第5部

統計的探究の実践 Ⅳ
～標本データから全体を推測する～

1　どの味のラーメンが好まれるだろうか？［標本誤差の評価］

ラーメンは、中国の麺料理が日本に伝わったものであるが、日本各地で独特の進化を遂げ、今や日本の国民食の地位を得ているといっても過言ではない。それにとどまらず、中国への逆輸出を始め、海外にも盛んに進出していて、世界のあらゆる都市で日本のラーメン屋が店を構えるほどになっている。

ラーメンの味には大きく分けて、みそ、しょうゆ、塩、とんこつの4種類があるが、はたしてどの味のラーメンが最も好まれているのであろうか。ラーメンの味に関する調査は数多く行われていて、オリコンスタイル・トレンドリサーチの調査によると、日本全国では4種類の味の中で最も好きなものの割合は表1の結果であった。

表1　ラーメンの味の好みの調査結果

しょうゆ	みそ	塩	とんこつ	合計
28.6%	25.3%	22.0%	24.1%	100.0%

表1は全国の結果であるが、ラーメンの味の好みには地域差があり、北海道・東北地方ではみそ味が好まれ、九州・沖縄地方ではとんこつ味の人気が高いとの調査結果となっている。さらに細かく見れば、ご当地ラーメンがそれぞれの地方にあり、味の好みも千差万別のようである。そこで、あなたのいる地方では、どの味のラーメンの味が好まれるのかに興味がわくのではないか。これは単なる興味ではなく、地域のラーメン屋にとっては死活問題であるし、スーパーマーケットなどでの即席ラーメンの仕入れ量にもかかわってくる問題である。

STEP 1：Problem 問題　課題の設定

◇ 自分の住む地域において、ラーメンの4種類の味（しょうゆ、みそ、塩、とんこつ）のうち、どれが最も好まれるかを知りたい

STEP 2：Plan 計画　どのようなデータ・統計資料を集めて分析するか

◇ 4種類の味のうちどれが最も好まれるかについて、その結果が妥当かを検討する

データの集め方には種々のものが考えられる。たとえば、ラーメン屋に行ってどの味のラーメンの売り上げが最も多いかを調べる、スーパーマーケットに行ってどの味の即席麺が最も売れているかを調べる、などである。しかし、街のラーメン屋がみそラーメン専門店であったり、スーパーの売り上げは企業秘密であったりと、望ましいデータが得られないことも考えられる。そこで、**アンケート調査**を実施することとする。

　アンケート調査は比較的手軽に実施できることから、あまり計画性もなく行われることも多いが、正しい結果を得るためには相応の計画と準備が必要である。まず、母集団をある程度特定する必要がある。母集団をあらかじめ特定し、そこからデータを得るのが原則であるが、逆に、実際に得ることができるデータは何かという考察から母集団が決まることもあるだろう。母集団からの無作為（ランダム）な標本抽出を原則とするならば、調査可能でない部分集団は母集団に含め得ないことになる。

データは母集団から無作為に抽出するのが理想的であるが、実際問題ではなかなか無作為抽出は難しい。したがって、性別や年齢層など、結果に影響を及ぼすと想定される要因については、なるべく偏らないようにすること！
また、回答拒否などによるデータの欠損（欠測）への対処法も考えるのが大事だね。実際のデータ収集では、計画通りにデータが得られることは稀であり、それへの対応もできる範囲で事前に想定しておくことが必要だ。

Q１：母集団からの偏りのない調査法にはどのようなものがあるだろうか。また、はたしてそれは実施可能か。もし、実施が難しかった場合、代わりにどのような調査法があるだろうか。

Q２：インターネット調査では偏りが起きる可能性が否定できない。起きうる偏りにはどのようなものがあるだろうか。

STEP 3：Data 収集　必要なデータ・統計資料を集める

◇ アンケート調査でデータを収集しよう！

ここでは、高校の生徒たちが自分自身を含め、彼らの親兄弟や身近な人からデータを得ることを想定する。高校のあるクラスの生徒のラーメンの味の好みが地域の他の集団と著しく異なるのであれば、この調査結果を地域の住民全体に一般化することはできないが、そのようなことはないと仮定する。

何人分のデータを集めたら良いかは重要な問題である。データ数が少ないと確かな結論は得られないが、多くのデータを集めるには時間も費用もかかる。そこで、参考となるのが標準偏差の値である。サンプルサイズを n として母集団比率 p を標本比率 $\hat{p}$ で推定する場合、$\hat{p}$ の標準偏差は $\sigma(\hat{p})=\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ である。この値は p が分からないと計算できないが、その最大値は $p=0.5$のときであり、$\sigma(\hat{p})=\frac{1}{2\sqrt{n}}$ である。仮に、標準偏差の目標値が５％であったとすると、$\frac{1}{2\sqrt{n}}\leqq 0.05$ より、$n\geqq\frac{1}{4\times 0.05^2}=100$ が得られる。そこで、サンプルサイズを100以上にすれば、すべての p に対して標準偏差を５％以下にできることになる。

