「私たちは、あとどれくらいで大人？」

標本データを利用した２群の差の検定

賢明女子学院中学校３年Ａ組　野中麻央

要旨

中学３年生の身長は、成人女性と比べてどれくらい差があるかを調べました。賢明女学院中学３年生の身長データ（n=88）をアンケート調査して同学年全員の協力のもと集めました。そのデータを一般的な中学３年生女子の標本データと考え、身長分布を作成しました。次に、一般成人女性の身長データを街頭で集めました。（ｎ＝190）街頭アンケートでは、カテゴリーデータとして年齢も同時に集めました。20 歳から 80 歳までのデータが集まりました。

身長は正規分布するので、この２つの集団の平均値に差があるかどうかを調べました。ここでは、「ｔ検定」という統計手法を用いました。

個人情報の取り扱いに注意しながら、学内と学外で身長データを収集しました。検定の結果として中学３年生の身長と一般成人女性との平均身長には 3.4 ㎝の差があり、両群には有意な差が存在するということがわかりました。つまり、私たちはまだ身体的に大人にはなりきっていないという結果となりました。

ただし、賢明の中学３年生の人数は少ないためか、成長のアンバランスによるものか分かりませんが、データが正規分布しているかどうかが曖昧であり、両群が同じ正規分布をしているという検定条件を満たせているかまでは分かりませんでした。

Key words：　標本調査とデータ活用・平方根・正規分布・平均値・２群における差の検定

目次

１．序章

背の高い両親の間に生まれた私は、昔から背が高く小学生の頃は男の子より背が高かったです。そして中学 1 年生の時には祖母より背が高くなっていました。しかし最近、まわりのみんなも背が高くなってきて、私より背の高い子も増えてきました。

これは、学年が上がるに連れて同年代の平均身長が高くなってきているということになります。そこで今回、中学 3 年生の身長が一般的な成人女性と比較してどれくらい差があるのかということを調べたいと思いました。

これは、定期試験の結果のように自分が全体のどの位置にいるかということを計算するのではなく、集団と集団の差を比較するということになります。よって、平均値と自分との差を計算するというようなことはできません。

研究では、アンケート調査により同学年と街中の成人女性のデータを比較し、中学 3 年生の身長はすでに大人に追いついているかどうかということを調べます。

２．文研研究

研究は標本データを集めるところから始めます。標本データとは、調べたい集団から無作為に抽出したデータのことです。対して、もともとの集団を母集団といいます。統計上、標本データ[1]が「どれくらい正しく母集団からサンプルを収集できているか」ということは非常に重要で、それを「代表性」といいます。

今回の調査における各グループの全体のデータ（母集団）は、中学校 3 年生女子のデータと一般女性です。無作為抽出[2]とは、対象の母集団から偏りなく選び出すということです。（図 1 参照）

街中におけるアンケート調査において一般女性データの代表性を担保するにはサンプル数を大きくする必要があります。（図 2、 3 参照）今回は、190 名のサンプルを集めました。一方、3 年生全体における賢明女子中学 3 年生のデータは上限があり 88 名となりました。それぞれの母集団の平均身長は一般女性が 158.0 ㎝[3]で中学 3 年生が 156.9 ㎝[4]です。母集団と標本データの差は、一般女性で 0.3 ㎝・中学 3 年生で 0.6 ㎝となりました。双方正規分布しているとみなし、母集団に対する代表性は満たしていることにしました。

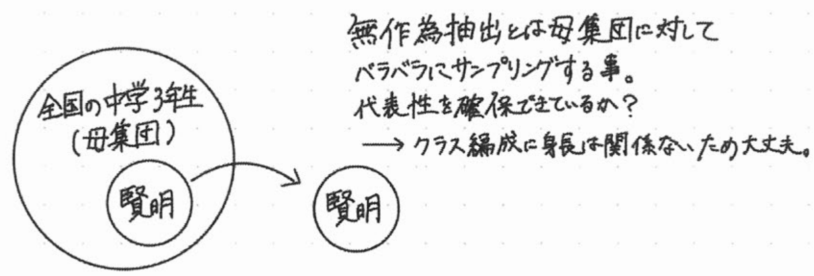


図 1　抽出データと母集団イメージ

---

[1] 標本データとは、ある母集団（全体の集まり）から無作為に選ばれた一部のデータのことです。標本データは、母集団全体を代表する性質を持つと仮定され、その性質や特徴を分析したり推測したりするために使用されます。（ChatGPT）

