第５章
推測統計学

Koishikawa Philosophy II

12 標本調査とは？

集団の一部分のみを取り出して，全体を知る方法とは？

第 1 章で学んだ内容は，統計学の中でも特に「**記述統計学**」と呼ばれる分野です。これは例えば，中学生女子**全員の**運動能力の数値を調べ，体力や運動能力の低下が見られるか見られないかを判断したように，**すべてのデータを**工夫して整理することによって，全体がどのような傾向にあるか，どのような性質をもっているかを探るものでした。

このように，集団のすべてを対象に調査することを，統計学では**全数調査**と呼びます。

ところが，集団の性質を調べたいとき，いつもその集団のすべてを調査することができるとは限りません。

例えば，ある缶詰製造工場で，製品に不良品が混じっていないかを調べようとするとき，製造した缶詰を片っ端から開けていけば，全数調査によって完璧なデータが得られますが，製品としてはダメになってしまいます。つまり，この場合，全数調査はナンセンスです。

そこで，集団の一部分のみを調査することによって，全体の性質を知る方法はないのか，考えていくことにしましょう。

1．一部だけを取り出して調べる方法

対象とする集団の一部を調べ，その結果から，集団全体の状況を推測する調査を［　　　　］といいます。

この調査において，本来調査したい対象全体を［　　　　］といい，調査のために母集団から抜き出されたものの集まりを［　　　　］，母集団から標本を抜き出すことを標本の［　　　　］といいます。

また，母集団に含まれるものの数を［　　　　］，標本に含まれるものの個数を［　　　　］といいます。

このように，集団の一部のみに注目して全体の性質を推論する統計学の方法を［　　　　］といいます。

問題 1

次のそれぞれの調査は，全数調査と標本調査のどちらが適切か考えてみましょう。

(1) 学校で行う健康診断……［　　　　］調査

理由は…

(2) 真空パックされた食品の中身の品質に関する調査……［　　　　］調査

理由は…

(3) 届いたお年玉くじ付き年賀はがきが当選しているかどうかの調査……［　　　　］調査

理由は…

問題2

次の(1)，(2)は標本調査の方法として適切か考えてみましょう。

(1)　A さんが働いている工場では，果物の缶詰をつくっている。ある日，出荷する缶詰のうち，どれくらいの割合で不良品が含まれているかを調査するため，標本調査を行うこととなった。

A さんは，出荷される予定の 10000 個のうち，適当な 4 個の缶詰を開けて調べたところ，そのうちの 1 個が不良品であることがわかった。ということは，不良品が含まれる割合は 4 個に 1 個，つまり約 25 %であると言えそうだ。

(2)　B 町に暮らしている C さんは，お寿司が大好きである。そこで，B 町にどのくらいの寿司好きがいるのかを知りたくなり，標本調査を行うこととなった。

ある日，B 町にあるすべてのお寿司屋さんに，来店する客にお寿司が好きか嫌いかのアンケートをとるように伝え，後日その回答を受け取ったところ，客の 95 %が，お寿司が好きであるということがわかった。

ということは，B 町に住む人の約 95 %が寿司好きだと言えそうだ。

上の例を考えてもわかるように，標本調査は

① 適切な標本の大きさで

② まったくバラバラの（性質に偏りのない）集団を選んで

行うことが大切です。

くじ引きなどの方法で，母集団から偏りなく標本を選ぶこと，標本を ______ といいます。

無作為抽出では，人間の意志がまったく入らないように，さいころを投げたり，コンピュータでランダムにつくった数を並べた**乱数表**などを使ったりするのが一般的ですが，その手間を惜しんで頭の中でランダムであろうと考えたランダム化を行ってしまうことがよくあるそうです。

乱数さい

実験

0 から 9 までの数字を思いつくままにスラスラと，ただしランダムになるように意識しながら，1 分間書き並べてみましょう。

0… ___ 個，1… ___ 個，2… ___ 個，3… ___ 個，4… ___ 個，
5… ___ 個，6… ___ 個，7… ___ 個，8… ___ 個，9… ___ 個