Q３：母集団比率が0.3程度とみなされるとき、標準偏差を２％とするためには、サンプルサイズ n はいくつ以上必要か。

世論調査の有効性

2016年には、当初の調査結果を覆す「事件」が２つ起きました。１つは英国の国民投票による EU 離脱の決定であり、もう１つは米国大統領選挙でのトランプ候補の勝利です。とくに、トランプ候補の勝利は、それまでの世論調査の結果が概ねクリントン候補の勝利と出ていただけに、大いなる驚きを与えたと同時に、世論調査への疑問も投げかけられる結果となりました。

大統領選挙当日まで多くの調査が行われましたが、選挙直近ではどの調査においても、両候補の支持率は極めて接近したものでした。したがって、世論調査で導くべき結論は「支持率は接近していて、どちらが勝ってもおかしくない」であったはずです。その証拠に、全米での支持率はクリントン候補が上回っていましたが、トランプ候補が大統領に選出されたのは米国の大統領選挙の仕組みによるところが大きかったといえます。

日本でのマスコミ論調では、トランプ候補を揶揄したような報道が多く、それに接した日本人の導いた結論では、圧倒的にクリントン候補に支持が集まっていました。日本でのデータに基づかない判断よりも、米国でのデータに基づく調査結果のほうが正しく結果を言い当てていたのです。

いろいろな調査結果を比べてみると、好みの味の割合が調査ごとに異なっていることが分かる。これは、どの調査結果が正しくて他の調査結果が間違いということではない。日本のすべての人のラーメンの味の好みなどというのは、もちろん調べることはできない。調査は、日本全体の人々の中から選んだごく一部の人に対して行われるもので、調査ごとに対象者が違うため調査結果がそれぞれ異なるのは当然であるといえよう。

統計用語では、調査対象となる全体の集団を母集団といい、実際に得られたデータを標本（サンプル）という。そして母集団から標本を得ることを**標本抽出**（**サンプリング**）という。調査では、何人の人から得たデータであるか（これを**標本の大きさ**、あるいは、**サンプルサイズ**という）が重要な鍵である。また、得られた標本が母集団の特徴を偏りなく表しているかどうかも重要な視点となる。

標本抽出で、標本に含まれる人が調査ごとに異なることによる不可避的な調査結果のバラツキを**標本誤差**という。調査の計画が妥当なものであれば、標本誤差はサンプルサイズが大きいほど小さくなり、調査結果は信頼に足るものとなる。しかし、標本抽出が妥当なものでなく、標本が母集団の特徴をうまく表していないとすると（これを**非標本誤差**という）、サンプルサイズをいくら大きくしても、調査結果は信用のおけないものとなってしまう。

STEP 4：Analysis 分析　統計量で傾向を捉える

◇ アンケート結果を母集団における値の推定値とみなし、その標準偏差を計算する

データが得られた場合には、まず、データのクリーニング、すなわち年齢が180歳などというあり得ないデータを拾い出し、訂正すべきものは、根拠をもって訂正しなくてはならない（根拠のない訂正はデータの改ざんにつながる恐れがある）。その後、集計表の作成やグラフ化により結果を分かりやすく表示する。そしてその後、必要に応じて標準偏差の計算など推測統計的な手法を適用する。

たとえば、100人の生徒への調査で、表２のようなデータが得られたとする。みそ味の人気が高いように見えるが、そのように結論付けて良いだろうか。

表２ ラーメンの味の好みの割合と標準偏差

好みの味	しょうゆ	みそ	塩	とんこつ	合計
割合	22%	42%	19%	17%	100%
標準偏差	4.52%	4.35%	4.14%	4.28%	

各味の好みの割合の標準偏差を計算すると表２の下段のようになる。これによると、各好みの割合の平均は４％から５％程度の標本誤差を含んでいることが分かる。しかし、みそと２番人気のしょうゆとの差は20%もあることから、この調査から、みその首位は動かないものとみても差し支えないと考えられる。

コンピュータ・シミュレーション

標本誤差の大きさを実感するため、コンピュータを使って実験してみましょう。表１の割合を母集団における値として、ここからランダムに50人および200人を抽出して、それぞれの味を好む人がどれくらいの割合となるかについて、コンピュータの乱数を用いてシミュレーションします。ここでは、標本抽出のシミュレーションを100回繰り返します。