[2] 無作為抽出とは、全ての個体や要素が同じ確率で選ばれる方法です。これにより、抽出されたサンプルが母集団を代表し、統計的な信頼性を確保することができます。（ChatGPT）

[3] 「日本女性」の「平均身長」は？ 世界の「平均身長」や推移もあわせて紹介！ | Oggi.jp　日本女性の平均身長

[4] https://www.suku-noppo.jp/data/1804_height.html 身長データバンクより　中学 3 年生女子平均身長

図２　街頭調査（ハローズ香寺店）

図３　街頭調査（フレッシュバザール福崎店）

３．研究の方法

まず初めに、全員の身長データを集めたら、全員の身長の分散を計算します。分散とは、データのばらつきのことで、背の低い子もいれば高い子もいるということです。（図４参照）もし、全年齢のデータを集めれば分散は大きくなり、中学３年生だけとなれば小さくなるということです。

しかし、本当に知りたいのは標本間の差ではなく、想定する母集団の差なので、母集団からできるだけ無作為に抽出する必要があります。

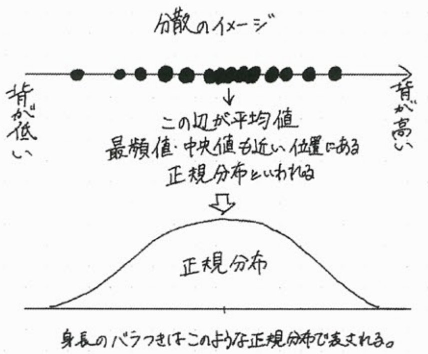


図４　分散のイメージ

この分散はどのように計算するかというと、まず、抽出したデータの平均値をとります。次に、その平均値から個々のデータ（身長）がどれくらい離れているかということを計算します。このとき、平均値より 10 ㎝背が高い人もいれば、10 ㎝低い人もいます。これの分散全員分のデータを合計し、人数分で割り戻すと標本分散（Ｓ）となります。ちなみに、母集団の分散[5]をデルタ（δ）といいます。

ただし、この分散を合計するにはひと手間加えないといけません。それは、平均値より 10 ㎝背が高い人と 10 ㎝背が低い人を合計すると０になってしまうからです。つまり、全データの平均値から＋方向の合計とマイナス方向合計の総和シグマ（Σ）[6]は０になるのです。（図５参照）そこで、＋と－を合計するには、双方を２乗し自然数にすることで上下の広がりを足すことができます。

例えば＋10 ㎝と－10 ㎝をたすには(10 ㎝）^{2}＋（－10 ㎝）^{2}とするのです。

---

[5] 分散とは、分散は、データのばらつき度合いを表します。データのばらつきが大きいと、分散も大きくなり、小さくまとまったデータだと分散は小さくなります。(https://kotodori.jp)

[6] Σ（シグマ）は、数学や統計学で使われるギリシャ文字の記号です。これは「総和」を意味し、数式の中で複数の項を足し合わせることを示します。

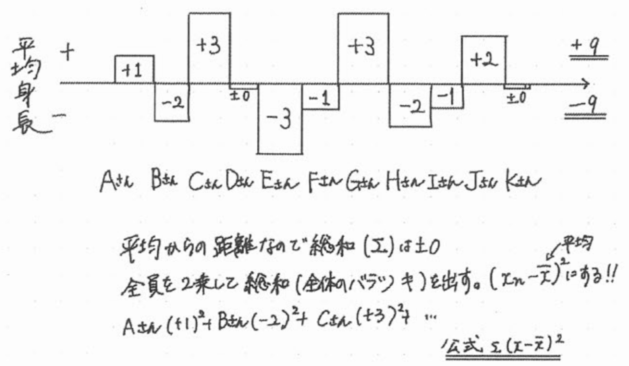


図５　２乗して合計する意味

そして、データの広がりの総和が出たら２乗をもとに戻してやる必要があり、単純に√（平方根）を付けます。これを個々の数値の平均にするためにサンプル数（n）で割り戻すとデータのばらつき具合（標準偏差[7]）になります。（図６参照）