第１章　第２章　第３章　第４章　第５章

2．標本調査の例

標本調査の例を 2 つ紹介します。

例 1　**母比率の推定**

袋の中に大きさが等しい白玉と黒玉が合計 200 個入っています。この袋の中の玉をよくかき混ぜてから 15 個の玉を取り出したところ，白玉が 6 個，黒玉が 9 個でした。

このとき，最初に袋の中に入っていた白玉の個数は，およそ何個と考えられるでしょうか？

(1)　話を整理してみよう

母集団
（袋の中の玉 200 個）
白玉の数……　？　個
黒玉の数……　？　個

無作為抽出 →

標　本
（取り出した＿＿＿個の玉）
白玉の数……＿＿＿個
黒玉の数……＿＿＿個

全体に対する白玉の比率は，$\dfrac{?}{200}$

全体に対する白玉の比率は，＿＿＿

(2)　計算してみよう

標本における白玉の全体に対する比率（**標本比率**）は　$\boxed{}$

これが，**母集団における白玉の全体に対する比率と同じだと推定する**ことができる。

よって，最初に袋の中に入っていた白玉の個数は，およそ

$200 \times \boxed{} = \boxed{}$（個）

と推定できる。

問題 3

(1)　袋の中に大きさが等しい白玉と黒玉が合計 300 個入っています。この袋の中をよく混ぜてから 20 個の玉を取り出したところ，白玉が 13 個，黒玉が 7 個でした。このとき，最初に袋の中に入っていた白玉の個数を推定してみましょう。

(2)　ある工場で生産した 9500 個の製品から，無作為に 70 個を取り出して品質検査を行ったところ，不良品が 4 個ふくまれていました。9500 個の製品の中に，不良品は約何個あると考えられるでしょうか。

例2　母集団の大きさの推定

ある池にいる鯉の数を推定するために，次のような調査を行いました。

① まず，池のあちらこちらから全部で50匹の鯉を捕獲し，それらに印をつけて，池に放した。

② 2 週間後に，同じようにして池から全部で 80 匹の鯉を捕獲したところ，そのうちの 14 匹に印がついていた。

この結果から，池にいる鯉の数を，10 匹単位で推定してみましょう。

第1章
第2章
第3章
第4章
第5章

(1) 話を整理してみよう

母集団

（池の中の鯉　？　匹）

印付きの数……＿＿＿匹

印なしの数……　？　匹

全体に対する印付きの比率は，$\dfrac{50}{?}$

無作為抽出 →

標　本

（取り出した＿＿＿匹の鯉）

印付きの数……＿＿＿個

印なしの数……＿＿＿個

全体に対する印付きの比率は，＿＿＿

(2) 計算してみよう

標本における印付きの鯉の全体に対する比率は　$\boxed{}$

これが，**母集団における印付きの鯉の全体に対する比率と等しいと推定する**ことができる。

よって，池にいる鯉の数は

$\boxed{}$

と推定できる。

※　この方法は**標識再捕法**（**捕獲再捕獲法**ともいう）と呼ばれ，昔から行われていた推定法です。この方法は，湖の中の魚のように，限られた条件の中で用いられています。

問題4

(1) ある湖にいる魚の数を推定するために，次のような調査を行いました。

① まず，湖のあちらこちらから全部で 100 匹の魚を捕獲し，それらの印を付けて，湖に放した。

② 10 日後に，同じようにして湖から全部で 200 匹の魚を捕獲したところ，そのうちの 11 匹に印がついていた。

この結果から，湖にいる魚の数を 100 匹単位で推定してみましょう。

$\boxed{}$

(2) 袋の中に，白い碁石と黒い碁石が合わせて 400 個入っています。これをよくかき混ぜて，20 個取り出し，白い碁石の数を調べてから，取り出した碁石を袋の中にもどすことを 3 回行いました。右の表はその結果です。

