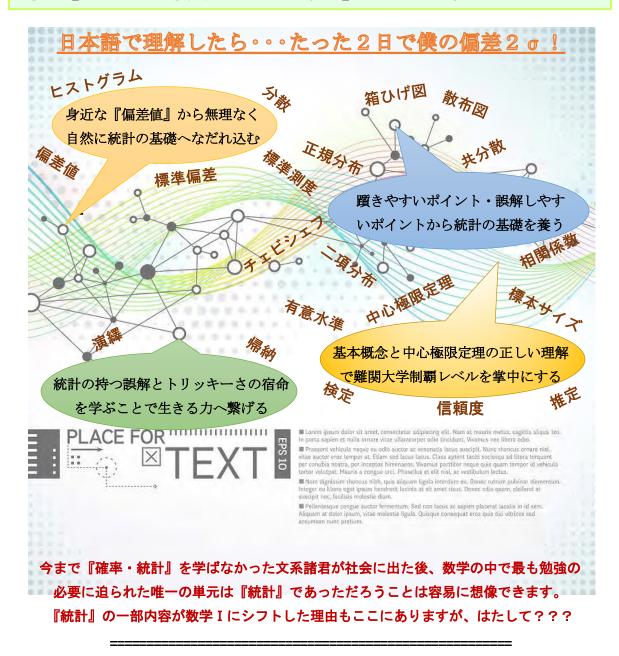
2012年 新課程より

『統計』のはしりは『数学 I:データの分析』として事実上誰もの必修科目へ!



発行日: 2013 年 10 月 25 日

著者: 二三五 八十三(フミコ ハトミ) 発行元: 「帝都大学へのビジョン」事務局

https://e-littlefield.com/teito-vision

teito@e-littlefield.com

Copyright (C) 2013 「帝都大学へのビジョン」事務局 All Rights Reserved.

『偏差値』発 賢者の統計 · 愚者の統計 著作権 ■2

-感動!ガリレオ流 帝都大学への数学 vol.7-

著作権について

「帝都大学へのビジョン」及び付属著作物全て(以下、本書と表記します)は著作権法によ り保護されています。

本書の取り扱いについては、以下の点にご注意ください。

本書の著作権は、「帝都大学へのビジョン」事務局にあります。

また、本書を購入された本人及びご家族のみが読む事を許され、購入者本人のみが内容を元 に活用出来るものとします。

(購入者本人が保護者の場合はお子様が活用できるものとします。)

許可なく無断で、本書の一部、または全部を複製及び転載することを一切禁じます。

コピー、ビデオや電子メディア、メール、メルマガなどでの複製、転載、転売 およびインタ 一ネット上で公開する事を禁止します。

またインターネット上で転売(オークションなどを含みます)することも禁止します。

万一上記の禁止行為を発見した場合は、日本国の法律に従い解決をいたします。

上記規定に違反した場合、違約件数×違約発見時の本書の販売価格と法律に従い適正な賠償 額を請求いたします。

上記規定の禁止行為を発見された方は、直ちに連絡頂ければ幸いです。

【免責事項】

本書は、著者のこれまでの指導実績と理念を元にまとめあげられたものであり、また、多く の中高生の成績アップを願ってつくられたものですが、全ての読者の利益を保証するもので はありません。

本書により、いかなる損害が発生しましても一切の責任を負いません。 すべては自己責任で行われることとします。

本書をご購入していただきました方に、このようなことをお伝えするのは心苦しいことなの ですが、法律上の必要記載事項ですので、ご理解の程お願いいたします。

本書を開封した時点で、購入者は、これに同意したものとします。

『偏差値』発 賢者の統計・愚者の統計 目次 3

-感動!ガリレオ流 帝都大学への数学 vol.7-

目 次

はじめに	5~
1 時間目:【偏差値】を求める・・・偏差という見方	
1) 【学力偏差値】を日本語でほぐす	8~
2) 【標準偏差】を日本語でほぐす	14~
3) 統計用語を日本語でほぐす	17~
4) [ケース・スタディ 1] 国語の得点分布から偏差値を求める	19~
5) [ケース・スタディ 2] 数学の得点分布から偏差値を求める	26~
6) 【偏差値】から見えること	31~
2 時間目:【箱ひげ図】を書く・・・四分位という見方	
1) 【箱ひげ図(box-and-wisker plot)】とは?	33~
2) 【箱ひげ図】を書いてみる	34~
3 時間目:【正規分布】と【二項分布】・・・統計学の大黒柱	
1) 【正規分布(normal distribution)】について	41~
2) 【正規分布】から【標準正規分布】への基準化	44~
3) ミスターZ が全てを決める【標準正規分布】	47~
4) 【正規分布表】を駆使する	49~
5) 統計に関する定義・公式のまとめ	54~
6) チェビシェフの不等式	56 ~
7) 【二項分布(binominal distribution)】について	60~
8) 【二項分布】発ーパスカルの三角形経由ー【正規分布】行き	62~
9) 大学入学後の宿題 :【二項分布】から【正規分布】導出	65~
10) 【二項分布】の公式は二項定理と ${}_{\scriptscriptstyle n}C_{\scriptscriptstyle k}$ の復習・理解とともに	66~
11) いろいろな分布	68~
4 時間目:相関というもの・・・【散布図】【共分散】【相関係数】	
1) 相関関係を【散布図(scattergram)】で表す	69~
2) 【共分散(covariance)】を日本語でほぐす	79~
3) 【相関係数(correlation coefficient)】を日本語でほぐす	77~
4) 【共分散】【相関係数】を実際に求めてみる	68~

-感動!ガリレオ流 帝都大学への数学 vol.7-

5時間目:誤解しない【中心極限定理】・・・【推定】【検定】へ 1) 【推定】【検定】とは何をするのか? 88~ 2) 中心極限定理 91~ 3) 標本数と中心極限定理 94~ 4) 具体的に推定と検定をなぞる(平均値型) 96~ 5) 比率型の推定をなぞる 103~ 6 時間目:愚者の統計・賢者の統計・・・統計の立ち位置 1) 大陸合理論とイギリス経験論 108~ 2) データ改ざん・捏造というもの 110~ 3) 統計情報のトリッキーな主役たち 113~ 4) 統計の力 117~ 5) 最後に 119~120



-感動!ガリレオ流 帝都大学への数学 vol.7-

はじめに

本講座の意図は3つあります。

第1点目は、文字通り今の君たち(保護者さんの場合はお子さん)に一番縁が深い 『偏差値』について、誤解のない理解をしておいてもらいたいという願いです。

実は、私たち日本人には『偏差値』という言葉は定着していますが、これを "deviation value" や "Z-score" や "standard score" と訳しても英語圏の一般の方にはさっぱりと通じません。

統計を使っている人や専門家なら、"standard score"という英語を理解はしますが、日本の『偏差値』の意味とは少し違って捉えられるのです。

何故なら『標準偏差』などの世界共通の公式な統計用語とは異なり、日本で『偏差値』なる概念を教育に生かすようになってから初めてこれを説明するために英語表現ができたというような経緯がある言葉だからです。

マルチリンガルの松平先生(<u>『ユダヤ式記憶術』</u>などの著作者)にお聞きしたことなのですが、面白いことに『偏差値』をそのまま中国語読みすると中国人は理解するそうで、しかも、日本人と同じ意味に捉えるそうです。

さすが、「科挙」の国ですかね。

また、韓国でも理解はされると思いますが、中国・韓国いずれの国も『偏差値』システム が一般的に使われていることはありません。

ともかくも、自分に関わる身近なデータ『偏差値』の正確な意味を知らずに、世間の使い 方から何となく理解している諸君や保護者さんが多いのではないでしょうか?

