日本赤十字九州国際看護大学/Japanese Red

Cross Kyushu International College of

Nursing

Mean and standard deviation

メタデータ	言語: jpn 出版者: 公開日: 2020-08-04 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://jrckicn.repo.nii.ac.jp/records/716

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.

第３章 平均, 偏差, 分散, 標準偏差

今日のテーマは平均値、偏差、標準偏差、分散の考え方です。どれも中学校あるいは高校の数学ですでに身につけているはずです。以前学んだことを思い出し、さらに考えを深めてください。最初にお話しするのは平均値の歴史です。

１　平均値の考え方;歴史的発展

・一人の特徴の数値化

健康や保健に関連して、人の特徴を数値で表して理解することは、医療従事者は一般的に行っています。看護師が患者さんを見た場合に、その人の体温、身長、体重、血圧、心拍数などをすぐに思い浮かべるでしょう。ここでは特徴として身長を取り上げます。皆さんの目の前にいる一人の友人の身長、例えば 160 センチとします。

・集団（例:10 人の友人）の特徴の数値化

さて目前に、一人ではなく、10 人の友人がいるとします。10 人の身長は同じではありません。一人ひとり値が少しずつ異なります。ではこの 10 人の特徴を何か 1 つの値に代表させて捉える事はできるでしょうか。これは保健統計学の基礎になる考え方で、平均値を用います。そんなの常識だよ！と皆さんは言うかもしれません。しかし平均値の考え方は昔から常識だったわけではありません。

・古代～15 世紀の考え方

15 世紀までは、一人一人身長や胸囲が異なる 10 人の人がいた時、その 10 人の平均値で代表させるというような考え方はなく、「いろいろな身長や胸囲の人がいる」という事実認識にとどまっていました。

・16 世紀以降の, 考え方の転換

状況が変わったのは 16 世紀以降です。統計的な考え方が発展し、個体数を 2 以上、n までの集団に拡張し、その集団に代表値があると考え、代表値の推定値として算術平均を使うと言う考え方が現れてきました。たとえば天文学では「惑星の位置や月の直径を求める際、何回も観測して計測を繰り返し、その平均値を取ると計測の誤差を減らせる」など、平均値の考え方が様々な分野で「計測の誤差を減らす考え方」として 16 世紀以降ヨーロッパに普及し始めました。

・19 世紀，ケトレーによる革新

人間に対して平均値を使うという革新を始めたのは 19 世紀前半に活躍した天文学者・統計学者であるベルギー人アドルフ・ケトレー(1796-1874)です。ケトレーは多くの人々を観察する中で「一人ひとりの人は個々別々であっても、観察の対象となる個人の数を増やしていくと、人（人々）の平均的な特徴がだんだんに明らかになってくる」と考え、そのような特徴を持つ人を平均的人間と呼びました。観察の数を増やして得られた分布がどのようなものになるかに関連して「ケトレーには釣鐘型の正規分布曲線が至るところに見えた」とされています。ある実験でケトレーは 5,738 人のスコットランド人兵士の胸囲を測定し、その値から正規分布図を作成し、得られた結果

と理論から導かれる分布図とがほとんど完璧に対応することを示しました。

ケトレーは身体的データの計測を科学的に発展させたことでも知られ、体重（ｋｇ）を身長（m）の二乗で割ったＢｏｄｙ　Ｍａｓｓ　Ｉｎｄｅｘは、ケトレー指数とも呼ばれ、人の肥満度を表わす体格指数として医療や看護の分野でも広く使われています。

こうしてケトレー以後は、人間に関連した様々な科学において、集団に平均値の考え方さらに標準偏差、また正規分布といった統計学の分野で発展されてきた考え方を当てはまることが一般化しました。

母集団から標本を抽出し、ある 1 時点において横断的な標本調査を行う場合、対象とする集団の様々な健康の特徴をとらえる上で、平均値は最もよく利用される指標の 1 つです。皆さんもこの講義の後半で米標本により計算演習を行いますが、その際も出発点は平均値の計算になります。

２　基本統計量;平均, 偏差, 分散, 標準偏差

さて天才的な数学者ケトレーが観察と推論から見出した「平均的な人間」という考え方は現代の統計学において集団を捉える際の基本です。この考え方は様々な集団に当てはめることが可能です。例えば学生の皆さんが属しているこの大学の二年生という集団、皆さんが将来就職する病院の入院患者という集団、様々な集団が考えられます。皆さんも勉強する時に、なんとなく平均値・標準偏差値などと考えるのではなく、具体的な特定の集団をぜひイメージしてみてください。

・要約統計量（基本統計量）

平均的な人間を数値で要約して示すのが要約統計量（基本統計量）です。特に大切なのは、データ内の特定の位置を示す「代表値」とデータのばらつきを示す「散布度」です。

１）代表値（データ内の特定の位置を示す量）

・平均値 mean

平均値は、データＸの合計（Σ）をデータ数（データの個数、ｎ）で割った数値です。算術平均とよび、Ｘの上にバーをつけて、またはμ（ミュー）で表します。たとえばデータが {1, 2, 4, 6, 9} ならばΣは 22、ｎは 5、平均値は 4.4 です。

・中央値 median

中央値は、データを大きさの順に並べたとき、真ん中の値です。データ数が奇数のときは、ちょうど真ん中の値です。データ数が偶数なら、真ん中の 2 つの値の平均値です。たとえばデータが {1, 2, 3, 4, 5} ならば中央値は 3、データが {1, 2, 3, 4, 5, 6} なら中央値は 3.5 となります。

・最頻値 mode

最頻値とは、データから度数分布表やヒストグラムを作ったとき、最も度数が多い値のことです。たとえば、データが{1, 2, 3, 4, 4, 5}ならば最頻値は 4 です。