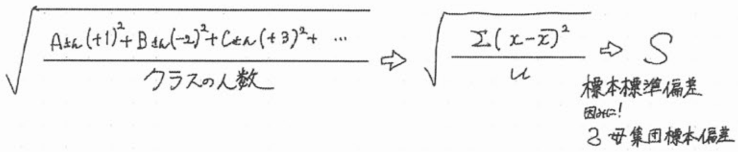


図６　平方根で元の数値に戻す

標準偏差は、統計学にとって非常に重要な概念で、我々にとても身近な数値である偏差値もこの計算のもとに成り立っています。つまり、成績などの偏差値[8]も全員の点数のばらつきを加味した数値であり、平均点からの乖離だけではわからない、もっと高度な評価方法であることがわかります。
次に、データの比較です。
２つの異なる集団のばらつきを比較するには、双方の集団に統計上有意な差があるかどうかを調べる必要があります。（図７参照）これを検定といいます。２群の距離が近いと、その違いは見えにくくなります。そのため、より多くのデータを収集しないといけません。

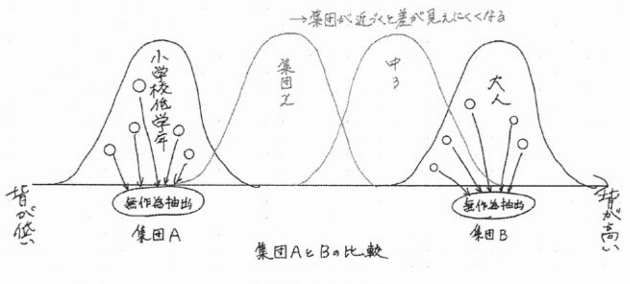


図７　二つの分散を比較するイメージ

---

[7] 標準偏差とは、データのばらつき具合を示す数値です。具体的には、データが平均値からどれだけ散らばっているかを測る指標です。例えば、クラスのテストの点数が平均的にどれくらい離れているかを知るために使います。標準偏差が大きいほど、個々のデータが平均から遠く離れていることを示し、逆に標準偏差が小さい場合はデータが平均に集まっていることを意味します。（ChatGPT）
分散の平方根(現代統計学小事典 190 1)

[8] 偏差値（Standard Score）は、ある個人やグループのデータが平均からどれだけ離れているかを示す統計的な指標です。具体的には、平均が50、標準偏差が10の場合、偏差値60は平均より10ポイント上の位置にあることを意味します。偏差値が高いほど、平均よりも高いスコアであることを示します。（ChatGPT）

小学校低学年の平均身長と成人の平均身長には大きな開きがあるため、検証しやすく、中学３年生になると見えにくくなるのです。（図８参照）実際にお母さんより背が高くなったクラスメイトも多くなり、私自身も 165 ㎝なので、データとしては２群の差を近づけてしまうサンプルとなります。

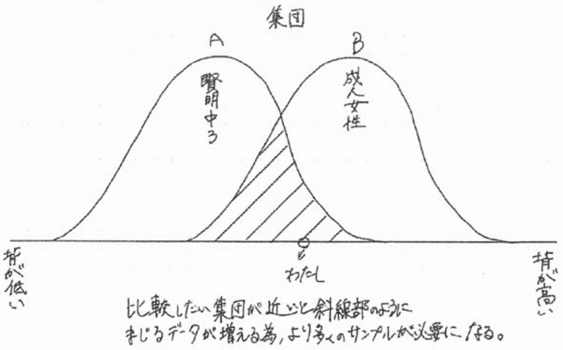


図８　２群の距離が近い場合の問題点

身長の分散データは正規分布[9]といわれる釣り鐘型の関数で表されます。今回は、この２つの正規分布が分かれているかどうかをＴ検定[10]という手法を用いて計算します。
Ｔ検定とは仮説検定ともいわれ、双方の分布に有意な差がない！とは言えない！という可能性を表す、帰無仮説[11]を利用した統計検定です。