【保護者さんも必読!】という意味も込めて、「数学」と言えるほど難しいものではありませんから、なおさら、世間の間違った物言いだけで一喜一憂することのないように、その意味を正しく理解し、正しく参考にしていただきたいという願いをもって記しました。

-感動!ガリレオ流 帝都大学への数学 vol.7-

第2点目は当然ながら、数学としての『統計』の基本を押さえてもらう目的です。

そのために、世界でも日本だけに特有の概念である『偏差値』、教科書には掲載されていない『偏差値』から入っていくわけですが、このことが害になるというわけではなく、むしろ身近な分だけ『統計』へのイントロダクションとして理解への障壁を低くしてくれる優れものと言えるものです。

特に、文系に進まれる諸君の多くは(『統計』が付き物の経済学部や経営学部を目指す方は除いて)、大学に入ってしまえば数学とは無縁だと思っている方も多いのではないでしょうか?

しかし、現実社会は善くも悪くも、『統計』というものを軸にしながら動いています。

文系諸学も含め、統計データを基にした論文や、何らかの商品での安全性が統計データによって示されているカタログなどは非常に多く、世にある多くのものは統計学的根拠で、その有意性を示されますので、『統計』に関する基本概念を持ち合わせていないと理解できなかったり騙されたりすることにもなります。

何より、サラリーマンになられたら最も身近となる「販売戦略」の世界においては、必ず 『統計』を拠り所とした戦略が立てられる筈です。

そういう意味からも、高校生レベルの【統計】の基本概念ぐらいは勉強されておかなければ、 「出来る奴」と一目置かれる存在になることは難しいかもしれません。

もっと直近の可能性で言えば、君たちが卒業するために書く卒業論文でも『統計』のお世 話になる方は至って多い筈なのです。

にもかかわらず、高校数学で実質的に学習することが極めて少ないようなシステムになっていましたし、こと入試においては『確率・場合の数』は必ず出題される反面、その分 『統計』の出題頻度は非常に小さいのも事実でした。

ですから、今までの文系の諸君なら習うことすらなく卒業された人も多かったはず。 そのため、数学の基本理解度が貧しくて、経済学部では講義で『統計』の前提となる基礎 数学を教えるのに困っておられる先生も多いと聞きます。

-感動!ガリレオ流 帝都大学への数学 vol.7-

ところが、2012年より、『統計』のさわり部分が、やっと「データの分析」として数学 I の履修単元となりました。

と言うことは、実質上、より多くの生徒の必修単元となったということですね。

僕自身は『統計』が好きなわけでも何でもありませんけれども、最も縁がある可能性の高い単元『統計』のさわりが必修化されたことは良いことではないかと考えています。

但し、実質的に数学ⅡB「確率分布と統計的な推測」を続いて学ばなければ、学んだ内容が 完全に浮いてしまい、結局何もしなかったことと同じ結果になる恐れは十分あります。 と言うか確実にそうなることは必定です。

ですから、せっかく「データの分析」を履修した諸君が、たとえ数学IIB「確率分布と統計 的な推測」を履修しなかったとしても、本講座で記した一連の流れをなぞっていただくこ とで、少なくとも『統計』の基礎ブロックが完成していることを意図しています。

最後の3点目は『統計』の立ち位置をしっかりと認識してほしいという願いです。

曖昧な説明で誤解されやすい<mark>『中心極限定理』</mark>及びこれに依拠する<mark>『推定・検定』を省察</mark> しながら、『統計』というものに内包されるトリッキーな側面を認識しておいてもらうこ とが目的です。

さらに、『統計』というものの立ち位置を世の中の思考形式との対応図式で整理しておくことで、そのトリッキーな面への注意を怠ることによって安易に統計学を信頼してしまうリスクへの歯止めとなるのではないかと考えるわけです。

2012年の新課程導入に際しては、東京大学と京都大学が統計分野の入試への出題は無いことを公表したというニュースもありましたが、ある意味、安易や俗に流れやすい『統計』 及びその流れを作りたいかのような新課程編成への反抗が垣間見れます。

※だからと言って、全く勉強しないという態度を取るようでは合格できませんよ! そして、どうせ大学では学ばねば通過できないですから、後で苦労するだけです。 優先度は下げても必ず学ぶべきは基本をしっかり学んでおいてください。

1時間目:【偏差値】を求める・・・偏差という見方

1) 【学力偏差値】を日本語でほぐす

本講座のスタートは、君たちが受験の際に参考にする【学力偏差値】というものを、日本語で表現した計算式で説明するところから始めます。

尚、以降は、【学力偏差値】を単に【偏差値(Z-score)】と記述することとします。

君は、ある模試を受けました。

しばらくすると、採点結果として各教科の素点や平均点や順位とともに【偏差値】を受け取ります。

その【偏差値】とは、

君の偏差値 = <mark>君の得点 – 平均点</mark> ×10 + 50 ・・・(1)

によって計算されている値なのです。

各教科での【偏差値】を算出する場合は、君の得点は素点そのもので、平均点も普通に考える平均点そのもです。

5 教科総合での【偏差値】を算出する場合は、君の得点は 5 教科の素点の総計を考え、当然、 平均点は各人の 5 教科の総計の平均点で考えればいいわけです。

意外に簡単な式ですよね。

記号で書かれた式よりもはるかに身近に感じ取れるのではないでしょうか?

ここまでに何度も繰り返し言ってきたことですが、ここでも、

『どんな複雑に見える式でも、<u>日本語で理解しようとするクセ</u>をつけることが大切』 だということを再認識してください。

さて、(1)の計算式では、「君の得点」は君だけのデータですが、「平均点」と「標準偏差」は、 君も含めた模試受験者全員のデータから割り出されるデータだということを最初に区別し ておいてほしいのです。 さらに次のステップとして、この式(1)を見たときに、

君の得点はすでに決まっているとして、

- i) 平均点が低いほど僕の【偏差値】は高くなるなぁ!
- ii)【標準偏差】の値が小さいほど僕の【偏差値】は高くなるなぁ!

という具合なことが思い浮かべば nice ですよ。

こういう言い換えができることこそが数学という学問なんですよ。

さて、【偏差値】を決める(1)の式を理解する上で、君が知らないから理解できないのは現段 階では【標準偏差】という言葉だけだと思いますが如何ですか?

実は、高校数学の『確率・統計』では【偏差値】という概念は残念ながら出てきません。 教科書に登場するのは、【標準偏差】という概念なんです。

【標準偏差】こそが、『統計』を学習する上で世界共通の最も重要で且つ基本要素なんです。

ただ、教科書には出てこない【偏差値】ですが、入試問題では過去には何度か出題されてい ますから、発展的な知識として知っていて当然と考えられているわけです。 自分の最も身近な値ですから、ここからしっかり学習しておいてくださいね。

では、第1ステップとして【偏差値】のイメージを掴んでいただきましょう。

【標準偏差】という概念をまだ知らないのですが、このステップでは、その【標準偏差】を おぼろげにイメージできればしめたものという目的で進めますね。 まずは、気楽に軽く通過してください。

- I) さて、例えば、君の偏差値が70だったとします。
 - (一般的には難関医学部を狙える高い偏差値ですね!)

これを(1)式に当てはめてみましょう。

君の得点 – 平均点 = 2 であれば、偏差値がちょうど 70 になる計算になりますね。 標準偏差

そうです、

『偏差値 70』とは、『【標準偏差】の2倍分を平均点よりも上回った点数 get!』 ということだと言い換えられますね。

Ⅱ) もし、君がちょうど平均点を取っていたとすると、

【標準偏差】がいくらであろうと

君の得点 - 平均点 = 0 となりますから、【標準偏差】

には関係なく君の偏差値は50だったはずです。

そうです、

『偏差値 50』とは、『平均点と同じ点数 get!』

ということだと言い換えられますね。

Ⅲ)さらに、もし、君が平均点以下の点数を取っていたとすると、

 $rac{ 書の得点 - 平均点 }{ < 0 }$ になりますから、偏差値が 50 以下だった筈ですね。

そうです、

『偏差値 50 未満』とは、『平均点より低い点数 qet!』

ということだと言い換えられますね。

Ⅳ)もう一丁、この式から考え得ることがありますよ。

それは、もし、君の点数が【標準偏差】と呼ばれる値と同じ点数だけ平均点を上回って いたとすればどうかということです。

この場合、

君の得点 – 平均点=1となりますから、君の偏差値は 60 だった筈です。

そうです、

『偏差値 60』とは、『【標準偏差】と同じ分を平均点よりも上回った点数 get!』 ということだと言い換えられるのです。

【偏差値】の式(1)を、このように具体的な数値を当てはめて、且つ、日本語で理解すると、 【標準偏差】の意味合いもおぼろげに見えてきたのではないでしょうか?