	1 回目	2 回目	3 回目
白い碁石 (個)	9	x	6

この 3 回の操作の結果から，標本における白い碁石の割合が碁石全体の 40 ％以上 45 ％未満と推測されたとき，表の x にあてはまる数として，考えられる値をすべて求めてみましょう。

$\boxed{}$

13 標本調査の実践

テレビ番組の視聴率を模擬調査してみよう。

標本調査の例として，テレビ番組の視聴率調査の模擬実験を行ってみましょう。

1．視聴率とは？

視聴率は，テレビの番組や CM がどのくらいの世帯や人々に見られているかを示す 1 つの指標です。

関東地方の世帯数はおおむね 1987 万世帯（2016 年 1 月 1 日現在，総務省公表）です。視聴率は全数調査ではなく，無作為に 900 世帯を抽出して標本調査を行っています。

2．実験

⑴ **準備**

6 つのグループに分かれ，視聴率調査キットをもっていく。キットの中身は

① 白玉と赤玉がたくさん入っている袋

② 標本を取り出すための容器（1 回で約 8 個，12 個，15 個，25 個，30 個すくえる）

⑵ **方法**

① 6 つのグループを，標本の大きさがそれぞれ 4 個，8 個，12 個，15 個，25 個，30 個で推定するグループとする。

② 班の中で，取り出す係，記録係などのように分担する。なお，各班の記録係の生徒の記録をもとに，班員全員が次ページの記録表を完成させる。比率の計算は，下の「比率の換算表」を使う。

③ 袋の中身をよくかき混ぜ，玉を決められた容器で（4 個の班は手で）取りだす。決められた個数に足りなかった場合は，袋の中身を見ずに手で何個か取り出して調整する。

④ 赤玉が何個入っていたか記録する。

⑤ 容器の中身を袋に戻した後，よくかき混ぜ，③をくり返す。

⑥ 結果をヒストグラムにまとめて傾向を調べ，赤玉の割合（視聴率）を推定する。

［比率の換算表］

標本の大きさが 4 個の場合

個	1	2	3	4
%	25	50	75	100

標本の大きさが 8 個の場合

個	1	2	3	4	5	6	7	8
%	13	25	38	50	63	75	88	100

標本の大きさが 12 個の場合

個	1	2	3	4	5	6	7	8	9	10	11	12
%	8	17	25	33	42	50	58	67	75	83	92	100

標本の大きさが 15 個の場合

個	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
%	7	13	20	27	33	40	47	53	60	67	73	80	87	93	100

標本の大きさが 25 個の場合

個	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
%	4	8	12	16	20	24	28	32	36	40	44	48	53	56	60	64	68	72	76	80	84	88	92	96	100

標本の大きさが 30 個の場合

個	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	…
%	3	7	10	13	17	20	23	27	30	33	37	40	43	47	50	53	57	60	63	67	70	73	77	80	…

⑶　結果

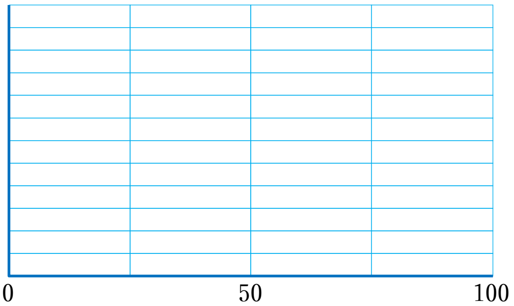
＜実験の記録＞

回	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
個																									
%																									

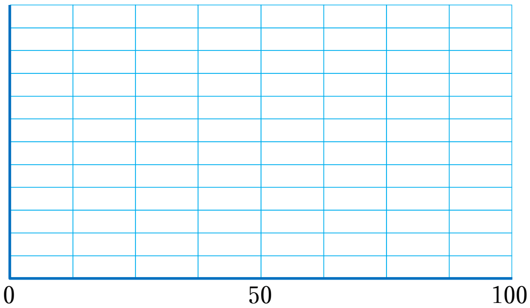
回	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
個																									
%																									

