

# 高等学校における確率分布と統計的な推測の教育

M2020SS004 大橋彪正

指導教員：小藤俊幸

## 1 はじめに

私たちの周辺には非常に多くの統計データが存在している。情報化時代と言われる今、社会生活においても多くのデータを整理し活用することが重要視されている。[1]

統計処理のルールを身につけていないために、間違った使われ方や人をだますような使われ方をされた場合にはそれを見抜く十分な力を持っていないとの不安にかられることも少なくない。こういった点から初等・中等教育段階で現実のデータに触れ、高等学校で本格的に統計教育に取り組むことになっていると私は考える。

新学習指導要領の目的は統計学の必修化、つまりはすべての高校生に統計学を学ばせることを目的としている。しかし私は問題点があると思う。それは十分な数学的な知識が準備されていないうちに勉強させないといけないという点である。特に文系の生徒は中々興味関心を持ってないと予想される。

そこで歴史とICTを用いた教育の方法を考えることを研究目標とする。歴史について各分野で触れることにより文系の生徒たちに興味関心を持ってもらうことが出来るのではないかと私は考える。また現在、GIGAスクール構想という言葉をよく耳にするだろう。ICT技術の社会への浸透に伴って、教育の現場でも最先端技術の活用が求められている。GIGAスクール構想はこうした社会の変化を受けて教育現場で生徒各自がパソコンやタブレットといったICT端末を活用できるようにする取り組みである。だからこそタブレットなどを用いてエクセルを取り扱うことによりICT教育を進めることが出来る。ただし、小学校・中学校でタブレットに慣れた生徒達が高校に進学してくるのでICT教育が可能であることを前提として考えた。

今回の授業案の基本として出来る限りその単元の歴史に触れ、内容を理解した後にエクセルを用いて計算・証明などを行っていくものとする。また授業を受ける生徒は文系の生徒を対象にするものとしている。

## 2 授業案

どう授業を行うのかを考えていくものとする。初めにその分野での目的を述べ歴史に触れ、問題に取り組んでエクセルを用いて考え、最後にまとめを述べた。中には問題ではなく導入・展開から述べてあるものもある。

### 2.1 外れ値 (数学 I)

外れ値において次のことを目的とする。

1. エクセルを用いて外れ値を理解する。

統計ダッシュボードに国や民間企業が提供している主要な統計データをグラフで提供している話をする。[5] この

サイトを開き、様々な統計データがあることを見て感じてもらう。このサイトから宿泊人数のデータを開ける。その他の機能からエクセルの状態データをダウンロードし、今回の問題へと移る。

【問題】2019年度の宿泊者数に外れ値は存在するのか。存在する証明をせよ。また存在するとするならばどのような理由であるのか。

【外れ値におけるエクセルでの計算】まずは2019年度のデータだけ取り出し表にまとめる。箱ひげ図を作りたいので月ごとの人数の部分を選択し挿入のところから箱ひげ図を選んで図を表示する。この時、周りが見ても分かるようにデータラベルの書式で値を表示にすると外れ値や中央値、最小最大値などの値が表示され見やすくなる。

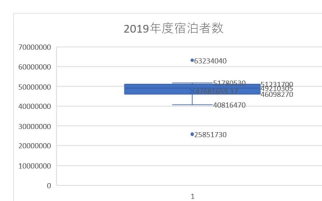


図1 excel における箱ひげ図

外れ値がある場合、図1のような箱ひげ図を用いる。

【まとめ】1. 本当に外れ値なのか。「異常値」と区別することも教える。外れ値は除外すべき値と捉えがちだが、その背景を探ることが大切である。今回は8月はお盆や学生の夏休み、2020年の3月といえばコロナウイルスの広まりなどと言った背景が読み取れる。背景に問題発見や問題解決の手がかりがあることもある。何か原因があるのではと考えることが出来る。

### 2.2 仮説検定の考え方 (数学 I)

仮説検定の考え方において次のことを目的とする。

1. ペアワークなどを通して具体例を考えさせ、仮説検定の考え方を理解する。

【仮説検定における歴史】仮説検定の考え方はどこから生まれたのかについて話す。アーバスナットという人物が男女の数の相違について考えたのがきっかけである。1712年にアーバスナットが「両性の出生において観察される一定の規則性からとられた神の摂理に対する推論」で神の摂理が持ち出される。[2] ここで述べられていたのが「今、保たれている男女それぞれの人数の間の正確な均衡状態に注目すべきだ。これは神の摂理以外何ものでもない。このことを証明しよう」と述べた。これを証明するために男が女より多く生まれる年を男子年、そうでない年を女子年