４．差の検定

２つのグループに差があるかどうかについてですが、集計結果（図 10）を見ると二つのグループには僅かな差があるように見えました。
また、平均値・中央値・最頻値[12]を比較しても・・・→正規分布の条件を満たしていることがわかります。次に、この差が偶然なのか必然なのかを統計的に調べる方法があるそうです。これは統計的に、２群の間に有意な差があるかどうかということを調べるということで、「検定」といいます。統計的検定にはいろいろなアプローチがありますが、データの種類によってどのようなアプローチが必要かということを決めることができます。（図９）

---

[9] 正規分布とは、左右対称の釣鐘型もしくは山型の分布を指します。平均値と最頻値と中央値が一致するという特徴があります。平均値を中心にして左右対称であることも特徴です。(https://qiqumo.jp)

[10] T 検定とは、2 つの母集団の平均値を検定する方法です。同様に平均値の検定である Z 検定は母分散がわかっている場合にのみ使用できるのに対し、T 検定は母分散がわからない場合にも用いることができます。(https://qiqumo.jp)

[11] 帰無仮説（Null Hypothesis）は、統計学で重要な概念です。具体的には以下のような意味を持ちます：
- **仮説検定の出発点**: 研究者が行う統計的な検定では、何らかの主張を行う前提として、帰無仮説を立てます。帰無仮説は通常、「何も起こっていない」「差がない」「関連性がない」というような中立的な主張です。
- **検定の対象**: 帰無仮説は、通常、対立仮説と対比して検定されます。対立仮説は研究者が実際に興味を持つ仮説であり、「差がある」「関連がある」といった主張を含んでいます。(ChatGPT)

[12] 最頻値とは、データの中で出現する回数が最も多い値。(1 歩前からはじめる「統計」の読み方・考え方。[第 2 版]P84 l22)

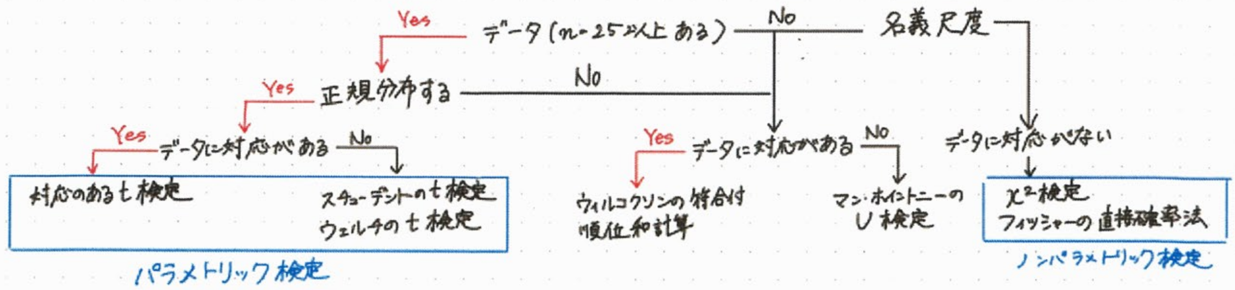


図９　２群を比べる場合の前提条件

今回の場合は、データ数が25以上で正規分布します（両群の分布も一致すると仮定しました）。そして、データには対応がありません。これは、それぞれのデータが同一人物でないからです。以上の条件から今回の差の検定は「ｔ検定」という方法を利用するということになります。今回は「スチューデントのｔ検定」を行いました。ちなみにパラメトリック[13]とは、パラメータのことで、これは分布を決めるときの重要な要素のことを言います。今回のように身長の分布は正規分布になるため事前にパラメータが既知の場合に使えるということらしいです。