以上の第1ステップを考察してみると、

- ① 【偏差値】とは、「君の得点が平均点からどの程度±に隔たっているのか?」を示す尺度。
- ② その「どの程度±に隔たっているのか?」を表すために、【標準偏差】というものを尺度 としているみたい!
- ③ ということは、【標準偏差】とは、「標準的な受験者の得点が平均点からどの程度隔たっ ているのかを表す数字なのではないか? と推測することができそうですね。

ただ、「何をもって標準的な受験者なのか?」は現段階では分からないですけれど・・・。

けれども、

標準的な受験生の隔たり具合と同等の点数を平均点より上回れば、偏差値は 60 なんだ というイメージは掴むことができましたね!

また、逆に、

標準的な受験生の隔たり具合と同等の点数を平均点より下回れば、偏差値は 40 なんだ というイメージも掴むこともできましたね!

どうやら、【標準偏差】は平均値を基準とした±両側への『得点の隔たり具合の標準値』を意 味していると考えてよさそうですよね?

『隔たり具合』 という言葉は少し小学生っぽいですから、ここからは『散らばり具合』とい う言葉を使っていきましょうか!

ともかくも、

受験生全体の平均点と得点の散らばり具合(分布の仕方)で君の偏差値が決まってくる ということが言えそうです。

普段、各個人の学習における達成度を判断するには、通常、『平均点』と比較するという尺度 で考えていますよね。

平均点よりどの程度高いか低いか、即ち、(君の得点 - 平均点)で、成績の大雑把な感覚的 判断をしていると思いますが、ただ、全体の中でのポジションは分かり兼ねましたね。

仮に80点で答案が返ってくれば、通常は「よくできた!」と判断しますが、平均点が90点 だったのであれば、実際には決して「よくできた!」とは言えない成績だと言えますから、 平均点からの高低を目安にする方が正確な判断ができます。

-感動!ガリレオ流 帝都大学への数学 vol.7-

でも、これだけでは相対的なポジションが分かりづらいことに変わりはありません。

そこで、その「みんなの中でどの程度成績が良いのか悪いのか?」、即ち、全体の中での相対 的な位置を数値化したもので表そうという試みが【偏差値】と言えそうです。

具体的には、【自分の偏差】を【標準的な偏差】で割った数値を尺度として使おうということ にしたと見て取れます。

ところで、大学受験・高校受験に向けてまっしぐらの諸君にとっては、この【偏差値】なる 言葉はほとんど日常用語化していますね。

「君が目指す大学(高校)に合格するためにどれほどの偏差値が必要か?」は日常生活の最 大の関心事ではないでしょうか?

君が参考にしている各大学(高校)の【偏差値】は、大手の予備校や大手の模試実施会社か ら発表されています。

これは、各予備校 etc が実施する全国統一模試におけるデータから算出された偏差値です。

ですから、各予備校が発表するデータには少々の差がありますし、たとえ同じ予備校であっ ても、実施される模試の度に合格基準となる【偏差値】は変動するのです。

ただ、各予備校によって模試を受験する層の分布がそれほど変わることがなければ、予備校 間の差はそれほど大きくは出ないでしょうし、まして、同じ予備校の第1回模試、第2回模 試での差異もそれほど出ないでしょう。

即ち、大学への合格可能性の判定は、昨年度までの模試あるいは前回までの模試のデータと 昨年度までの合格者の偏差値、本年度志望者の偏差値を対照して統計的に割り出されていま すが、当然、模試受験生が前年度と同じような層で形成されている母集団であることが前提 となります。

従って、正確に且つ厳密に言えば、『○○予備校の全国統一模試での偏差値○○が、○○大学合格 のために必要な基準偏差値』となるのです。

ただ、大手の予備校になると、その値がそれほど大きくは変わらないというところで、

「東大文Ⅰの偏差値は約○○」などと一般的には言われているわけです。

それでも、駿台・河合塾・代々木ゼミナールが発表する合格基準となる【偏差値】は 0~5 程 度の相違は見られるのです。

『偏差値』発 賢者の統計・愚者の統計 1時間 13

-感動!ガリレオ流 帝都大学への数学 vol.7-

一般的に言えば、母集団から抽出したデータの一部から母集団のデータ分析を行います。 この取り出されたデータを『標本=サンプル』と呼びます。

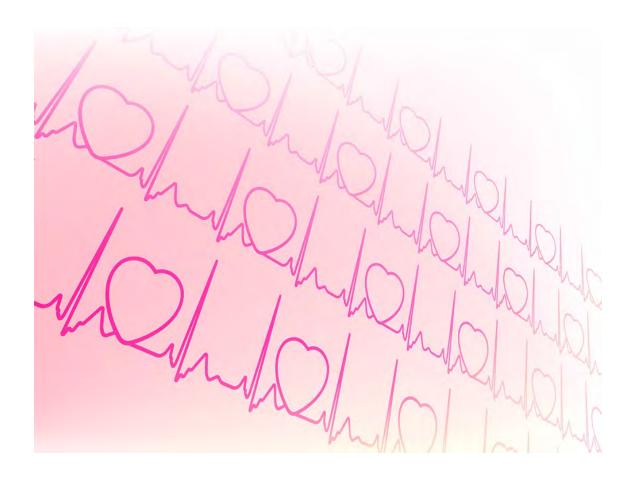
母集団の構成が違うデータでは、偏差値は大きく変わる。

このことを先ず理解しておかなければ、大きな誤解をすることになります。

学校内での偏差値と大手予備校の偏差値を比較しようとする諸君は居られないでしょう。 それと同じように、中学受験模試の偏差値と高校受験模試の偏差値、大学受験模試の偏差値 は全く同じ土俵で比較できないのです。

中学受験模試は有名中学受験生層が多く受験しますが、高校受験模試は広く一般駅な受験生 層になりますからね・・・。

高校受験のための模試でいつも偏差値が 70 以上あったから、これなら楽勝で東大にも合格 できそうなどと考える諸君もいないでしょうけれども念のために言っておきますね。



2) 【標準偏差】を日本語でほぐす

もう一度、(1)式を眺めて、もう少し詳しく解説をした後、具体的に【偏差値】を求める手順 を追うことにしましょうか。

<mark>君の偏差値 = <mark>君の得点 - 平均点</mark> ×10 + 50 ・・・(1)</mark>

先ほど、大雑把な感覚的判断はできると書いた(君の得点 - 平均点)のことを、君の【偏差 (deviation)】と呼びます。

意味合いを考えても実にフィットした言葉ですから、すぐに頭に入りますね。

【標準偏差(standard deviation)】は後述しますが、先に結論だけ述べておきますと、君を も含めた対象となる全ての受験生の【偏差】の二乗平均平方根です。

よって、 $\frac{ 君 \mathcal{O}$ 得点 - 平均点 は、君が獲得した君だけに固有の【偏差】を、全員の【偏差】

から演算された【標準偏差】で割った値ということです。

即ち、「君の得点は平均点から標準偏差の何倍隔たっているか?」を表す数値です。 「君の成績の偏りの標準的な偏りに対する感受性」といった表現もいいですね。 このように、日本語で理解しておくことがとても大切です。

- この数値は、<mark>【標準測度(standardized measure)】</mark>と呼ばれ、
 - ・君の得点が平均点の場合、【標準測度】=0となります。 前項Ⅱ)の偏差値 50 の場合です。
 - ・君の偏差(君の得点ー平均点)が標準偏差と等しい場合、【標準測度】=1 となります。 前項Ⅳ)の偏差値 60 の場合です。

ところで、君の成績は【標準測度】が0だとか1だとか言われても、なんだか淋しいですね。 そこで、【標準測度】を10倍して50を加えて【偏差値】にしたのですね。

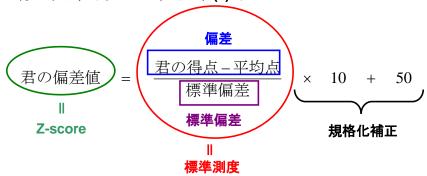
それは、

- 平均点であれば偏差値が50となるようにする
- ・オーダー的に妥当な数値として標準偏差を 10 として換算する

ために補正し規格化するための操作だったのです。(日本の【学力偏差値】の統一規格)

