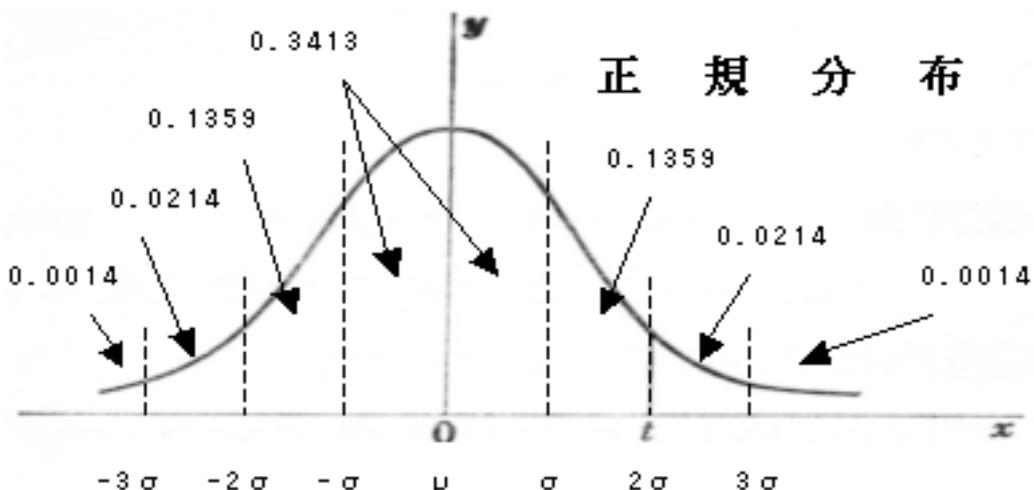
数学者のガウスが詳細を明らかにしたので、「ガウス分布」とも呼ばれる。

正規分布は左右対称の形をしていて、平均 $\mu$ 、標準偏差 $\sigma$ の正規分布は以下の特徴を持つ。

平均 $\pm$ 標準偏差の範囲内には、全体の 0.683 のデータが入る

平均 $\pm$ 標準偏差 $\times 2$ の範囲内には、全体の 0.954 のデータが入る

平均 $\pm$ 標準偏差 $\times 3$ の範囲内には、全体の 0.997 のデータが入る



## 推測統計の考え方

全国の失業率を調べたい  
新薬の効果を検証したい  
勉強法の効果を調べたい

など、全体（全国民や、全患者、全学生など）が大きすぎて、すべてを調べることが不可能な場合、その大きな集団（母集団）から「無作為」に抽出した一部のデータから、集団の特性などを推測することで、全体の様子を推測する。「点推定」「区間推定」については、前回の資料参照。

## 仮説検定の考え方

たとえば、勉強法Aと勉強法Bで、どちらが効果があるかを仮説検定する例を考えてみる。効果を見るために、グループを2つ作り、それぞれの勉強法で勉強し、試験を受けてもらう。試験結果を比較することで、どちらの勉強法が効果があるか、あるいは変わらないかを調べられる。ただし、仮説検定では「差がある」ということを直接結論づけることはできない。

### 前提となる条件：

勉強法の実施前には、その試験に対する能力は2つのグループで差がないこと。  
各グループの母集団は正規分布をしていることが仮定できること。

### 帰無仮説：

勉強法AとBでは効果に差がない。

### 検定結果：

試験結果の得点を検定すると、「差がない」と仮定した場合に、この結果を得られる確率は1%以下である。

### 結果の解釈：

1%以下ということは、ほとんど起こりえないことなのだから、そもそも「差がない」という仮説が間違っていたのだ。結論としては、「勉強法AとBでは効果に差がない」ということは無い。（つまり効果に差がある）

前回の補足：グラフの各データに「ラベル」を付加する方法

- 1 ラベルをつけたグラフの系列全体を選択する
- 2 データラベル：データラベルオプションから  
ラベルの内容：セルの選択をえらび、ラベルにしたいデータ系列を選択する。