５．実験データ

集めたデータはエクセルに入力し計算します。各自の身長は小数点以下を四捨五入し、１㎝刻みの身長に当てはまる人数を集計しています。

（表１．身長の標本データ）

身長（㎝）	成人	中３
140	1	0
141	0	0
142	0	0
143	0	0
144	0	1
145	1	0
146	0	1
147	1	2
148	2	6
149	3	2
150	4	4
151	2	5
152	4	2
153	9	6
154	9	6
155	4	5
156	12	9
157	14	2
158	17	7
159	5	4
160	19	7
161	10	1
162	10	3
163	15	2
164	7	4
165	13	3
166	9	1
167	2	0
168	7	4
169	0	0
170	5	0
171	2	1
172	1	0
173	0	0
174	1	0
175	1	0
n	190	88

表１が収集された身長のデータです。

賢明中３の平均身長は156.3㎝で成人女性の平均身長は159.7㎝となりました。このデータを比較してグラフにしたものが、図８です。２群の平均値の差は小さく見えますが、成人女性の分散は裾野が広く、少し右側にずれているように見えます。次に図11は、成人女性の年齢と身長の関係を表したものです。２０歳から80歳までのデータから、年齢が上がると身長は少し下がるように見えましたが、全体的に正規分布しているので、このままデータを利用することにします。ちなみに、図10のデータは、両群のサンプル数が違うため、各データをサンプル数で割り戻し平準化して各カテゴリーに分布するパーセンテージ（%）を表示しています。

---

[13] パラメトリック検定とは、「事前にデータの分布を仮定している検定」のこと。Ｔ検定が有名。ｔ検定はデータは正規分布に従っていることが前提。（いちばんやさしい、医療統計より）

また、中３の身長分布は一般女性に比べて、正規分布が乱れているように思いました。特に、148 ㎝と 168 ㎝で突出しています。原因として２つ考えられるのは、単純にサンプル数が少ないというものと、成長スパートの時期[14]です。この時期の女子は「成長のスピードやタイミングに個人差が大きく出る時期」とされているのです。