表３は、シミュレーションの最初の５回分の結果です。各値は、それぞれの味を好んだ人の割合と、表１の設定で最も人気の高かったしょうゆを好んだ人の割合の４つの味の中での順位です。サンプルサイズ n＝50の５回のシミュレーションでは、しょうゆが最も好まれたのは１回だけですが、n＝200では３回となっています。n＝50の第５回目のシミュレーション結果では、しょうゆの割合が最も小さくなっています。それに対して、n＝200では、しょうゆの順位は１位もしくは２位のみです。ここからも、標本誤差はそれなりにあること、およびサンプルサイズが大きいほど結果は安定することが分かります。

表３　シミュレーションの最初の５回分

n＝50	しょうゆ	みそ	塩	とんこつ	順位
1	0.26	0.28	0.22	0.24	2
2	0.22	0.28	0.20	0.30	3
3	0.28	0.36	0.22	0.14	2
4	0.28	0.20	0.24	0.28	1
5	0.22	0.30	0.24	0.24	4

n＝200	しょうゆ	みそ	塩	とんこつ	順位
1	0.320	0.275	0.185	0.220	1
2	0.310	0.240	0.210	0.240	1
3	0.260	0.285	0.215	0.240	2
4	0.295	0.280	0.220	0.205	1
5	0.255	0.235	0.230	0.280	2

表４に100回のシミュレーションにおける標本ごとの割合の平均と標準偏差を示します。表４からは、シミュレーション100回の平均は、n＝50と n＝200のいずれについても、表１の値に近いことが分かります。それに対して、値のバラツキを表す標準偏差は、n＝200のときの値が n＝50での値の約半分になっていることが見て取れます。また、図１は、100回のシミュレーションにおけるしょうゆの割合のヒストグラム（n ＝50, 200）としょうゆの順位の棒グラフ（n＝50, 200）です。図１の（a）と（c）より、n＝200では n＝50のときに比べて、しょうゆの割合の値のバラツキが小さいことが見て取れ、図１の（b）と（d）からは、しょうゆの順位が正しく１位となった回数は、n＝200では n＝50のときに比べて大きくなり、４位となってしまうことはなかったことが分かります。

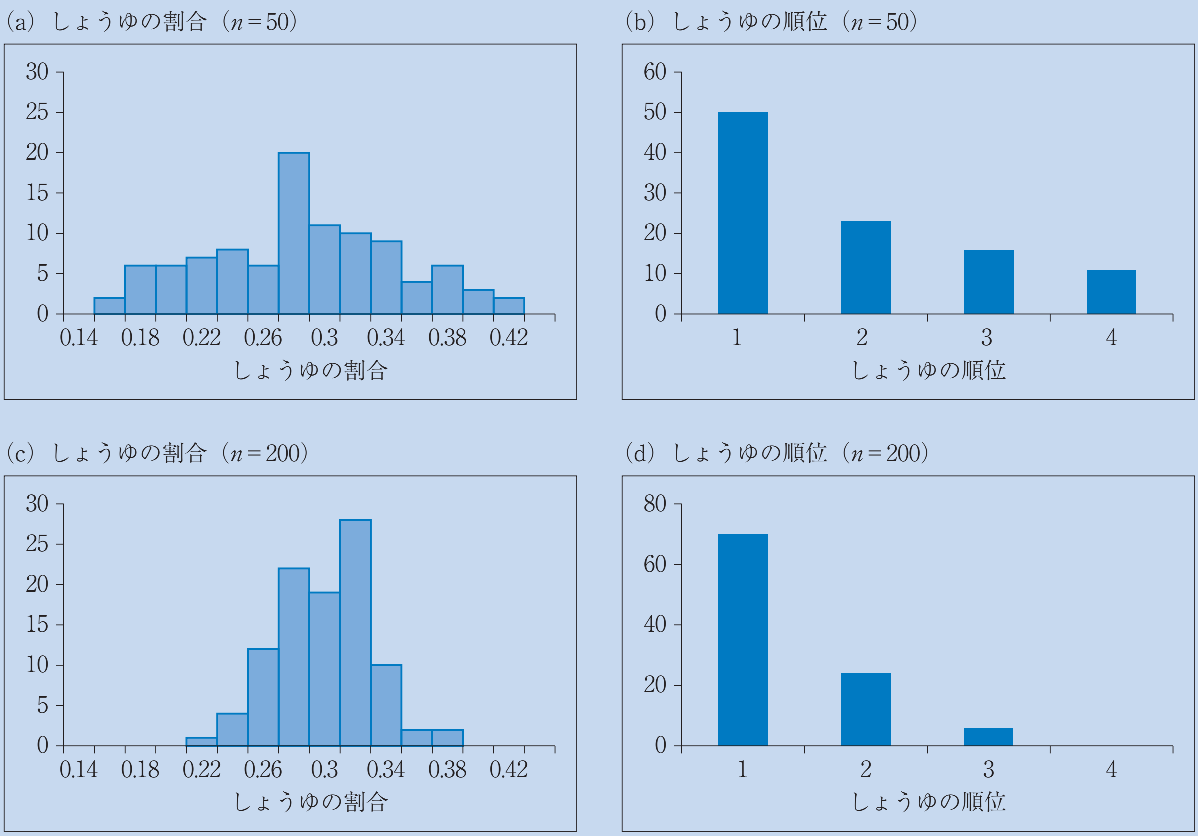
このように、標本抽出が妥当であれば（このシミュレーションでの標本抽出は完全無作為抽出といって、最も妥当な標本抽出法であるとされる）、サンプルサイズが大きければ大きいほど標本誤差は小さく、母集団での特徴を正しく表すことができることになります。