※注意:この操作は国や目的によって補正値が変わります。

分かりやすくまとめておくと、(1)式は



ということになります。

あらゆる『統計』の基本は、この【標準測度】なんですね。

いよいよ、【標準偏差(standard deviation)】を理解しておくべき時が来たようです。 先ほどは、対象となる全ての受験生の【偏差(deviation)】の二乗平均平方根だと結論だけ 示しておきました。

【二乗平均平方根】は、電力や電気、振動をはじめとする信号の強度として物理系の計測では「Root Mean Square を取る、あるいは、RMS を取る」という言葉で日常的によく使われ、その値は『実効値』と呼ばれます。

僕たちは、通常「Root Mean Square」を使っていましたが、日本語で言う場合は、平方根をとる前の「二乗平均値」という言葉を単に使っていました。 厳密に言えば「二乗平均平方根」ですが、単に「二乗平均値」と呼んでいたんですね。

この辺り、日常的には暗黙の了解の下に厳密には正しくない言葉が使われることも多いですけれど(テキストや参考書は大丈夫です)、当人たちは常識的に区別しているのです。 そんな場面に初めて遭遇しても、よく意味を考えて誤解がないようにしてください。

では、【偏差値】も含めて社会学分野の統計で使われる場合の多くは物理信号ではなく離散 的な数値でのお話が多くなりますので、その土俵で分かりやすく説明していきましょう。

受験生が N 人とし、それぞれの受験生の得点を $\left\{x_1,x_{2,}x_3,\cdots,x_{N-1},x_N\right\}$ とすると、

単なる【平均値】(=平均点)は、 $\frac{x_1+x_2+x_3+\cdots+x_{N-1}+x_N}{N}$ で計算できますね。

一方、【二乗平均平方根】とは、 $\sqrt{\frac{x_1^2+x_2^2+x_3^2+\cdots+x_{N-1}^2+x_N^2}{N}}$ のことになります。

その名前の通り、各人の得点を2乗した値の平均値(二乗平均値)のルートを取った値です。

以下、【平均値】(=平均点) は通常 m という記号で表されますのでこれに従います。 (英語の mean の頭文字: あるいは μ が使われることも多い)

いよいよ、本題です。

【標準偏差(standard deviation)】とは【偏差(deviation)】の二乗平均平方根でした。

即ち、
$$\sqrt{\frac{(x_1-m)^2+(x_2-m)^2+(x_3-m)^2+\cdots+(x_{N-1}-m)^2+(x_N-m)^2}{N}}$$
で定義されます。

この式のルート内の式、

即ち、【偏差】の二乗平均値(2 乗の相加平均)のことを【分散(variance) 】と呼びます。 よって、【標準偏差】は、【分散】の平方根であるということになりますね。

まとめておきますと、

【平均値 (mean)】・・・各データの相加平均

【偏差 (deviation)】・・・各データと【平均値】との差

【分散(variance) 】・・・各データの【偏差】の2乗の相加平均

【標準偏差(standard deviation)】・・・【分散】の平方根

以上のことが理解できれば、【偏差値】が理解できますし、高校数学範囲での『確率・統計』 もほとんど理解できたようなものです。

ここまでを読んで、『相加・相乗の関係』を見直してみようかという気になった人は、もう大 したものです。

この機会に是非復習しておいてください。

尚、【標準偏差】の記号としては、通常 σ が使われます。

文中では、SD (standard deviation の略) がよく使われています。

3) 統計用語を日本語でほぐす

これから具体的なケース・スタディを通してイメージを掴んでいただきますが、その前に、 『統計』で使う用語をしっかりと把握していただくために、ケース・スタディで使うサンプ ルデータのお話から始めましょう。

(ここまでもそうですが、英単語の語彙力や英語の勉強にも繋げてください。)

15,000 人が受験した模試の成績を観測して分かりやすいデータにすることを考えます。 【偏差値】の概念を伝える目的での便宜上、この模試は 5 点刻みの採点としました。 (現実では、60 点~64 点に何名、65 点~69 点に何名といった統計になります。)

15,000 人のデータを分析するのは大変ですから、例えば、無作為に 100 人の成績を抽出し、 今回の模試の統計とすることにしましょう。

15,000 人の成績を、この統計の【母集団 (population)】と呼びます。 抽出した 100 人の成績を、この統計の【標本 (sample)】と呼びます。

今回は、無作為に 100 人の成績を抽出し、国語と数学の 2 つの教科に分けて、それぞれの教科ごとに集計したデータを題材にしてみましょう。

この場合、受験生全体という『【母集団】から【国語の標本】と【数学の標本】を抽出する』 という統計データになります。

サンプル 100 人の成績を、【階級 (class)】に分けます。

模試の場合、獲得した得点が【階級】となります。

今回は 5 点刻みの採点ですから、0 点と 100 点を含めて 21 の階級ができます。 各得点に、これを獲得した受験生の人数を集計します。

各得点の人数は、この統計における【度数 (frequency)】と呼ばれ、階級と度数の対応を【度数分布 (frequency distribution)】と呼びます。

統計を比較する場合、サンプル数が違うと【度数分布(frequency distribution)】では同じ 土俵で比べることができません。

そこで、サンプル全体に対する度数の割合で分布を整理しておくと便利ですね。

-感動!ガリレオ流 帝都大学への数学 vol.7-

これを、この統計における【相対度数 (relative frequency)】と呼び、階級と相対度数の対 応を【相対度数分布(relative frequency distribution)】と呼びます。

【相対度数分布 (relative frequency distribution)】は、数学的に【確率分布 (probability distribution)】と同様の扱いで分析していけます。

何故なら、相対度数の総和も確率の総和も1となるからです。

現実に起こった事象であっても全体に対する比率で考えることによって、確率と同じ考え方 ができるということになります。

現実に起こった事象のデータや実験やアンケートで得られたデータは実際のデータであり、 可能性ではありませんけれど、考え方は確率の考え方を当てはめますから、実際に起こった ことの統計データなのに、【確率変数 (probability variable)】という言葉が使われます。

正確に言えば、【相対度数変数】と呼ぶべきものではないかという点で、大いに違和感を感じ られるかと思いますが、数学的に【確率理論】の考え方を適用しているのだと考えておけば いいですね。

では、聞くより慣れろということで、実際の手順を追っていきましょう。

ここから、具体的なケース・スタディで【偏差値】を求めるステップを追いながら、『統計』 というものの基本イメージを掴んでいただきます。



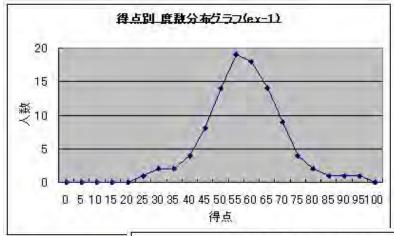
-感動!ガリレオ流 帝都大学への数学 vol.7-

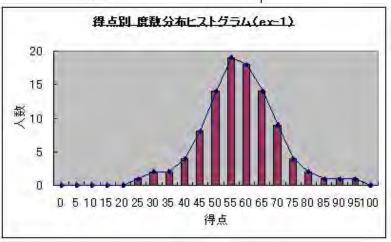
4) [ケース・スタディ 1] 国語の得点分布から偏差値を求める

i) サンプル 100 人の成績から得点と人数を集計し、表にします。···度数分布表

得点	人数	得点	人数
【階級】	【度数】	【階級】	【度数】
0	0	55	19
5	0	60	18
10	0	65	14
15	0	70	9
20	0	75	4
25	1	80	2
30	2	85	1
35	2	90	
40	4	95	
45	8	100	
50	14	合計	100

ii)度数分布表を下記のようにグラフ化します。・・・度数分布グラフ or ヒストグラム





Copyright (C) 2013 「帝都大学へのビジョン」事務局 All Rights Reserved. https://e-littlefield.com/teito-vision ※当レポートのテキスト・画像等すべての転載転用、商用販売を固く禁じます。