とした。男女出生が同等のチャンスによって起こるとすれば、男子年と女子年の分布も公正なコイントスの結果の分布と似るべきだと考えた。これが史上最初の統計的仮説、仮説検定の考え方である。

【導入・展開】授業者が仮説検定の具体例を話し、帰無仮説や有意水準の説明をする。その後、この世の中にある仮説検定が行われているものをペアワークやグループワークを通して何がわかるのかを考えてもらい、何人が指名し発表してもらい、意見を交流しあう事により具体的に仮説検定とはなにか、日常の何に使われている考え方なのかという理解を深めることが可能である。

【まとめ】1. ペアワークなどを通して仮説検定がどういうものなのかを理解を深めることが可能である。また最後にコロナに対するワクチンの話が発表の中で出なかった時に一番生徒の身近なものとして仮説検定を経て現在出回っているのがコロナに対するワクチンというものであると話をする。私達が使用する薬やワクチンなどの効果はある程度保証されているのも、実は仮説検定による基準が設けられているからであると話を進めていく。

### 2.3 期待値 (数学 A)

期待値において次のことを目的とする。

1. 歴史・問題から期待値を理解する。
2. 日常的なデータを用いてエクセルを利用し求める。

【期待値における歴史】期待値の考え方はどこから生まれたのか。まずは生徒に考えて答えてもらう。その後答え合わせとして賭博から生まれたものである話から始める。[3] パスカルとフェルマーによるサイコロ賭博をテーマにした書簡のやりとりは多くの人々の関心の的であった。それもお金儲けの為に皆興味があったのである。こうして確率論という学問が始まったのである。

【問題】「企業 A」と「企業 B」の2つの企業がくじ引きを開催している。どちらもくじは100本あり、1000円で引くことが出来るとする。どちらかの企業でくじ引きを引くとした時、どちらで引く方が得であると言えるのか。

企業A(賞)	1等	2等	3等	外れ	計
賞金	10000円	5000円	1000円	0円	
本数	5本	15本	30本	50本	100本

企業B(賞)	1等	2等	3等	外れ	計
賞金	50000円	10000円	2000円	0円	
本数	1本	4本	20本	75本	100本

【期待値におけるエクセルでの計算】確率と賞金の掛け算を行いその後 sum を使い期待値を求める。

【まとめ】「企業 A」の期待値は1550(円)、「企業 B」の期待値は1300(円)となり、期待値が大きいのは「企業 A」である。よって企業 Aの方が得である。

1. 期待値を使った計算は何から始まったかの話をしていく。それは賭博(賭け事)から確率論は学問として始まっ

てきた。そのような話をして興味関心を深める。

2. エクセルを使って求めるのは簡単である。エクセルを用いることによりエクセルで出来ることの視野も広がっていくことが可能である。また日常的なデータをもエクセルで扱うことが出来た。

### 2.4 二項分布 (数学 B)

二項分布において次のことを目的とする。

1. エクセルを用いて反復試行の計算かつ二項分布のデータの表示を出来るようにする。

【二項分布における歴史】確率論という学問がどこから始まったのか期待値での説明の時話していた事を復習として生徒に問う。それは賭博(賭け事)から始まったのである。[3] 二項分布こそ賭博に用いられていた考え方である。例えばサイコロの目が次は1が出る確率を分布的に考えたりすることが可能である。こういった話をして復習かつ興味関心を持たせる。

【問題】打率3割バッターが5打席中4打席以上でヒットを打つ確率を求めよ。

【導入・展開】「ヒットを打つ」または「打たない」のいずれかの事象しか持たないためこの確率は二項分布を用いて計算できる。

【二項分布におけるエクセルでの計算】combin 関数を使いヒットを打つ確率をそれぞれ求める。例えば  $COMBIN(5, B6) * (3/10)^{B6} * (7/10)^{(5-B6)}$  を打ち込むと5打席中ヒットを打つ回数が0回であるときの確率(今回で言うと0.16807)と表示される。ヒットを打つ回数を1, 2, 3, 4, 5回と繰り返していき、その後 sum 関数を使い4回ヒット打つ時と5回ヒット打つ時の合計を求める。combin 関数を使いヒットを打つ確率を求めてその後 sum 関数を使い4回ヒット打つ時と5回ヒット打つ時の合計を求める。