表４　100回のシミュレーションにおける標本別割合の平均と標準偏差

n＝50	しょうゆ	みそ	塩	とんこつ	順位
平均	0.285	0.253	0.219	0.242	1.880
標準偏差	0.062	0.054	0.054	0.064	1.047

n＝200	しょうゆ	みそ	塩	とんこつ	順位
平均	0.292	0.250	0.218	0.241	1.360
標準偏差	0.030	0.026	0.029	0.029	0.595

図１　100回のシミュレーションの結果のグラフ表示



コンピュータを用いて、母集団からの標本抽出をシミュレーションしてみよう。乱数発生は Excel の分析ツールの「乱数発生」で容易にできるよ。

STEP 5：Conclusion 結論　結論を導き、新たな課題を見出す

◇ 調査結果が妥当なものかどうかを吟味しよう！

データの収集と統計分析の結果、表２のデータからは、表１の全国における値に比べて、みそ味の人気が高いことが分かった。プラスマイナス４％程度の誤差はあるが、この地域でみそ味が好まれる割合はおおよそ40％である。

この結果を次の段階につなげることが重要だ。より一層みそ味のラーメンの人気を高めるべきなのか、あるいは、他の味のラーメンをさらに拡充していくべきなのか。いずれにせよ、意思決定には客観的なデータと統計分析の利活用が必要となる。

2　フライドポテトの重量は公表値と同じ？［区間推定］

ハンバーガー・チェーンは、M 社が1971年に東京・銀座に１号店を出店して以来日本全国に展開され、今やファーストフードの王様的存在になっている。ハンバーガー店での人気メニューの１つがフライドポテト（フレンチフライ）であり、通常、L、M、S の３サイズが用意されている。それぞれのサイズの重量は概ね決められていて、あるチェーン店では M サイズのフライドポテトは135g と公表されている。実際にハンバーガー店に行って注文すれば分かるが、フライドポテトの重量を逐一測ってお客さんに提供していたのではサービスに時間がかかるため、店舗スタッフが目分量で判断して提供していることが多い。そこで、本当にフライドポテトの重量が公式発表の135g となっているかどうかの疑問がわく。

STEP 1：Problem 問題　課題の設定

◇ あるハンバーガーショップのフライドポテト M サイズの重量は、公表値の135g なのだろうか？

STEP 2：Plan 計画　どのようなデータ・統計資料を集めて分析するか

◇ このハンバーガーショップから、M サイズのフライドポテトを実際に10個購入した上で重さを測定し、得られたデータに基づいて、公表値が適正かどうかについて区間推定を使って検証する

STEP 3：Data 収集　必要なデータ・統計資料を集める

◇ 購入日と時間帯を変えて、M サイズのフライドポテト10個の重量データを得よう

10個のフライドポテトの重量の測定結果は表１のとおりであった。

表１　10個のフライドポテトの重量

重量（g）	118	122	125	126	128	130	132	135	136	138

STEP 4：Analysis 分析　統計量で傾向を捉える

◇ データのバラツキについて評価しよう

購入したフライドポテトの重量の平均129.0g は公式発表の値よりも小さかったとはいえ、たった10個のデータであり、実際、表１からは公表値の135g 以上のフライドポテトも10個中３個あるのも事実である。データのバラツキが考慮されていないので、標本誤差も加味した上での分析が必要であろう。統計学の手法を用いて精緻に検証しよう！

前節では、調査には標本誤差が不可避的に生じることを学んだ。前節では比率のデータを用いたが、ここでは、フライドポテトの重量のような連続型のデータのバラツキとその評価法を考えよう。

データが得られたら、その特徴を示す基本統計量を計算するのが統計分析の第一歩である。表１のデータに Excel の「分析ツール」の「基本統計量」を適用すると表２が出力される。本節ではそれらの基本統計量のうち標準誤差、標準偏差、**信頼度**（95.0%）に着目する。

表２ Excel の「基本統計量」

重量	
平均	129
標準誤差	2.03
中央値（メジアン）	129
最頻値（モード）	＃N/A
標準偏差	6.43
分散	41.34
尖度	－0.80
歪度	－0.24
範囲	20
最小	118
最大	138
合計	1290
データの個数	10
信頼度（95.0%）	4.59