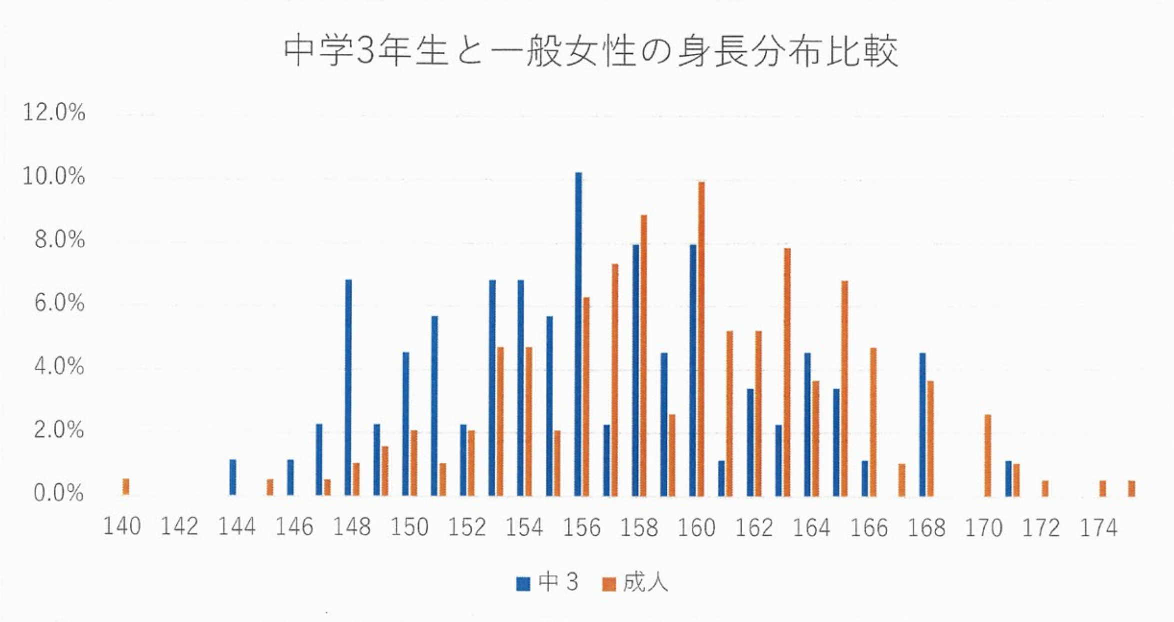


図 10　２群の身長分布

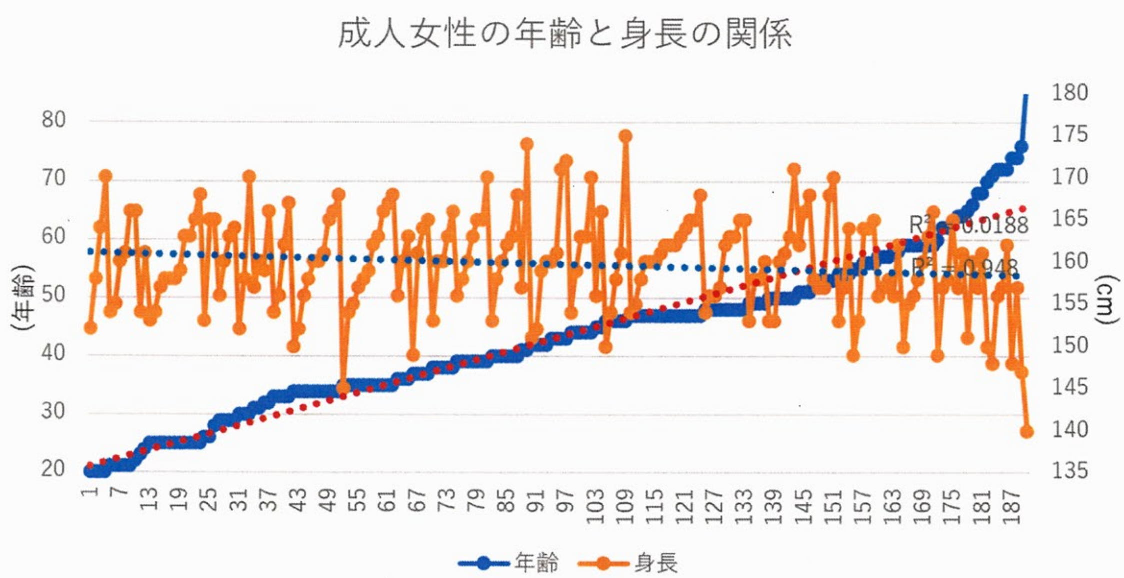


図 11　年齢と身長の関係

[14] https://koreyokatta.net/chuugakusei-jyoshi-heikin-shinchou/#i-13
『中学生女子の身長はどのくらい？』

以上のように、少し正規分布がずれていますが、そのまま図 10 の両群が統計的に有意な差があるかどうかを計算します。データのばらつき（標準偏差）の計算式は以下の通りです。

$$s = \sqrt{\Sigma (xi - \mu)^2 / n}$$

「S」とは、標本標準偏差です。身長は全員のデータをとったものではないので、標本となり、全体（母集団）となる場合は「δ(デルタ)」になります。「n」とは、標本のサイズです。今回の場合ではA組の人数やアンケート調査のサンプル数となります。ちなみに母集団の数となると「N」と表記します。「xi」とは、標本の観測値といって「i」は添え字です。今回の場合では個々の身長を意味します。「μ」とは、平均値を意味し、$\bar{x}$（エックスバー）ともいい、平均身長を意味します。「Σ」とは、すべて足すという意味です。これらを数字に当てはめると表 2のようになります。