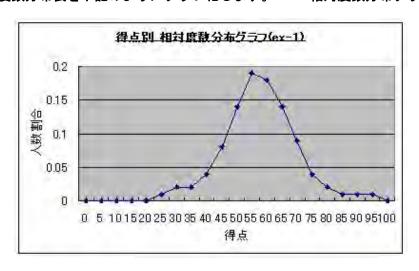
『偏差値』発 賢者の統計・愚者の統計 1時間 20

-感動!ガリレオ流 帝都大学への数学 vol.7-

iii) 度数を全体に対する割合の表示に変換します。· · · 相対度数分布表

	人数割合		人数割合
得点	【相対度	得点	【相対度
【階級】	数】	【階級】	数】
0	0	55	0.19
5	0	60	0.18
10	0	65	0.14
15	0	70	0.09
20	0	75	0.04
25	0.01	80	0.02
30	0.02	85	0.01
35	0.02	90	0.01
40	0.04	95	0.01
45	0.08	100	0
50	0.14	合計	1

ⅳ〉相対度数分布表を下記のようにグラフ化します。・・・相対度数分布グラフ



ここまでは、統計を視覚的に捉えるための作業と考えてください。

ここからは、iii)の相対度数分布表から、確率理論に基づいて計算を進めます。

そのために、ここで、

- ·【階級】 \rightarrow 【確率変数 X】
- -【相対度数】 \rightarrow 【確率 P】 (probability の頭文字) と名称を変え、且つ、記号を付与しておくことにしますよ。

-感動!ガリレオ流 帝都大学への数学 vol.7-

即ち、

得点		得点	
【確率変数	人数割合	【確率変数	人数割合
X	【確率 P 】	X]	【確率 P 】
0	0	55	0.19
5	0	60	0.18
10	0	65	0.14
15	0	70	0.09
20	0	75	0.04
25	0.01	80	
30	0.02	85	0.01
35	0.02	90	0.01
40	0.04	95	0.01
45	0.08	100	0
50	0.14	合計	1

v)【期待値(expectation value)】を計算します。

【期待値】の記号には通常Eが付与されます。(expectation の頭文字)【期待値】と【平均値】の意味は同じです。

可能性を論じる場合には【期待値】、現実に起こったデータを論じる場合には【平均 値】を使うのが自然とぐらいに考えておいてください。

ただ、後者の場合も確率理論の考え方を使いますので、現実に起こったデータなのに 言葉として【期待値】が使われる場合も多いのです。

- ※i) の度数表を用いて、平均値(=平均点)を出すことと同じことになります。 【階級値】の【平均値】(現実事象)であり、式は、∑(【階級値】×【度数】)÷全サン プル数ですね。
- ※iii)の相対度数表を用いた場合、∑(【確率変数】×【確率】)で算出できます。 即ち、【確率変数】の【期待値】(確率事象)の意味になります。
 - i)のように全サンプル数で割る必要がありません。 (相対度数自体が、すでに度数を全サンプル数で割った値ですから・・・)。

【確率変数】の【期待値】(可能性) =【階級値(相対度数)】の【平均値】(現実)で あり、後者の場合も前者の用語が使われると理解しておけばいいですね。

本来は、現実事象を論じる場合と確率事象を論じる場合は、下表のように区別しておくと より厳密と思うのですが、実際には【度数変数】という言葉が使われることは極めて稀で、 現実事象を論じる場合も確率事象としての言葉(黄色)が通常使われます。

現実事象		確率事象
度数変数	\Leftrightarrow	確率変数
相対度数	\$	確率
平均值	⇔	期待値
偏差		
標準偏差		
分散		

但し、【平均値】と【期待値】に関しては双方が使われます。

■【平均値】と【期待値】の使い分けのイメージング

例えば、サイコロを 600 回振って、各目が 100 回ずつ出たとします。

出た目の【平均値】は、

$$m = \frac{1 \times 100 + 2 \times 100 + 3 \times 100 + 4 \times 100 + 5 \times 100 + 6 \times 100}{600} = 3.5$$

一方、サイコロを振った場合には、各目の出る確率は等しく $\frac{1}{\kappa}$ と考えますと、

出るであろう目の【期待値】は、

$$E = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

【偏差値】を扱う本ケースでは【平均値】を使う方が適していますが、ここでは、確率理論 として計算のステップを踏んでいきますので【期待値】を使うことにします。

『偏差値』発 賢者の統計・愚者の統計

-感動!ガリレオ流 帝都大学への数学 vol.7-

では、国語の相対度数分布表から【期待値(平均値)】を求めましょう。

【確 率変数 X】	【確率P】	$X \bullet P(X)$
0	0	0
5	0	0
10	0	0
15	0	0
20	0	0
25	0.01	0.25
30	0.02	0.6
35	0.02	0.7
40	0.04	1.6
45	0.08	3.6
50	0.14	7
55	0.19	10.45
60	0.18	10.8
65	0.14	9.1
70	0.09	6.3
75	0.04	3
80	0.02	1.6
85	0.01	0.85
90	0.01	0.9
95	0.01	0.95
100	0	0
合計	1	$E(X) = \sum \{X \bullet P(X)\}$

集計結果より、【期待値(平均値)】E(X) = 57.7

【期待值(平均值)

VI)【分散(variance) 】と【標準偏差(standard deviation)】を計算します。

【分散】とは、【確率変数】と【期待値】との隔たり、即ち、【偏差】の二乗平均値で

そして、【標準偏差】とは【分散】の平方根でした。

今後、

- -【分散】は、V (variance の頭文字)という記号が充てられ
- ・【標準偏差】は、 σ という記号が充てられます。

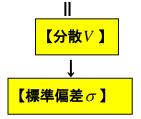
よって、【分散】と【標準偏差】の関係を式で表しておけば、

-感動!ガリレオ流 帝都大学への数学 vol.7-

 $\sigma = \sqrt{V}$ あるいは $V = \sigma^2$ ということになります。 これは、しっかりと頭に入れておいてくださいね。

【確率変数	【確率₽】	$(X-E(X))^2$	$(X - E(X))^2 \bullet P(X)$
X 1 0	0	3329.29	0
5	0	2777.29	0
10	0	2275.29	0
15	0	1823.29	0
20	0	1421.29	0
25	0.01	1069.29	10.6929
30	0.02	767.29	15.3458
35	0.02	515.29	10.3058
40	0.04	313.29	12.5316
45	0.08	161.29	12.9032
50	0.14	59.29	8.3006
55	0.19	7.29	1.3851
60	0.18	5.29	0.9522
65	0.14	53.29	7.4606
70	0.09	151.29	13.6161
75	0.04	299.29	11.9716
80	0.02	497.29	9.9458
85	0.01	745.29	7.4529
90	0.01	1043.29	10.4329
95	0.01	1391.29	13.9129
100	0	1789.29	0
合計	1	合計	$V(X) = \sum \{(X - E(X))^2 \bullet P(X)\} =$

集計結果より、【分散】V(X) = 147.21



よって、【標準偏差】 $\sigma = \sqrt{V(X)} = \sqrt{147.21} = 12.133$

さぁ、【期待値(平均値)】Eと【標準偏差】 σ が出ましたから、これで準備万端です。

『偏差値』発 賢者の統計・愚者の統計 1時間 25

-感動!ガリレオ流 帝都大学への数学 vol.7-

VII) 最後に、【偏差値】を求めましょう。

X の偏差値をZ(X)で表すことにします。

実際の偏差値は整数で出ますけれども、ここでは小数点3桁まで表現しました。

【確率変数 X】	【期待値 E 】	【標準偏差	【偏差値 Z 】 $Z(X) = (X - E(X))/\sigma(X) \times 10 + 50$
0			2.444
5			6.565
10			10.686
15			14.807
20			18.928
25			23.049
30			27.170
35			31.291
40			35.412
45			39.533
50	57.7	12.133	43.654
55			47.775
60			51.896
65			56.017
70			60.138
75			64.259
80			68.380
85			72.501
90			76.622
95			80.743
100			84.864