＜ヒストグラム＞（横軸は％）

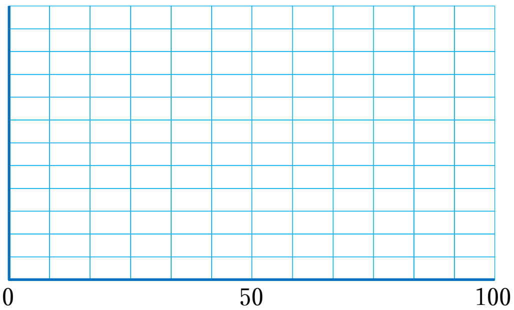
標本の大きさが 4 個



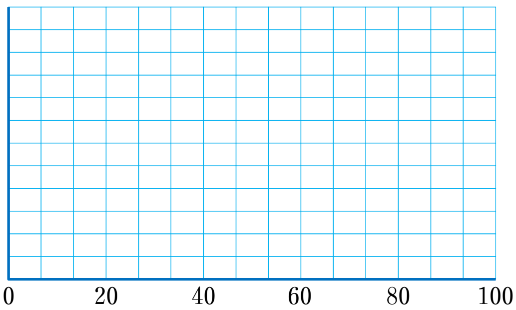
標本の大きさが 8 個



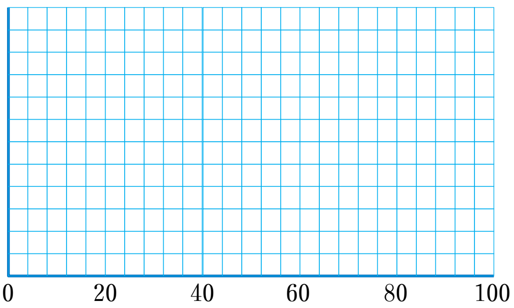
標本の大きさが 12 個



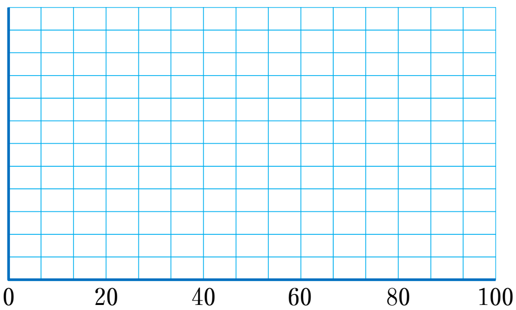
標本の大きさが 15 個



標本の大きさが 25 個



標本の大きさが 30 個



⑷　結論

これらの結果から，私たちのグループの予想視聴率は約＿＿＿％です。

◎　理由

14 正規分布とは？

自然界に多く見られる分布を探ろう。

ここから先は，やや内容が高度になります。それもそのはずで，本来ならば高校 2 年生で学習する内容で，その前提となる知識は「確率」，「関数」，「積分」など，今後の数学で少しずつ学習するものばかりだからです。

中学 2 年生の時点でもっている知識では理解するのは至難の業ですが，難しい部分を思い切ってカットして雰囲気だけでもつかんでみよう，というのがここから先の話になります。

現時点で理解できなくても，なんとなく「こんな感じなのかなぁ」程度につかんでもらえれば十分です。一部，用語の使い方などが本来のものではない場合がありますが，わかりやすさを追求した結果だと思って許してください。

1．まずは「偏差値」から

中学受験，大学受験などの「受験」のときに，気になる数値が**偏差値**。みなさんも小石川の受検にあたって，偏差値を気にすることが多くあったでしょう。

偏差値とは，**全データに対する相対的な位置を示す指標**のことです。自分の得点が平均点と等しければ偏差値は 50，平均点以上であれば偏差値は 50 以上，平均点以下であれば偏差値は 50 以下となります。