（表２．検定の結果）

	賢明中３	一般女性
平均身長	156.3 ㎝	159.7 ㎝
2 群の平均の差	3.36 ㎝	
標本データのばらつき	5.898	5.765
ばらつきの 2 乗	34.782	33.243
n 分のΣ	0.3953	0.1749
平方根	0.755133	
自由度	（190－1）+(88-1)=276	
t 値	4.45377（分布中心からの距離）	
P 値	0.001227% (2 群に差がない場合に t 値がこの位置に来る確率)	

上記で算出された数値（表 2）の中で注目する数値が t 値[15]と p 値[16]です。

[15] t 値（t-value）は、データを比較するときに使う数字です。例えば、あるクラスのテストの平均点を調べて、全体の平均点とどれくらい違うかを知りたいときに使います。
t 値を計算するには、次の 3 つのことが必要です：
1. **標本の平均点**（例えば、クラスの平均点）
2. **母集団の平均点**（例えば、全校生徒の平均点を推測する）
3. **標本のばらつき**（データの散らばり具合）

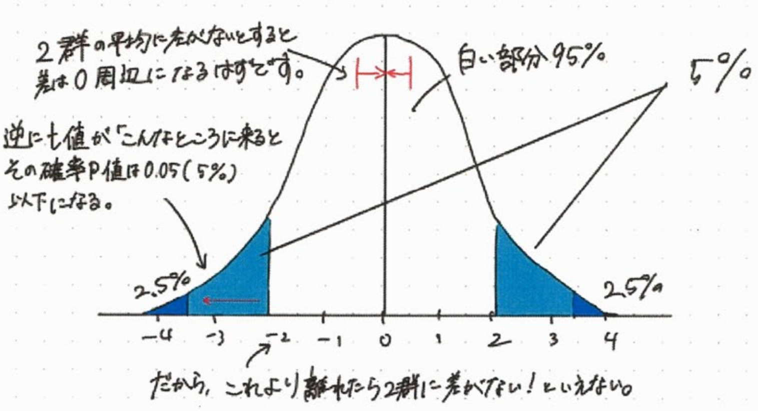
計算式は少し複雑ですが、この数字を使うと、例えば「新しい教材が効果があるかどうか」などの重要な質問に答えることができます。（ChatGPT）

[16] p 値（p-value）は、統計学で仮説検定の結果を評価するための指標です。具体的には、以下のような役割があります：
1. **仮説検定の結果の評価**: 仮説検定を行った後、得られたデータが偶然に起因するものかどうかを評価します。
2. **統計的な有意性の判断**: p 値が小さいほど、観察されたデータが帰無仮説（例えば、2 つのグループの平均値が等しい、ある処置が無効であるなど）に反すると考えることができます。(ChatGPT)

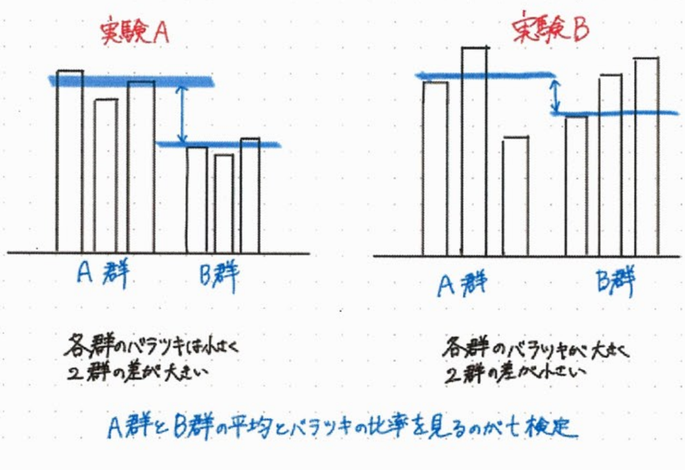
t 値が絶対値 2 以上で、p 値が「5%」未満の場合、2 群の平均には有意な差があると考えられるようです。(図 12 参照) t 検定とは 2 群の差があるかないか？という計算です。もし両群に差がないとすると、t 値は正規分（t 分布）の中心（0）周辺に来るはずで、t 値がそこから離れていく可能性は低くなります。その可能性を表したものがＰ値ということらしいです。