【まとめ】つまり3割バッターは3%の確率でその日の試合で4打席以上ヒットを打つことが可能である。

1. combin 関数を今回利用することにより、エクセルで解けるようになる。また打席数を増やすとどうなるのかを考えさせ、解く時間を設ける。実際にエクセルを用いて打席数25回の時と打席数50回を時のグラフを表示してみる。すると少しずつ正規分布曲線に近づく事が分かる。

### 2.5 正規分布 (数学 B)

正規分布において次のことを目的とする。

1. 歴史・問題からデータが多いほど正確になることを学ぶ。

【正規分布における歴史】なぜ保険会社が今でもあるのかについての話をする。それは1600年代にハレーという人物が死亡年齢の統計的解析を行ったからである。[3] 何十年何百年をかけて人が死亡する年齢というものをデータとして集め続けてあったおかげで保険会社は購入者の年齢に合わせた値段で販売し、安定した運営が出来るようになった。肝心なのはデータが多ければ多いほど正確になるということ伝える。また、データの数が少なければ正規

分布になるということも言えないだろう。模試のデータが点数が高得点・低得点の人しかいなかったら正規分布に従うとは断定するのは難しい。

【問題】次のうち、正規分布から離れていると予想されるものを選び。

1. 全国高校2年生男子の身長分布
2. 200年分の岐阜市の降水量分布
3. 年収分布
4. 大規模な模試の点数分布

【導入・展開】まず正規分布の性質の話をしてからこの問題に取り組む。よって正規分布がどのようなものなのかのイメージをつけさせるため今回このような問題を扱った。

1 は生物学的に 2 は自然現象より平均値の近くに多く数があることが予測でき、平均値から離れるほど数が少なくなることが予測される。またそれは同様に 4 でも言えるだろう。よって 1,2,4 は正規分布に近いものであると予測される。3 については具体的なデータを見せて話をしていく。平均から大きく離れた年収 1000 万以上稼ぐ人は一定数いたり、それ以上と稼ぐ人もいる。よって今回 3 のデータはグラフより正規分布であると言うことが出来ない。

【まとめ】問題文に正規分布に従うと書いていない時や実際に実験しグラフを作ってみない限り正規分布であると確実に言うことは出来ない。

1. 歴史について触れることによって興味関心を持たせることができる。データの数が多ければ多いほど的確になることも理解できるだろう。また実際の年収について日常の話に触れることによってより正規分布について理解が深まり、統計学に対して興味を持つことができるであろう。

## 2.6 中心極限定理 (数学 B)

中心極限定理においては実際に指導案を作成した。

数学 B 学習指導案				
授業者 大橋 直正				
1. 本時の目標	中心極限定理について具体例を用いて理解させる。具体的にグラフやエクセルなどを用いて理解させる。			
2. 指導に当たって	<ol style="list-style-type: none"> <li>1) 生徒観 高校2年生文系クラスに対して授業するものとする。数学Ⅲ+Ⅳを今後扱う生徒ではないことに注意し授業を行う。</li> <li>2) 教材観 パソコンまたはタブレットなどを使いエクセルやグラフを用いて中心極限定理を理解させる。</li> <li>3) 指導観 次回の推定、検定に繋がる内容なので理解している小規模的に机間指導・発問を意図して行いたい。</li> </ol>			
時間	学習内容	生徒の学習状況	評価 指導上の留意点	
導入 (5分)	<ul style="list-style-type: none"> <li>・宿題の確認</li> <li>・母集団と標本の違いの確認</li> <li>・標本平均の求め方の確認</li> </ul>	<ul style="list-style-type: none"> <li>・ノートを授業者に見せる。</li> <li>・違いについて積極的に発言を求める。</li> <li>・標本平均についての公式を覚表形式で答えてもらう。</li> </ul>	<ul style="list-style-type: none"> <li>主体的に学習に取り組み態度</li> <li>知識・技能</li> </ul>	<ul style="list-style-type: none"> <li>扱った生徒は次回必ずやるように指導。</li> </ul>
展開 (20分)	<ul style="list-style-type: none"> <li>・中心極限定理について説明する。</li> <li>・nの値が10, 30の時のヒストグラムを書かせる。</li> </ul>	<ul style="list-style-type: none"> <li>・n=5のゲーム、n=10のゲームをそれぞれ何回かくり、PCまたはタブレットを用いてエクセルで標本平均を求めてそれをヒストグラムにさせる。</li> <li>・サンプル数は多い方が正確に出るので100-200行してもらおう。</li> </ul>	<ul style="list-style-type: none"> <li>主体的に学習に取り組み態度</li> <li>知識・技能</li> <li>思考・判断・表現</li> </ul>	<ul style="list-style-type: none"> <li>各グループ、エラーが出ていないか、間違えていないかを注意し机間指導に時間をかけよう。</li> <li>またサンプル数が多い方が良いという点とnの値が大きいと正規分布に近づくと(中心極限定理)点を理解させることを目的としている事を忘れないでいく。</li> </ul>
確認次へ (11分)	<ul style="list-style-type: none"> <li>・様々な母集団のデータをを用いて他でも同様に考えさせる。</li> <li>・次の検定へと移る。</li> </ul>	<ul style="list-style-type: none"> <li>・前回の年収のデータや箱がずって同じデータ、二項分布なども同様に考えさせ、発問させる。</li> </ul>	<ul style="list-style-type: none"> <li>主体的に取り組み態度</li> <li>知識・技能</li> <li>思考・判断・表現</li> </ul>	<ul style="list-style-type: none"> <li>先ほどのエクセルでの証明からどんな母集団でも出来る事を再確認する。</li> </ul>