「統計情報」と「平均の信頼度の出力」にチェックを入れると求められる。数値は小数第３位を四捨五入している。
Excel において量的データの場合、最頻値は意味がないため、表２の最頻値の欄には＃N/A の利用不可のマークが表示されている。

n 個の測定値を $x_1,\ \cdots,\ x_n$ として、それらの標本平均を $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ とする。そして、**標本分散**および**標本標準偏差**（standard deviation）をそれぞれ

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,\ s=\sqrt{s^2}$$

とする。また、標準誤差（standard error）は $\frac{s}{\sqrt{n}}$ で求められる。

高等学校の教科書では標本分散の定義式において除数は n であるが、ここでは、大学以降での定義式および Excel での計算式に合わせて n−1としている。n が大きい場合には数値的にあまり差はない。

Q１：表１のデータを用いて Excel で表２を出力しなさい。また、標準誤差の値2.03は、標準偏差の6.43を $\sqrt{10}$ で除して得られることを確かめなさい。

Q２：ハンバーガーショップはフライドポテトの公表値135g をどのように捉えるべきであろうか。平均値が135g であれば良いのか。最小値が135g、すなわち135g を下回る商品があってはならないのか。135g を下回る商品が５％程度あるのは許容するのか。それぞれの基準を満たすためには、どのように商品の提供をするのが良いか。

標本標準偏差は測定データのバラツキを表す指標であり、母集団全体での値のバラツキを反映した値である。すなわち、標本標準偏差は母集団全体でのバラツキを示す標準偏差（母標準偏差 σ）の推定値であるともみなされる。したがって、サンプルサイズを大きくすると標本標準偏差は σ に近づく。それに対して、標準誤差は、その定義式の分母にサンプルサイズ n を含むことから、n を大きくすると０に近づく。

標準誤差の意味の理解には、前節でも議論した母集団からの無作為抽出が重要な役割を果たす。母集団からの無作為な標本抽出によって n 個の測定値を得て、それらから標本平均 $\bar{x}$ を計算する、という作業を考えよう。この「n 個の測定値から標本平均 $\bar{x}$ を計算する」という作業は、実際は１度きりしか行われていなくても、概念的には何回でも繰り返すことができる。そして、その度ごとに $\bar{x}$ が得られるが、標本抽出は無作為であるので、標本誤差により $\bar{x}$ はそれぞれ異なる値を取り、何らかの分布に従って分布することになる。これを $\bar{x}$ の標本分布という。母集団分布の**母平均**が μ で、**母分散**が σ^2 のとき、$\bar{x}$ の期待値は同じく μ で、分散は $\frac{\sigma^2}{n}$ となることが示される。この分散の平方根 $\frac{\sigma}{\sqrt{n}}$ を $\bar{x}$ の標準誤差と呼ぶのである。母標準偏差 σ は通常未知であるので、標本標準偏差 s をその代わりに用いた $\frac{s}{\sqrt{n}}$ を標準誤差と呼ぶことが多い。

前節では、推定値に、その精度の情報を加味して（推定値±推定値の標準偏差）と表示した。ここでは、推定値の標準偏差のことを標準誤差と呼んでいるので、これは推定値±標準誤差と表現することができる。表２の結果については、129.0±2.03となり、区間（126.97, 131.03）を意味している。その区間に確率的な意味を持たせたものが**信頼区間**である。

信頼区間

母集団を特徴付ける定数は**パラメータ**と呼ばれるが、ここではパラメータの推定について扱う。推定は１つの値で行われることが多い。たとえば、母平均 μ の推定値として標本平均 $\bar{x}$ といった１つの値が用いられる。このように、１つの値でパラメータを推定することを、１点で推定するという意味で点推定という。しかし、推定値の提示だけでは、それが精度の情報を持たないため適切であるとはいえない。そこで、標準誤差の値を加味した（126.97, 131.03）のような表示が用いられるが、そこで示される区間は確率的な意味が不明確である。そこで、確率的な意味を持つ区間を導出し、その区間を用いて母集団パラメータを推定しようというのが区間推定である。そして、そのときの区間を信頼区間という。n 個の測定値が正規分布に従う場合、α を**有意水準**として、$\mathrm{Z}(\alpha/2)$ を標準正規分布の上側 $100(\alpha/2)$ ％点としたとき、 $\mathrm{P}(\mathrm{Z}\leqq-\mathrm{Z}(\alpha/2))+\mathrm{P}(\mathrm{Z}\geqq\mathrm{Z}(\alpha/2))=\alpha$ である。ここで有意水準とは、ある事象が起きる確率が偶然であると判断する確率をいい、0.05または0.01が用いられることが多い。母集団パラメータのうち、母平均 μ の信頼度（**信頼係数**）$100(1-\alpha)$％の信頼区間は、