これで、国語の模試テスト得点に対応する【偏差値】が出揃いました。

得点 70 点の生徒 ⇒ 偏差値≒60 得点80点の生徒 ⇒ 偏差値≒68

次に、数学の模試テスト結果に対しても同様に行います。

『偏差値』発 賢者の統計・愚者の統計 1時間 26

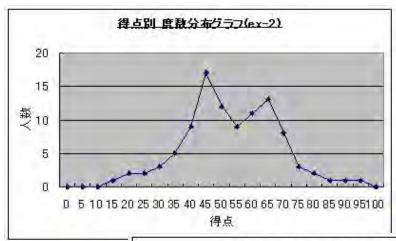
-感動!ガリレオ流 帝都大学への数学 vol.7-

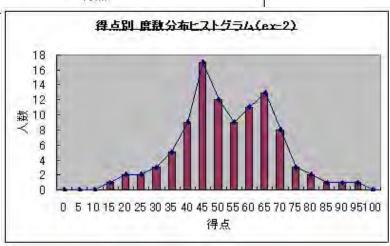
5) [ケース・スタディ 2] 数学の得点分布から偏差値を求める

i) サンプル 100 人の成績から得点と人数を集計し表にします。・・・度数分布表

得点	人数	得点	人数
【階級】	【度数】	【階級】	【度数】
0	0	55	9
5	0	60	11
10	0	65	13
15	1	70	8
20	2	75	3
25	2	80	2
30	3	85	1
35	5	90	1
40	9	95	1
45	17	100	0
50	12	合計	100

ii) 度数分布表を下記のようにグラフ化します。・・・度数分布グラフ or ヒストグラム



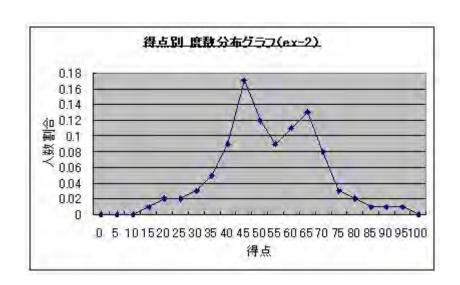


-感動!ガリレオ流 帝都大学への数学 vol.7-

iii) 度数を全体に対する割合の表示に変換します。· · · 相対度数分布表

	人数		人数
得点	【相対度	得点	【相対度
【階級】	数】	【階級】	数】
0	0	55	0.09
5	0	60	0.11
10	0	65	0.13
15	0.01	70	0.08
20	0.02	75	0.03
25	0.02	80	0.02
30	0.03	85	0.01
35	0.05	90	0.01
40	0.09	95	0.01
45	0.17	100	0
50	0.12	合計	1

ⅳ)相対度数分布表を下記のようにグラフ化します。・・・相対度数分布グラフ



『偏差値』発 賢者の統計 - 愚者の統計 1時間目 28

-感動!ガリレオ流 帝都大学への数学 vol.7-

v)【期待値(expectation value)】を計算します。

【確 率変数 X】	【確率₽】	$X \bullet P(X)$
0	0	0
5	0	0
10	0	0
15	0.01	0.15
20	0.02	0.4
25	0.02	0.5
30	0.03	0.9
35	0.05	1.75
40	0.09	3.6
45	0.17	7.65
50	0.19	6
55	0.09	4.95
60	0.11	6.6
65	0.13	8.45
70	0.08	5.6
75	0.03	2.25
80	0.02	1.6
85	0.01	0.85
90	0.01	0.9
95	0.01	0.95
100	0	0
合計	1	$E(X) = \sum \{X \bullet P(X)\}\$

【期待値(平均値)E】

集計結果より、【期待値(平均値)】E(X) = 53.1

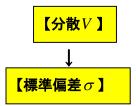
『偏差値』発 賢者の統計・愚者の統計 1時間 29

-感動!ガリレオ流 帝都大学への数学 vol.7-

VI)【分散(variance) 】と【標準偏差(standard deviation)】を計算します。

【確率変数	【確率₽】	$(\mathbf{x} - \mathbf{r}(\mathbf{x}))^2$	(())2 -()
X]	【唯一學》	$(X-E(X))^2$	$(X - E(X))^2 \bullet P(X)$
0	0	2819.61	0
5	0	2313.61	0
10	0	1857.61	0
15	0.01	1451.61	14.5161
20	0.02	1095.61	21.9122
25	0.02	789.61	15.7922
30	0.03	533.61	16.0083
35	0.05	327.61	16.3805
40	0.09	171.61	15.4449
45	0.17	65.61	11.1537
50	0.12	9.61	1.1532
55	0.09	3.61	0.3249
60	0.11	47.61	5.2371
65	0.13	141.61	18.4093
70	0.08	285.61	22.8488
75	0.03	479.61	14.3883
80	0.02	723.61	14.4722
85	0.01	1017.61	10.1761
90	0.01	1361.61	13.6161
95	0.01	1755.61	17.5561
100	0	2199.61	0
			$V(X) = \sum \{(X - E(X))^2 \bullet P(X)\} =$
	1		ll 229.39

集計結果より、【分散】V(X) = 229.39



よって、【標準偏差】 $\sigma = \sqrt{V(X)} = \sqrt{229.39} = 15.146$

-感動!ガリレオ流 帝都大学への数学 vol.7-

さぁ、【期待値(平均値)】Eと【標準偏差】 σ が出ましたから、これで準備万端です。

Ⅷ)最後に、【偏差値】を求めましょう。

【確率変数 X 】	【期待値 E 】	【標準偏差 σ 】	【偏差値 Z 】 $Z(X) = (X - E(X))/\sigma(X) \times 10 + 50$
0			14.941
5			18.242
10			21.544
15			24.845
20			28.146
25			31.447
30		15.146	34.748
35			38.050
40			41.351
45			44.652
50	53.1		47.953
55			51.254
60			54.556
65			57.857
70			61.158
75			64.459
80			67.760
85			71.062
90			74.363
95			77.664
100			80.965

これで、数学の模試テスト得点に対応する【偏差値】が出揃いました。

得点 70 点の生徒 ⇒ 偏差値 ≒61

得点 80 点の生徒 ⇒ 偏差値≒67

偏差値を求める作業はこれで充分理解できたと思いますが、如何でしょうか?

では、最後に肝心なことを見ておきますよ。

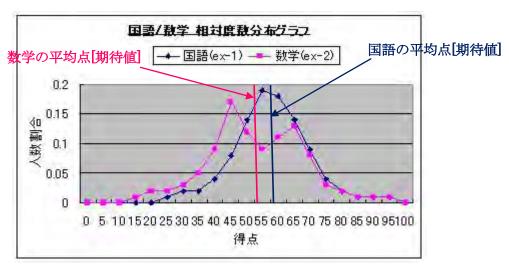
数学と国語の得点分布を比べながら、偏差値を比較してみましょう。

『偏差値』発 賢者の統計・愚者の統計 1時間 31

-感動!ガリレオ流 帝都大学への数学 vol.7-

【偏差値】から見えること 6)

国語と数学の得点分布を同じグラフ上に表すと、下記のようなグラフになります。 同じグラフに重ねると、分布の違いがよく分かりますね。



この分布の相違を日本語で表現しておくと、

- ・国語と数学では得点の上位層はほとんど一致
- ・数学では平均点~70点の分布が小さく、その分、平均点以下の分布が大きい その結果、得点の高い山と得点の低い山の2つのピークを持つ

少し抽象化して表現すると

- 国語は平均点近くをピークとして対称に近い自然な分布(後述の正規分布に近い)
- 数学は、できる人とできない人がはっきり分かれる二極化の傾向

このような把握の下で各得点の偏差値をピックアップした下の表を見てみましょう。

得点	偏差値		偏差値
汚点	ex-1 (国語)		ex-2(数学)
25	23	<	31
35	31	<	38
45	40	<	45
55	48	<	51
65	56	<	58
75	64	=	64
85	73	>	71
95	81	>	78

得点が同じ 55 点であっても、数学の 55 点の方が国語の 55 点よりも価値があること、 同じ点数を取っても、分布が違えば偏差値は結構変わってくることが分かりますね。

『偏差値』発 賢者の統計・愚者の統計 1時間 32

-感動!ガリレオ流 帝都大学への数学 vol.7-

数学では、国語に比して低い点数の方に皆の分布がシフトしています。

ですから、低い得点であっても意外に相対的なポジションは低くないというわけです。 従って、同じ点数でも偏差値は国語に比べて高目に出てきます。

これは、感覚的にも妥当な評価として受け入れられそうです。

一方、数学で高い得点を取った人の場合、同じ国語の得点よりも評価は高くなりそうな ものですが、実際の偏差値は少し低めに出てきます。

ここは、実際の感覚とは少し違和感を感じるのではないでしょうか?