図2 中心極限定理の指導案

ここで言うのは中心極限定理の説明とエクセルでの証明である。またこの定理は統計学において正規分布の果たす役割がとても大きいことを確認することが出来るものでもあることを伝え授業を終える。[6]

## 2.7 推定 (数学 B)

推定において次のことを目的とする。

1. 歴史を知り、興味関心を持たせる。
2. 推定の計算を出来るようにする。

【推定における歴史】20世紀にはいつてロナウド・フィッシャーという人物が集計されたデータは大きな母集団のうち小さな標本にすぎないと考えた。[3] 部分的であるとされたデータから母集団の性質を見極めようとする考え方が推測統計学、頻度論と呼ばれる考え方である。つまりは標本から母集団を推測する推定という概念が生まれたのである。だからこそ正規分布における歴史で話した話に繋がり、人の死亡年齢が推定することが可能となって運賃であった保険事業も購入者の年齢に合わせた確な値段で保険を販売し、今の保険会社が存在する。次の話は手計算で推定の問題を解いた後に触れるものとする。

【推定におけるエクセルでの計算】サンプル調査によりデータを収集した場合はその平均値に「どれぐらいの信頼性があるか」を検証しておくのが基本である。[4] だから最後に平均値の信頼性を確認するエクセルにおける「95%信頼区間」の算出方法を教えて授業を終えたいと考える。

テストの点数			
点数 (点)			
63	標本平均	80.2	
80	標準偏差	16.34503	
97	確	1.96	
55	標本の大きさn	10	
85	平均値の誤差	10.13075	
100			
53	平均値の95%信頼区間		
91	区間の最大値	80.33075	
59	区間の最小値	70.06925	
85			

図3 平均値の95%信頼区間

標本平均はあくまで「10名のサンプルの平均」を算出したものでしかなく、必ずしも母集団の平均値(母平均)に一致するとは限らない。むしろ誤差を含んでいると考えるのが自然である。

【エクセルで計算を行う場合】ここからエクセルでの推定の仕方を説明していきたいと考える。AVERAGE関数で標本平均を求めた後に95%信頼区間より1.96を使用するので書き込んでおく。また今回の標本の大きさは10なのでそれも書き込んでおく。その後平均値の誤差を $1.96 * (\text{標準偏差 } 16.34503) / \sqrt{10}$  で求める。あとはこの数値を「標本平均(80.2)」にプラスマイナスをすると区間の最大値と区間の最小値を求める事が出来る。

【まとめ】この結果を見ることである程度は標本平均の信頼性を探る事が出来る。今回の例における平均値の

95% は 70~90(点) となり, 全体で 20(点) もの幅がある。これは信頼性の高い指標ということが出来ないであろう。また今回のようにサンプル数が 10 といった非常に小さい時正確な値が出ないにも注意させる。

1. 歴史に触れることにより今後保険会社に勤めたいと考える生徒などに統計学の考え方の重要性を伝えることが可能である。また推定の考え方は標本からある程度, 母集団の平均を推定できることは他にも利用できるのではないかと考えることが出来る。
2. エクセルでも計算が出来るようになった。またこの部分は共通テストなど入試問題によく出る所である。そのことを理解させて多くの演習を行う。

