

# 表題 可視化で超速攻指導！実験付き統計・分析指導

## 副題(表計算ソフトウェア活用で時短実現)

神奈川県川崎市立橋高等学校・布村 覚

要旨 ネットワーク・DB・プログラミング・モデル化とシミュレーションなど20年前の教科情報開始当初からの指導項目であるのに対し、データサイエンスの単元が近年重要視されてきている。シミュレーションでは予測に対して様々なアプローチが存在することを表計算ソフトウェアにより可視化して理解へ導く取組を実践してきた。文中◎は1年必履修「情報の科学」、○は3年選択「社会と情報」

### 1. 確率分布 ◎

現在数学I及び情報の科学における「統計」単元は、ベルカーブありきで展開する。

事象を帰納法的に法則化し、その原理を基に未経験の課題に対して演繹的に解決する手順が科学であるとの視点で以下の実験を実践している。

#### 1.1 確率的現象の発生とガウス分布の発見

乱数により1万回の仮想コイントスを実行。これを1万回繰り返して結果検証を実施。

##### 1.1.1 正規分布の概念を認知させる。

裏を0、表を1として=RANDBETWEEN(0,1)をCells(1,1~10000)に実行。Cells(1,10001)に合計値を算出する。10,000回の試行結果をVBAでCells(2,1~10000)に転記するプログラムを書かせて実行。範囲を指定してヒストグラムを描画。

##### 1.1.2 ガウス分布確認

マクロの複数回実行及び周囲の生徒の結果から帰納法的に正規分布/ベル曲線を理解させる。

##### 1.1.3 パレート

同範囲を指定してパレート図を描画させ、重要要素や問題点抽出のポイントを理解させる。

##### 1.1.4 標準偏差 $\sigma$ と概形

標準偏差 $\sigma$ を任意に設定してガウス曲線描画を実行して概形との関係を理解させる。分析ツールから正規分布条件で乱数を発生(例：個数1,000、平均0、 $\sigma=15,25.35$ )させ、度数分布表を作成後ヒストグラム描画。概形を比較させる。

### 1.2 代表値・分散・標準偏差の算術計算 ◎

表から統計的代表値・分散・標準偏差・SS・四分位数を算出させ、ヒストグラム・箱ひげ図描画。

### 1.3 新体力測定結果分析 ◎

#### 1.3.1 各測定CSVデータからRDB作成

SQLでフィールド結合を実施し、テーブル作成後表計算ソフトウェアにエクスポート、分析開始。

#### 1.3.2 種目ごと代表値・標準偏差・分布図作成

ヒストグラムの二峰性やパレートにおける特異性に注目させ、理由を推測させた上で再検証。

#### 1.3.3 種目間の相関係数と散布図描画

関連性の強さを予測後、実行して検証。

### 2. モデル化とシミュレーション ◎

様々な題材でモデル化を思考させた上で、それぞれシミュレーションを実施する。

#### 2.1 数理モデル・シミュレーション

物理から「物体の投げ運動」を取り上げ、初速 $V_0$ と投射角 $\theta$ をパラメータとする水平 $H$ ・鉛直 $V$ 方向を要素にした軌跡を描画させる。

留意点：ラジアン角未習の場合は完成形を提供

#### 2.2 確率的シミュレーション

モンテカルロ法を $\pi$ の算出を主題にプログラムさせる。半径1の四分の一円に1,000,000回打点を6回試行した上で結果をグラフ化させる。

例題：理論値との誤差が5%以下になる試行回数

#### 2.3 選挙シミュレーション

比例代表制選挙制度における議席確保。政党合併によるメリットとその分岐点の考察させる。

#### 2.4 動的要素を含む数理シミュレーション

マウスの個体数初期値を $n$ 匹とする。出生率と死亡率をパラメータとする個体数増加シミュレーションをプログラムさせる。

### 3. 回帰分析 ◎○

人口推移・covid-19新規感染者数を題材に実践

#### 3.1 回帰分析の見方・描き方 ◎

総務省統計局データから1920年以降2020年までの我が国の人口推移を基に2060年を予測

データを散布図に描画後、線形近似曲線(回帰直線)・多項式近似曲線(回帰曲線)をR2乗値(決定係数)とともに追加し信頼性を考察させる。

#### 3.2 Covid-19新規感染者発生数推移 ○

NHK「新型コロナウイルス特設サイト」のデータから適宜範囲を抽出した上で、回帰分析を実施。R2乗値を評価して、近未来を予測。図3.2

### 4. 移動平均 ○

動的変化を伴う変量を短期及び長期移動平均とともにグラフ化し、可視的にトレンドを捉える。

#### 4.1 移動平均の見方・描き方

基となる変量を散布図で描画し、短期・長期移動平均曲線を追加。短期が長期の上にある時は上昇傾向にあり、逆なら下降傾向であることと「ゴールデンクロス、デッドクロス」を捉える。

#### 4.2 Covid-19 新規感染者発生数移動平均

2020年1月以降のNHK発信データが長期に亘るので、短期を7日・長期を28日とする。母集団の規模から東京を題材とする。図4.2

### 5. チャートグラフ ○

時間軸上で動的变化を伴う変量を等区間のローソク足に描画し、それを時系列データとして並べた陽線・陰線・上下ひげから予測の根拠とする。

#### 5.1 チャートグラフの見方・描き方

元となるデータから等間隔の区間に区切り、始値・最高値・最低値・終値の4点を捉えてローソク足を描画する。動的变化を捉えており、概念が箱ひげ図と異なる点に注目させる。陽線・陰線・上ひげ・下ひげに対する解釈を理解させる。