これは最初の(1)式から予想できますように、平均点が低いため得点の高い層には有利に 見えますが、逆に分母の標準偏差が大きくなっているため、即ち、散らばり具合が大き いということの影響の方を強く受けて低く出てしまうということなんです。

要するに、平均点と標準偏差の増減具合で決まるというわけですね。

「理系の偏差値は何故低い?」という疑問がよくありますが、ここまでの例で言うと、 国語(ex-1)の例が理系、数学(ex-2)の例が文系の分布により近いと定性的に考えると辻褄 が合うのではないでしょうか?

どちらかというと、理系は目的意識をはっきり持った受験生が多いですが、文系の場合 は「とりあえず大学」という層が裾野に広がっていると見受けられますから、学力分布 の二極化傾向が強く、本例のように一定のレベル層以下の受験生の偏差値が高めに出る と考えられます。

ともかくも、偏差値は、試験問題の難易にかかわらず、自分の素点と平均点の比較だけでは 分かりにくい「受験者の中でどのような位置にあるのか?」のメドを掴むためには便利な統 一規格ですし、素点だけでは判断できない「どの教科が弱いのか?」などを判断するにも便 利な指標となります。

同じ 55 点を取っても、数学の 55 点は国語の 55 点より価値があるということになり、どち らかと言うと国語を苦手科目として優先的に取り組む方がよいというような判断材料にも なるわけですね。

しかし、あまり細かいところを気にするのも意味がないということを頭に置いてください。 受験者層の違う模試での偏差値を比べても意味がないということ以上に、分布の様子が、こ の後説明する【正規分布】から外れれば外れるほど、【偏差値】は指標としての有効性がなく なって来るのです。

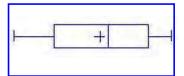
『偏差値』発 賢者の統計・愚者の統計 2時間 33

-感動!ガリレオ流 帝都大学への数学 vol.7-

2時間目:【箱ひげ図】を書く・・・四分位という見方

1) 【箱ひげ図 (box-and-wisker plot)】とは?

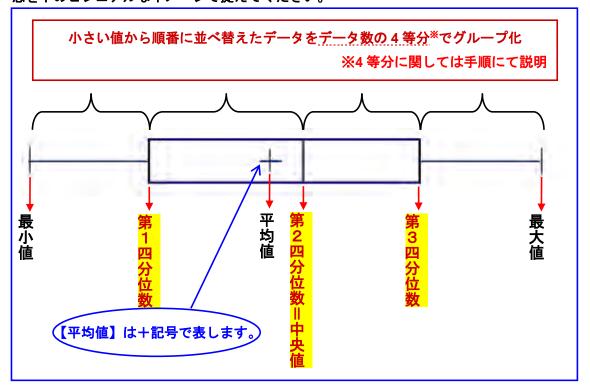
【箱ひげ図】は、全てのデータをデータ数が均等な4グループ に分けて、その分け目の値3つと最小値、最大値、平均値の合計 6個の値を一目見て分かるようにした図です。(右図)



ちょうど『箱』のような長方形と『ひげ』のような直線で表されることから【箱ひげ図(boxand-wisker plot) 】と呼ばれています。

図の向きは、上図のように横向きに書く場合もあれば縦向きに書く場合もあります。 データの性質によって、フィットする向きに書けばいいだけの話なんです。 教科書では、たいてい横向きに書かれていると思いますので、本講座でも横向きで説明して いきますね。

【箱ひげ図】を書く目的は、対象とする統計データの大よその分布具合をビジュアルに把握 するためのものだということで今は認識しておいてください。先ず、【箱ひげ図】の意味と概 念を下のビジュアルなイメージで捉えてください。



『偏差値』発 賢者の統計・愚者の統計 2時間 34

-感動!ガリレオ流 帝都大学への数学 vol.7-

2) 【箱ひげ図】を書いてみる

この意味と概念さえ理解していれば、あとは、数学と言うより小学校の算数です。 簡単すぎて拍子抜けしてしまうかもしれませんよ。

『大よその分布具合』と言った理由が、この概念図で分かりましたか? 沢山あるデータの内、たった6個のデータで分布を表現しようと言うのですからね。

さて、知らない言葉が出てきました。

『第〇四分位数(しぶんいすう:quartile)』と『中央値(median)』なる2つの言葉です。 いずれも『統計』でしかお目にかからない言葉です。

少なくとも僕は仕事上でも今に至るまで、どちらの言葉も使ったことがありません。

でも、『四分位数』なる言葉は「小さい値から順番に並べ替えたデータをデータ数の 4 等分 でグループ化」という言葉から、その由来も意味も想像がつくのではないでしょうか?

そこで、データを等しく4等分する手順に沿って説明して行きますが、具体的に分かりやす く理解してもらうために、サンプルデータで実際に作業しながら進めますね。 その元データは下表です。(10名の国語のテスト結果)

生徒番号I	1	2	3	4	5	6	7	8	9	10
国語得点 X	85	52	64	93	42	76	73	38	88	69

i) 元データを小さい値から順番に並べ替えたデータを用意します。

サンプルの元データからですと、

38	42	52	64	69	73	76	85	88	93
			•		. •	. •	-	-	

となりますね。

ii)次に、順番のままデータを2つのグループに等分します。

データ数は10個ですね。

ですから、「2 つのグループに等分する」ということは、小さい数 5 個と大きい数 5 個 のデータにきれいに等分割できることになります。

	-			_					_
				69 73					
. 30	12	52	61	60 X 73	76	Q.5	QQ	0.3	
<u> </u>	44	JZ	U- 1	09 73	10	OJ	00	93	

『偏差値』発 賢者の統計・愚者の統計

-感動!ガリレオ流 帝都大学への数学 vol.7-

iii) さて、この全データの『中央値』がここで決まることになります。

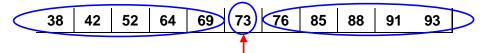
残念ながら、偶数個のデータの場合は、きれいに等分割はできますが、ちょうど真ん中 にデータが存在しません。

10 個のデータがあるサンプルデータの場合ですと、5 番目と6番目のデータが「真ん中 と言えば真ん中」と言えます。

こういった場合は、この2つのデータの平均を取って、これを『中央値』とするのです。 ですから、(69+73) ÷2=71 が『中央値』となります。

データが奇数個の場合は、等しく2つのグループに分割はできませんが、その分、ちょ うど真ん中に1つのデータが存在することになります。

例えば、下のように 11 個のデータになると



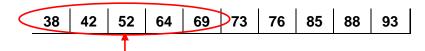
この場合は、このちょうど真ん中のデータ値そのものが『中央値』となります。 上の例では、73 そのものが『中央値』となります。

グループとしては3つに分割するような形で考えればいいわけです。

そして、この『中央値』は4等分した時にできる3つの分割点の内小さい方から2番目 の分割点になりますから、このデータは『第2四分位数』とも名付けられるのです。

iv) 次は、2 分割されたグループのうち小さい方のグループの『中央値』を求めます。 これは4分割した時の最初の分割点になりますから、『第1四分位数』と呼ばれます。

サンプルデータでは下記の 部ですね。



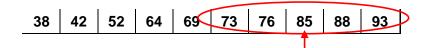
データが5個と奇数個ですから3番目のデータ52が『第1四分位数』となります。

『偏差値』発 賢者の統計・愚者の統計

-感動!ガリレオ流 帝都大学への数学 vol.7-

v) 同様に、2分割されたグループのうち大きい方のグループの『中央値』を求めます。 これは4分割した時の最後の分割点になりますから、『第3四分位数』と呼ばれます。