・最小値 minimum／最大値 maximum

データの中で最も小さい値が最小値、最も大きい値が最大値です。たとえばデータが {1, 2, 3, 4, 5, 6}ならば最小値１、最大値 6 となります。

２）散布度

データ全体のばらつきを示す値です。

・出発点としての偏差 deviation

あるデータの実際の値と平均値の差が、偏差です。たとえばＡさんの身長が 164cm,Ｂさんの身長が 158cm,クラスの身長の平均値が 160cm とすると、身長の偏差はＡさん＋4cm、B さん-2cm となります。さて偏差は、個々の値が平均値からどれくらい大きいか小さいかを直感的に知るために便利な値なのですが、集団全体について偏差を合計するとゼロになってしまいます。

そのため、データ全体のばらつきを示すためには、偏差をさらに加工する必要があります。

・分散 variance

そこで＋やーの値をとる偏差をそのまま用いず、２回掛け算して偏差二乗とするアイデアが出されました。偏差二乗は必ずプラスの値になります。この値（偏差二乗）をＡさん、Ｂさん、Ｃさんの場合・・・と合計（Σ）し、最後にデータ数ｎで割ると、偏差二乗の平均値が求められます。この値を分散 variance と呼び、ＶＡＲまたはσ2 と書きます。

分散は、確かにデータ全体のばらつきを示す値ですが、二乗したために、元のデータの単位（長さ、重さ）が「長さ×長さ」「重さ×重さ」に変わってしまい、扱い難いとの議論もあります。

・標準偏差 Standard deviation, ＳＤ, σ

標準偏差とは、分散の平方根です。平方根をとることにより、データの単位を元に戻したと考えられます。データが測定値の場合、標準偏差は通常、測定誤差をあらわすとされます。

動画上での計算演習

事例	身長
1	158
2	153
3	162
4	167
5	150

合計Σ＝790.0　　平均値μ＝790.0／5＝158.0

i	xi データ	xi 偏差	xi 偏差^2
1	158	0	0
2	153	-5	25
3	162	4	16
4	167	9	81
5	150	-8	64
合計Σ	790.0	0	186.0（←分散）
平均μ	158.0		37.2

標準偏差＝ $\sqrt{37.2}$ ＝ 6.09

３　平均値と標準偏差値の意味

さて今日は、ケトレーの平均的人間（平均人）という捉え方から出発し、集団の代表値やばらつきを示す値についてお話ししました。平均人の捉え方は、社会に大きな影響を与えています。平均人は建築物や交通機関のデザインをする上でも大切です。皆さんが用いる机や椅子、エレベータなども、平均人の身長や体重に合わせてデザインされています。

平均的な範囲に入っているか、そこから外れているかは、健康や疾病を考える時も大切です。健診で測定した皆さんの体重や血色素の値を思い出してください。自分の値がクラスの平均値より高いか低いかは偏差で捉えられます。一方、クラスの値が集団としてどのくらいバラつくか、自分はクラスのバラツキの中でどの辺りに位置付けられるか、を考えるためには、標準偏差ＳＤ（STANDARD DEVIATION）が大切です。

もう一度、正規分布曲線を確認しましょう。平均値プラスマイナス１ＳＤの間にデータの６８．３％、平均値プラスマイナス２ＳＤの間にデータの９５．５％、平均値プラスマイナス３ＳＤの間にデータの９９．７％が含まれます。

皆さんが将来看護師になったとき、基本的な検査値について、平均値と標準偏差から正規分布曲線をイメージできると、一人の人の値から、その人が医療を必要としているかの概要を判断できます。たとえば皆さんの同級生、Ｂさん２０歳・女性は、先日の健康診断で血色素１０．５Ｇ／ＤＬでした。またクラス全員の女性の血色素は平均値１３．０Ｇ／ＤＬでした。Ｂさんは貧血を心配する必要があるでしょうか。平均値１３．０より２．５低いという情報だけだと、判断できません。しかしクラス全員の血色素の標準偏差ＳＤが１．０Ｇ／ＤＬだと分かっていたらどうでしょうか。Ｂさんは２ＳＤ（標準偏差の二倍）よりも、さらに低い値だと判断できます。Ｂさんは直ぐに健康管理室に相談すべきでしょう。

さて平均値・偏差・分散・標準偏差などが、看護学を学ぶ上で実に大切な考え方であるということは十分に理解できたと思います。計算方法は中学校や高校の数学の時間に身につけたはずですが、思い出せたでしょうか。紙と鉛筆でも計算できるよう、復習しておいてください。

演習問題

１．ある標本６名（ＡさんからＦさんまで）の体重(kg)は以下の値でした。

45，47，52，54，54，62kg.

中央値、最頻値、最小値、最大値を答えてください。

２．先ほどと同じ６名の体重についての計算です。

45，47，52，54，54，62kg.

平均値、分散、標準偏差を求めてください。

３．あなたのクラスの身長の平均値を 160.0 ㎝、標準偏差を 3.0 ㎝とします。あなたのクラスメートの一人、Ａさんの身長は 163.0 ㎝、Ａさんは自分の身長が高すぎることを気にして、落ち込んでいるようです。あなたはＡさんにどんな言葉をかけますか。

ワークシート：基本統計量計算

（必要に応じてコピーして使ってください）。

変数（　　）

i	データ x_i	偏差 $x_i - \overline{x}$	偏差二乗 $(x_i - \overline{x})^2$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
合計 Σ		偏差 和	偏差二乗和
平均 Σ/n			(標本)分散 (偏差二乗の平均)

(標本)**分散** =＿＿＿＿　(標本)標準偏差=$\sqrt{}$= ＿＿＿＿＿＿

変動係数 = 標準偏差／平均値 = ＿＿＿＿＿＿