#### 5.2 Covid-19 新規感染者発生数チャート

NHK発信データから日曜～土曜の1週間毎に区切り、日曜日発生数を始値、土曜日発生数を終値に設定、週の最高値・最低値を関数で確定する。

これら4要素の時系列テーブルを作成し、株価グラフを描画したうえで評価させる。図5.2

#### 5.3 対前週比増減推移

対前週比の7日間移動平均もまた収束する方向である。収束後の動きに注目したい。

#### 5.4 グループ学習（県別分析と結果統合）

4～6人のグループの中で各都道府県を分担して分析。結果を結合して、担当地域と他県との相関、地域制、共通点など多角的学習を実践させた。

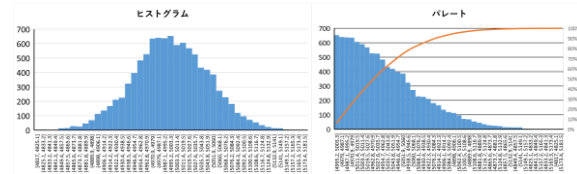


図 1.1.1 コイントス1万回試行・分布図パレート図

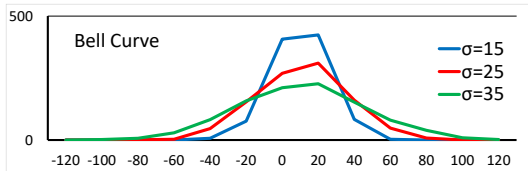


図 1.1.4  $\sigma$  値と Bell Curve 描画実験

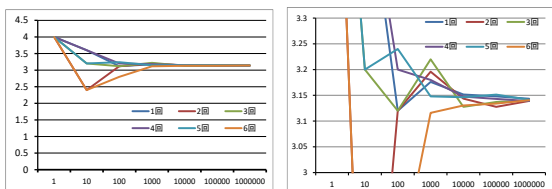


図 2 モンテカルロ法 100 万回試行実験  $\pi$  算出

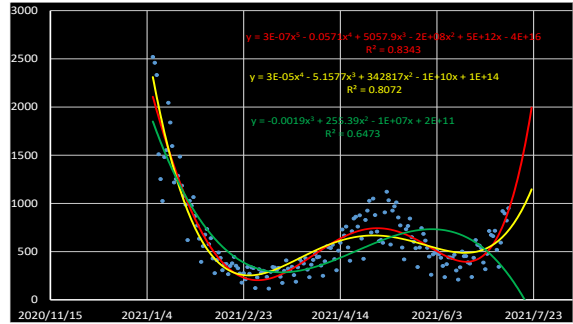


図 3.2 回帰分析(東京 2021.7.10)

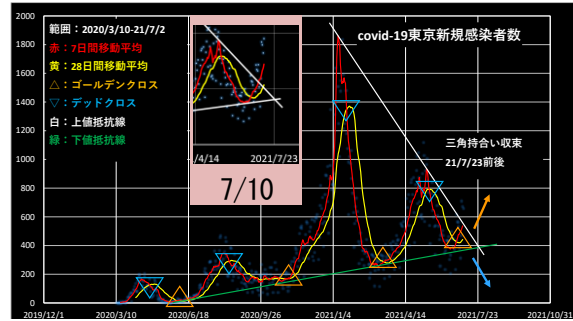


図 4.2 移動平均分析(東京 2021.7.3-10)

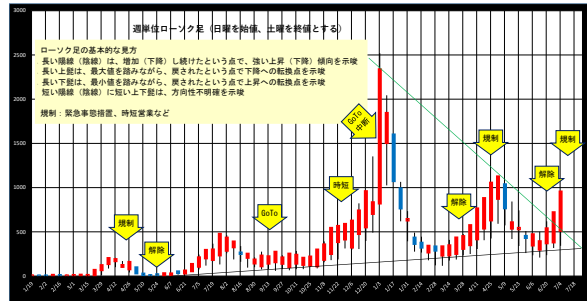


図 5.2 チャート分析(東京 2021.7.10)

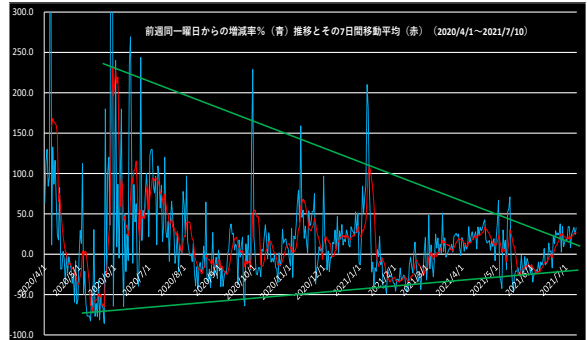


図 5.3 対前週比推移(東京 2021.7.10)

#### 5.5 NY DOW JONES 株価指数分析

コロナ禍で大統領選のあった米国、社会情勢や経済政策と株価指数の動きを分析。

参考文献および引用・参考サイト

- データ分析入門 慶應 SFC データ分析教育グループ著 (2014年9月1日第7班)
- EXCEL ビジネス統計分析 末吉正成・美喜著
- NHK 特設サイト新型コロナウイルス  
<https://www3.nhk.or.jp/news/special/coronavirus/>
- NY DOW JONES 株価指数 2年間推移  
<https://stocks.finance.yahoo.co.jp/us/chart/>