$\left(\bar{x}-Z(\alpha/2)\dfrac{\sigma}{\sqrt{n}},\ \bar{x}+Z(\alpha/2)\dfrac{\sigma}{\sqrt{n}}\right)$ で与えられる。

このハンバーガーショップで販売されているフライドポテトの母平均の信頼度95％の信頼区間を求める。母標準偏差 σ の値は与えられていないが、データから得られた値6.43が母標準偏差であるとしよう。すなわち、ここでは計算手順を理解する目的で $\sigma=6.43$ としておく。信頼度は95％であるので、α は0.05であり、標準正規分布表から $z(0.025)=1.96$ であるので、**誤差限界**は $1.96\times\dfrac{6.43}{\sqrt{10}}\approx3.99$ となる。したがって、信頼度95％の信頼区間は、$129.00\pm3.99=(125.01, 132.99)$ と計算される。

Ｑ３：信頼度を90％あるいは99％として信頼区間を求めなさい。

STEP 5：Conclusion 結論　結論を導き、新たな課題を見出す

◇ 公表値135g は怪しい！

母標準偏差を既知とした場合、信頼区間は公表値135g をその中に含んでいない。信頼度95％の信頼区間が公表値135g を含まないということは、信頼度95％でこのハンバーガーショップの販売するフライドポテトの母平均が135g ではないと言い切れる。ただし、10個の観測データには大きな標本誤差が存在することを踏まえたうえで、結論を導かなくてはならない。母平均についてさらに精度の高い推定が望まれるのであれば、サンプルサイズをもっと増やした調査が必要である。

Ｑ４：信頼区間幅を狭めるための手立てとして、どのようなものがあるのかについて議論しなさい。

3　フライドポテトの重量は公表値通りか？［統計的検定］

◇ 勘と経験よりも客観的なデータ分析！

勘や経験だけでなく、客観的なデータの分析や数学的な計算に基づく行動や決定は、多くの場合、良い結果を生むことになろう。良い結果を生まないまでも、データ分析や数学計算の限界を知ることができる。そのためには、基本的な事項を学習し、実際の現場に応用する力を養う必要がある。努力は必ず報われる。報われない努力はないと思いたい。

STEP 1：Problem 問題　課題の設定

◇「ハンバーガー店のフライドポテトの重量が公表値通りかどうか」を検証する

前節で、あるハンバーガー店のフライドポテトのＭサイズの重量が公表値135g かどうかについて、信頼区間を用いて検証した。ここでは、それについて統計的検定という方法論を用いて評価する方法を実践する。

STEP 2：Plan 計画　どのような方法で分析するか

◇ 統計的仮説検定を行う

「駅前のハンバーガー店のＭサイズのフライドポテトの重量が公表されている通りかどうか疑わしい」と考え、これを検証するために、帰無仮説として「フライドポテトの重量が公表値135g とおりである」と設定し、購入データに基づいて仮説検定を行う。

帰無仮説の棄却とは

統計的検定では、Ｐ値という値が有意水準（有意確率）よりも小さいとき、帰無仮説が正しくてありそうもないことが起きたとは考えないで、帰無仮説を棄却する。すなわち、帰無仮説は誤りであると判断する。しかし、Ｐ値が有意水準よりも大きいときでも、帰無仮説が正しいことが証明されたわけではない。そのことを表現するため、帰無仮説を棄却しないといい、帰無仮説を採択するとは言わない（「採択」に正しいと認めるという意味が含まれているため）。

たとえば、ある高校には男子生徒と女子生徒とが半分ずついるという帰無仮説を考えよう。すなわち、男子生徒の割合を p としたとき、帰無仮説は $H_0 : p=0.5$ である。いま、この高校からランダムに４人の生徒を抽出して性別を調べたところ、全員が女子であったとしよう。そうなる確率は $(0.5)^4=0.0625$ であり、有意水準を $\alpha=0.05$ とすると、確率はそれよりも大きいため、帰無仮説は棄却されない。このとき、帰無仮説が正しく、この高校には男子生徒が半数いるといって良いだろうか。実はこの高校は女子高であって、性別を何人調べても全員が女子である可能性はあるのである。なお、５人調べて全員が女子であれば、帰無仮説の下でそうなる確率は $(0.5)^5=0.03125$ であるので、H_0 は棄却される。

すなわち、帰無仮説が棄却されないときは、単なる情報不足という可能性を含んでいるのである。

STEP 3：Data 収集　仮説検定のためのデータの収集と整理

◇ 実際にフライドポテトを購入してその重さを測る

高校生のグループが手分けして、駅前のハンバーガー店のＭサイズのフライドポテトを10個購入し、その重量を計測した。その結果を下記の表１に示す。

表１　購入した10個のフライドポテトの重量

重量（g）	120	124	126	130	130	131	132	133	134	140

このデータについて、Excel の分析ツールの「基本統計量」で計算した出力結果が表２である。このデータを用いて、このハンバーガー店のフライドポテトの重量が公表値135g といえるかどうかを統計的検定の枠組みで検討する。まず、統計的検定について簡単に述べておく。