## 2.8 仮説検定 (数学 B)

仮説検定において次のことを目的とする。

1. 統計学の歴史に触れて重要性を改めて理解する。
2. 用語の意味を理解する。
3. 帰無仮説, 対立仮説を立て検定方法を理解する。

【今日における統計学の歴史】仮説検定の考え方はどこから生まれたのか。数学 I 「仮説検定の考え方」で一度歴史に触れているが忘れていた人も多いためもう一度話す。この際, 帰無仮説, 対立仮説がどうなるのか述べて現代の統計学についての話へと進む。また仮説検定は数学 B における最後の単元であることより現在の統計学における話に触れ, 仮説検定への説明へと移りたいと考える。統計学の概念というものは今まで触れてきたように昔から使われていたが, 学問としての統計学はまだ登場して日が浅い。[3] しかし統計質保証推進協会と日本統計学会による統計検定という資格試験が 2011 年から始まっており, その受験者数は年々と増加し続けている。このことから統計学は人気な学問になりつつある。医療や保険事業の運営など様々な場面で利用されて社会の発展に貢献している。これから先も自分たちの生活と密接に関わりながらますます発展していく分野であることを意識してもらうためこの話をして授業を進めていきたいと考える。

【導入・展開】用語を覚えているのかを確認するため, グループワークを通して話し合いをさせ発表してもらう。その後, 対立仮説の話を行って次に検定の方法についての話を行う。問題に応じて両側検定, 右側検定, 左側検定どれを用いて解いていくのかを考えなければならない。

【問題】ある機械が袋に詰める塩の重さ (単位 g) は正規分布  $N(100, 25)$  に従うように調整される。機械が正しく調整されているかどうかを確かめるために無作為に 10 個の袋をとって塩の重さを測ったところ, 平均  $\bar{X}$  は 101.5 であったという。このとき, 仮説「この機械は正しく調整されている」を検定する。この時の帰無仮説, 対立仮説はどうなるのか, 検定はどう考えるのか。[7]

【導入・展開】今回の問題は本来の母平均 100 が変化したかどうかを調べるのが目的であることから両側検定を行う。そのことから帰無仮説  $H_0$  および対立仮説  $H_1$  を次

のようにおくことが出来る。

$H_0 : m = 100$  (母平均は変化していない)

$H_1 : m \neq 100$  (母平均は変化した)

【まとめ】今回用語の意味の理解や帰無仮説・対立仮説を立てることが出来るのか, 検定方法の判別を出来るようにすることに重点をおいた。次回の授業で具体的な検定の計算を行っていきたいと考える。

1. 現在の統計学について触れることによって改めて最後に重要視されていることを伝えて意識的に統計学を勉強させたいと考える。また日常生活に密接に関わっている学問である話もして興味関心を持たせたい。
2. グループワークを通して考えることにより以前習った復習かつ用語の意味を再理解することにつながる。
3. 仮説検定においてすぐに問題を解くのではなく帰無仮説・対立仮説を上手く立てることが出来るのか, 検定方法を間違えていないかを丁寧に確認して問題へ取り組みたいと考える。

## 3 おわりに

今回, 新学習指導要領の実現にあたって歴史と ICT を用いた教育の方法を考えることを研究目標としていた。

歴史を用いて授業を行うことは文系の生徒に対して興味関心を持たせる良い効果が生まれる。また ICT を用いて授業を行うことは文系の生徒に対して ICT 教育の発展に繋がる。歴史と ICT を用いることによって統計学の考え方の本質に少しでも触れることが出来る。数学の考え方とはあえて別物として触れてあげるのが本質的に理解しやすいのかもしれない。重要視されている統計学についてより深く理解して社会に出た際に活かすことが出来ると考える。

このように歴史と ICT を用いることは生徒にとって良い効果を生むと言える。

## 4 参考文献

- [1] 柴山宮恵子: 『やさしい統計学』.2014.
- [2] 安藤洋美著: 『確率論の生い立ち』.1992.
- [3] y0he1 著: 『統計学の歴史 古代ローマから現代まで』. 2020.  
<https://ai-trend.jp/basic-study/basic/history/> (最終参照 2022.1/22)
- [4] 相澤裕介: 『第 59 回 Excel データ分析の基本ワザ』.2020.  
<https://news.mynavi.jp/techplus/article/excelanalytics-59/> (最終参照 2022.1/23)
- [5] 総務省: 『統計ダッシュボード』2017.  
<https://dashboard.e-stat.go.jp/graph?screenCode=00480> (最終参照 2022.2/1)
- [6] 石井俊全: 『意味がわかる統計学』.2019.
- [7] 高遠節夫ほか 5 名: 『新 確率統計』.2020.