サンプルデータでは下記の 部ですね。



データが5個と奇数個ですから8番目のデータ85が『第3四分位数』となります。

- vi)『最小数』と『最大数』を求めます。(と言うより、ピックアップですね) 小さい順に並べた表の左端の38が『最小数』、右端の93が『最大数』ですね。
- vii)『平均値』を求めます。 さんざんやってきましたので今さら説明する必要はありませんね。

最後にもう一度まとめておきますと、

生徒番号 I	1	2	3	4	5	6	7	8	9	10
国語得点 X	85	52	64	93	42	76	73	38	88	69

なるデータの【箱ひげ図】を書くために、

小さい値から順番に並べ替えたデータ

38 42 52 64 69 73 76 85 88	93
--	----

を作成。

【中央値】→※『第1四分位数』と『第3四分位数』を求め、『最小数』と『最大数』をピッ クアップして、さらに元データの『平均値』を計算します。

合計6つの値が求まれば【箱ひげ図】を書く準備はOKです。I

※【中央値】→『四分位数』の手順は必要ですが、その他の順番はどうでもいいですよ!

最後に、次のような表に数値をまとめておき、【箱ひげ図】の意味と概念に従って作図してい きます。

『偏差値』発 賢者の統計・愚者の統計

-感動!ガリレオ流 帝都大学への数学 vol.7-

今回のサンプルデータでは、

最小値	第 1 四分位数	第2四分位数(中央値)	第3四分位数	最大値	平均値					
38	52	71	85	93	68					
35 35 00										

表と対応させた【箱ひげ図】が完成します。(対応がよく分かるでしょ!) 沢山の【箱ひげ図】を書きたい場合は、上表のフォーマットを用意しておいて、求めた数値 を入れていくと便利ですよね。

ただ、【箱ひげ図】だけが宙ぶらりんで書かれていても、何だか意味不明ですよね。 単独で見ることは少ないでしょうが、それにしても尺度が何なのかがさっぱり分かりません。

ですから、単独図であっても、次ページのように絶対的な尺度の上に表現しましょう。 同一尺度の多くの統計を比較したい場合はなおさらです。

さて、ここまで来れば、【箱ひげ図】の概念を日本語でしっかりと表現しておきましょう。

要するに、【箱ひげ図】の根幹をなす【データを小さい値から昇順に並べて、順番どおりで等しく 4 グループに分ける】ということは、全体を 25%ずつ 4 グループに分けるということですね。

そう考えると、例えば偏差値データの場合で表現すれば

- ・左のひげは成績下位者 25%の得点ポジションを示す。
- ・箱全体は、成績中位者 50%の得点ポジションを示す。
- 箱の内部区切りは全体順位のちょうど真ん中順位の生徒の得点を示す。
- ・左室は中位者の下位半分、右室は中位者の上位半分の得点ポジションを示す。
- 右のひげは成績上位者 25%の得点ポジションを示す。

といった表現ができます。

次章で学んでいく【正規分布】においては、データが $m-\sigma$ と $m+\sigma$ の間に全体のデータの68.3%が分布するという事実と照らし合わせると興味深いですよ。

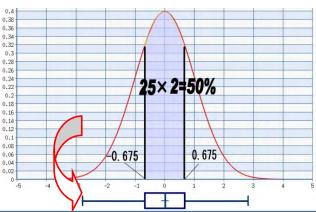
【正規分布】を学んだあと、ここに戻ってきて、下の枠内のことを検証してくださいね。

高校数学 『偏差値』発 賢者の統計・愚者の統計 2時間目 38

-感動!ガリレオ流 帝都大学への数学 vol.7-

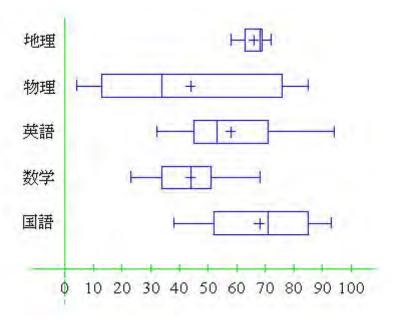
【箱ひげ図】の『箱』の部分は、すっぽりと【正規分布】の $X = m - \sigma$ と $X = m + \sigma$ の間、 即ち【標準正規分布】のZ = -1とZ = 1の間に収まっているということになります。 (全体のデータの 50%が分布する範囲は正規分布表から $-0.675 \le Z \le 0.675$)

ですから、【二項分布】でnp がそこそこ 0.34 大きいデータの【箱ひげ図】を書けば、 当然左右対称となり、右図のような対応 0.22 関係になります。



-感動!ガリレオ流 帝都大学への数学 vol.7-

最後に、生徒10名の5教科のテスト結果を【箱ひげ図】にして並べた図を見てください。 下に、テストの点数という絶対尺度を入れて絶対値が一目で比較できるようにしていますか ら、非常に比較しやすいですね。



この【箱ひげ図】の元になるサンプルデータは、あらかじめ小さい数値順に並べたデータと して、表を本項の最後に記しておきますので、【箱ひげ図】の作成に慣れるために実際に作成 してみてください。

その前に、君が生徒 10 名の担任の立場だったとして、上で示した 5 教科の成績の【箱ひげ 図】を見て、どう分析し、さらに何を調査する必要があり、各教科の先生に何を要望すれば いいのかを考え、書き出してみてください。

どうですか?

分析を終えたら、本項の最後に記した元になるサンプルデータとにらめっこしながら、自分 の分析が的を射ていたかどうか、分析できるはずのことが抜けていなかったかをチェックし てみてください。

この分析力こそが、諸君が社会に出てから、直接評価される能力に繋がる能力なんです。 こんな図が書けるぐらいの能力では誰も評価してくれませんよ。

『偏差値』発 賢者の統計・愚者の統計 2時間 40

-感動!ガリレオ流 帝都大学への数学 vol.7-

もちろん、【度数分布グラフ(ヒストグラム)】の方が見たままをそのままイメージ化できま すから、はるかに分析もし易いわけですが、多くのデータをこれで表すとカサ高くなってし まいますね。

概念からも分かるように、【箱ひげ図】は統計データの分布具合を全てイメージ化している ものではありませんから分析はし辛いですし精度も落ちるでしょが、比較したい沢山のデー タを一目瞭然に並べて大まかな簡易的分析ができるメリットはあります。

ただ、『統計』の表現としてはそれほど意味がない『四分位数』を使っていますこともあり、 概念通りの【箱ひげ図】が利用されることはそれほどないと思います。 また、パーセンタイル(25,50,75)などを独自に変更して使っている場合もあります。

さらに、正規分布に近いと考えられる場合には、『四分位数』ではなく『標準偏差』を使って 【箱ひげ図】的に表現するという方法が医学系や心理学系論文でも結構見られますように、 何を統計値として【箱ひげ図】を書いているのかで混乱を招く可能性もあります。

株式や為替市場のリアルタイムの動きを表すチャートにある『ローソク足』も、【箱ひげ図】 的に表現されたものと言えそうですね。(始値・最安値・最高値・終値)

最後に、先ほどの【箱ひげ図】の元となるサンプルデータ表を示しておきます。 本データを元に、ほんの 2.3 個も【箱ひげ図】を作成すれば、図の書き方は充分に身につく と思います。

データ No.	1	2	3	4	5	6	7	8	9	10	合計
国語	38	42	52	64	69	73	76	85	88	93	680
数学	23	28	34	39	43	45	47	51	62	68	440
英語	32	35	45	46	51	55	66	71	85	94	580
物理	4	9	13	28	33	35	74	76	83	85	440
地理	58	60	63	64	68	68	68	69	70	72	660

どうでしたか?

「図を書くぐらいなら誰でもできるじゃん!」と思ったでしょ! そうです、こんなもんが書けるだけでは何の足しにもならないのです。

必要なことは、【箱ひげ図】を見てデータをどう分析するかということにあるんです。 君の分析と元データをにらめっこして、分析力を磨いてみてください。