表２　Excel の「基本統計量」

重量	
平均	130
標準誤差	1.77
中央値（メジアン）	130.5
最頻値（モード）	130
標準偏差	5.60
分散	31.33
尖度	0.53
歪度	－0.14
範囲	20
最小	120
最大	140
合計	1300
データの個数	10
信頼度（95.0%）	4.00

STEP 4：Analysis 分析　帰無仮説の下での確率の評価

◇ 帰無仮説の下で検定のための統計量を求め、その実現する確率を導く

母集団の分布は、母数（パラメータ）によって規定される。いま、駅前店で販売するフライドポテトの重量の分布が正規分布に従っているとすると、分布の型は、平均 μ と分散 σ^2の２つのパラメータで定まる。ここでは、店で販売するフライドポテトの重量が135g であるかどうかが問われているので、$\mu=135$と判断することが適当か否かを統計的に検定することになる。

このとき、帰無仮説 H_0は H_0：$\mu=135$と表される。それに対して、帰無仮説が否定されたときに採用する仮説を対立仮説（alternative hypothesis）といい、それを H_1で表す。対立仮説には、**両側仮説**と**片側仮説**の２種類がある。両側対立仮説は H_1：$\mu\neq135$と表される。検証すべき値135g の両側を考えている。

一方、片側仮説は135g より多いか、少ないかの片側を考えており、H_1：$\mu>135$または H_1：$\mu<135$のいずれかとなる。両側と片側のいずれの仮説を立てるかは問題による。

ここでの設定については、対立仮説として、消費者側として μ が135g より小さいかどうかに関心があるのであれば、H_1：$\mu<135$の片側仮説を選択するが、ハンバーガー店としては、逆向きの H_0：$\mu>135$に興味があるかもしれない。そこで、ここでは中立的に両側仮説の H_1：$\mu\neq135$を立てることにする。

いま、店で販売するフライドポテトの重量が正規分布 $N(\mu, \sigma^2)$ に従っているとして、高校生のグループが無作為（ランダム）に購入した10個のフライドポテトの重量を（$X_1, X_2, \cdots X_{10}$）としよう。そのとき、それらの標本平均 $\overline{X}=(X_1+\cdots+X_{10})/10$ は正規分布 $N(\mu, \sigma^2/10)$ に従う。母分散 σ^2 が既知であるとすると、X を標準化した統計量 $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{10}}$ は標準正規分布 $N(0,1)$ に従う。

分散が既知で $\sigma=6$ とすると、帰無仮説 $\mu=135$ の下で、データから求めた Z^* の値は、$Z^*=\dfrac{130-135}{6/\sqrt{10}}\approx-2.64$ となる。したがって、両側対立仮説に対する（両側）P 値は、正規分布表から $P=P(|Z|\geq|-2.64|)\approx0.0084$ と求められる。

検定の有意水準を $\alpha=0.05$ とすると（有意水準は検定を実行する前にあらかじめ決めておくべきである）、$P<\alpha$ であるので、この場合の検定は有意水準５％で有意であり（有意水準１％でも有意である）、$\mu=135$ という帰無仮説は棄却される。すなわち、有意水準５％でこのハンバーガー店のフライドポテトの平均値は135g ではないことになる。

片側対立仮説の場合は、検定統計量の値は同じく $z^*=-2.64$ であるが、P 値は $p=P(Z\leq-2.64)\approx0.0042$ となり、この場合の結論は、有意水準５％でフライドポテトの平均値は135g よりも小さいといえる、となる。

統計的検定では、帰無仮説の真偽とデータに基づく判断では、下の表のように４通りの結果となる。すなわち、２通りの誤り方を区別するのである。

		人間の判断	
		H_0 を棄却する	H_0 を棄却しない
神のみぞ知る	H_0 が真	第１種の過誤	正しい判断
	H_0 が偽	正しい判断	第２種の過誤

誤りを２通り考える考え方は重要で、たとえば、病気の診断のとき、病気であるのにそれを病気でないとして見逃す誤りと、病気でないのに病気であるという誤りとを考える必要がある。どちらかの確率を小さくしようとすると片方が大きくなるという二律背反的な性格をもつ。統計的検定では、第１種の過誤を犯す確率をコントロールするという考え方を取る。実際、第１種の過誤を犯す確率が有意水準となる。