よって、上記の t 値は分布から 4.45 離れており、それは非常に小さい部分（確率）になります。そして、もしこの値が偶然によって出た数字だとすると、その確率は p 値であり、0.001%となります。「99.9%以上の確率で偶然とは言えない差が 2 群にはある」ということになりました。

またこの考え方は、帰無仮説[17]を棄却するといい、図 13 となります。



（図 12　棄却域のイメージ）



（図 13　帰無仮説のイメージ）

図 13 の実験 A の場合、「2 群には有意な差はない」という帰無仮説に対して、「でもこの結果が出るのは 5%以下」つまり「この 2 群の差は偶然ではない」という対立仮説を立て、帰無仮説を棄却します。よって、「2 群には有意な差がある」と考えます。一方実験 B は、帰無仮説に対して、35%くらいの頻度でこの結果が出たとすると、

[17] 帰無仮説（Null Hypothesis）は、統計学で重要な概念です。具体的には以下のような意味を持ちます：

- **仮説検定の出発点**: 研究者が行う統計的な検定では、何らかの主張を行う前提として、帰無仮説を立てます。帰無仮説は通常、「何も起こっていない」「差がない」「関連性がない」というような中立的な主張です。

- **検定の対象**: 帰無仮説は、通常、対立仮説と対比して検定されます。対立仮説は研究者が実際に興味を持つ仮説であり、「差がある」「関連がある」といった主張を含んでいます。(ChatGPT)

帰無仮説の棄却はできない。つまり、2群に見えている差は偶然であり、「2群に有意があるとは言えない」という結果になります。

6．結果と考察

今回の実験では、 t 値が中心から4以上離れており、これが偶然ならその可能性は 0.001%でした。つまり、賢明中学3年生と成人女性の平均身長には有意な差があり、私たちは、全体的にまだまだ成長できるということがわかりました。おそらく、身長分布から見えた 148 ㎝から 154 ㎝あたりの人たちが特に大きくなるんだろうということが仮説として考えられる結果となりました。

7．まとめと実験の限界

統計は、私たちの生活に密接に関係しています。頻繁に接することがあるアンケート調査もこの統計調査であり、実は非常に高度な技術がないと正しい答えが出ないということを知りました。また今回行った「t 検定」も、高校や大学で深く勉強するようです。当該研究においては、2 群の分散（正規分布）が同じであるという前提に基づいて「ｔ検定」を行いましたが、本来は2群の分散が本当に同じかどうかも調べる必要があります。それを「Ｆ検定」というらしいです。今回はその部分までできませんでした。

今回は、身近でありながら非常に奥の深い統計をネットやチャット GPT などを利用し、中学生の私でも自学しながら進めていくことが可能でした。

また、エクセルは表計算といわれ、様々な統計処理をはじめその他あらゆる計算に利用可能な万能ツールです。今から勉強しておいても早すぎないと思います。

しかし、本当にどこまで正確な分析ができたかはわかりません。ここで生まれた新たな疑問をいつか先生に教えてもらえるような環境に身を置きたいと感じました。

8．謝辞

今回の身長アンケートに協力していただきました同級生の皆さん、そして研究に協力していただきましたお母さんの知り合いの方々や、アンケート調査に協力していただいた街頭の皆さま、ありがとうございました。

また、数学自由研究のアドバイスをくださった前川さんありがとうございました。

参考文献

・神林博史（2019）『一歩前からはじめる「統計」の読み方・考え方〔第2版〕』ミネルヴァ書房
・栗原伸一・丸山敦史（2021）『統計学図鑑』㈱オーム社
・鈴木義一郎（1998）『現代統計学小事典』（株)講談社
・高橋信（2018）『マンガでわかる統計学』㈱オーム社
・安川正彬（1995）『経済学入門シリーズ　統計学入門基礎編〔第 27 版〕』日本経済新聞社