実際の問題では、２種類の過誤のコストを考えての判断が必要となる。病気の診断で、病気であるのにそれを見逃す誤りの確率をなるべく少なくしたいのであれば、ちょっとした兆候を持つ人をことごとく病気であると診断すればよい。そうすると病気を見逃す危険性は減るものの、病気でないのに病気と言われて再検査を受けるまで心配するのも精神衛生上よいものではない。

STEP 5：Conclusion 結論　統計的に検証された結論と新たな課題

◇ 仮説検定の結果、駅前店のフライドポテト平均重量は公表値よりも少ないと判断される

高校生グループからフライドポテトの重量が公表値以下であるとの指摘を受けたハンバーガー店の店長は、それでは平均値をどのくらいに設定したら良いかを、逆に高校生グループに問いかけた。もし、平均値を135g に設定して販売すると、およそ半数の客から自分のポテトの重量が135g より少ないとクレームがきそうである。かといって、平均の重量を135g より必要以上に重めに設定すると店の利益に響いてしまう。売り場に秤を置いて重さを逐一測るのが良いかもしれないが、それでは混雑時に時間がかかってしまいサービスが低下する。重量の多少のバラツキはやむを得ないとして、どうしたら良いか、ということである。

高校生グループは、授業で学習した確率計算の実践的な応用例として次のように考えた。まず、ポテトの重量のバラツキには特に理由があるわけではなく偶然的なものであることから誤差とみなすことにした。誤差であれば、そのバラツキへの正規分布の仮定は妥当性があることから、正規分布 $N(\mu, \sigma^2)$ に基づく確率計算を行うことにした。

X をフライドポテトの重量を表す確率変数とすると、それを標準化した $Z=(X-\mu)/\sigma$ は標準正規分布 $N(0, 1)$ に従う。X が135g 未満であるとクレームがくる可能性があるので、クレーム率として $q=P(X<135)=P(Z<(135-\mu)/\sigma)$ を求めることにした。

これまでの経験から、バラツキの大きさを表す標準偏差は $\sigma=6$ 程度であったとのことであるので、$P(Z<(135-\mu)/6))$ とすると、q はμを定めれば定まる値、すなわち μ の関数であるので、横軸に μ を取り、縦軸に q を取ってグラフに表すことにした。また、$\sigma=3$ としたグラフも同時に描くことにした。図１がそのグラフである。

図１　母平均（μ）とクレーム率

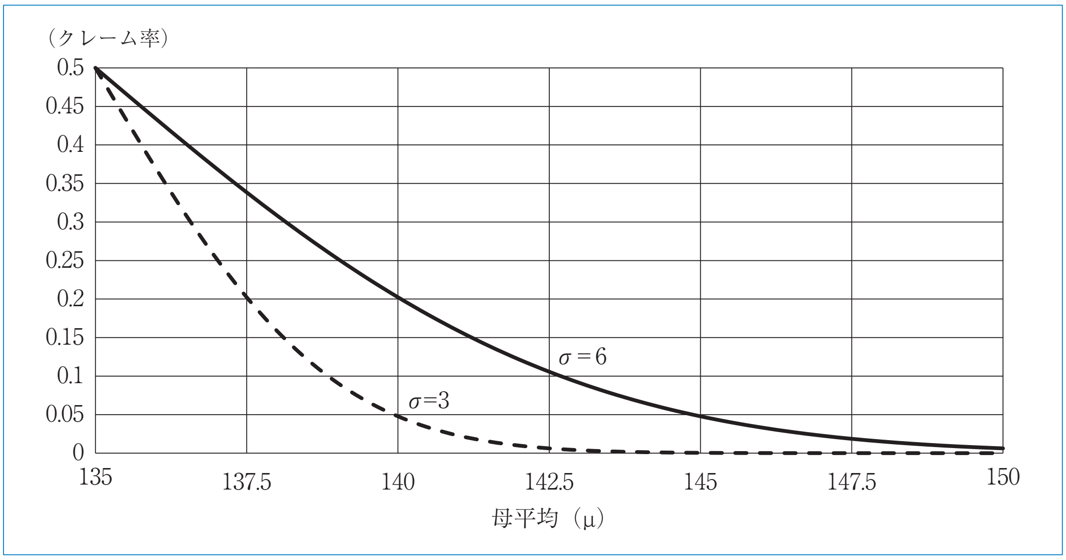


図１から、$\sigma=6$ とした場合、クレーム率を0.05程度にするためには母平均を145g 程度と、公表値よりも10g 多くしなければならないことが分かった。５g プラスの140g を平均値に設定すると、約20％のクレーム率となる。スタッフのサービスのスキルを向上させてバラツキを半分の $\sigma=3$とすると、140g を母平均に設定した場合のクレーム率は５％程度に下がることも、図１から分かる。たかが５g の差に見えるが、１日のフライドポテトの販売量はかなり多いので、かなりの節約になる。バラツキを抑える努力は報われるという一例であろう。