偏差値は，以下の式で定義されます。

$$\text{偏差値}=(\text{自分の得点}-\text{平均点})\times\frac{10}{\text{標準偏差}}+50$$

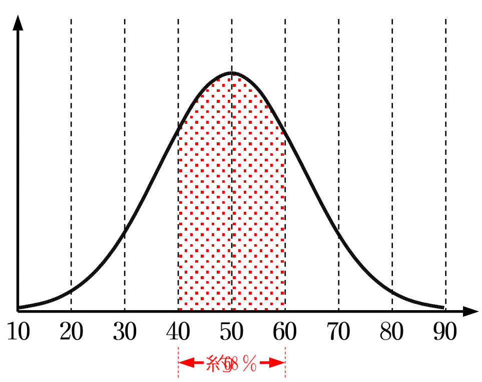
この計算式の「10/標準偏差」の部分により，標準偏差が 10 になるように補正されることになります。つまり，偏差値とは

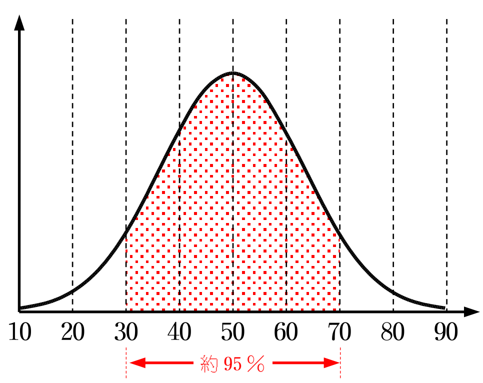
平均点が $\boxed{}$，標準偏差が $\boxed{}$

になるように補正した数値と考えることができます。

一般的な統計学※では，**(平均値)±(標準偏差)**の範囲に約 68 ％のデータが含まれる（細かくいうと約 68.27 ％）と考えます。偏差値に換算した場合は，平均値が 50，標準偏差が 10 となるから，(平均値)±(標準偏差)の範囲は偏差値 40～60 となります。つまり，偏差値 40～60 の範囲に約 68 ％のデータが含まれることになります。

同様に，**(平均値)±(標準偏差×2)**の範囲には約 95 ％のデータが含まれる（細かくいうと約 95.45 ％）と考えます。つまり，偏差値 30～70 の間に約 95 ％のデータが含まれることになります。





例　偏差値 70 以上とは？

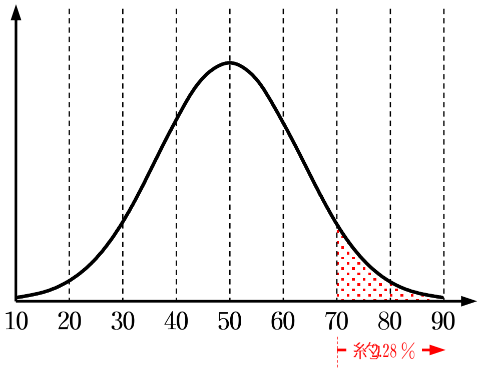
偏差値 30～70 の範囲に約 95.45 ％のデータが含まれる

⬇

偏差値　70　以上の範囲には約 $\boxed{}$ ％のデータが含まれる

⬇

偏差値 70 以上の人は「約 44 人に 1 人」



問題 1

東京大学の偏差値は，予備校などが公開している資料によると，67～70　程度とされています。最も難関な理科三類（医学部）でも，その偏差値は 72～73 程度だそうです。

仮に偏差値 70 として考えると，上の例から「約 44 人に 1 人」という割合になり，「クラスに 1 人くらいは，いずれ東大に行く」ということになりそうです。

しかし，実際は，そんなに東大生がいるわけもなく，進学校でもない限り，クラスに　1　人どころか同学年に 1 人もいない，という場合も少なくありません。

このように，「偏差値 70」が示す割合と，現実が異なる原因は何でしょうか？

注意　偏差値の取り扱いの注意

◎　一般公開されている統計データを見るときは，「その指標の分母は何か？」を見極めることが重要！

→東大に合格できる割合について，もっと現実的な数値を見たいのであれば，偏差値よりも統計データを調べた方が確実

⬇

日本の 18 歳人口は約 120 万人弱，東大の合格者数は 3000 人程度だから「約 400 人に 1 人」が合格するという計算になる。

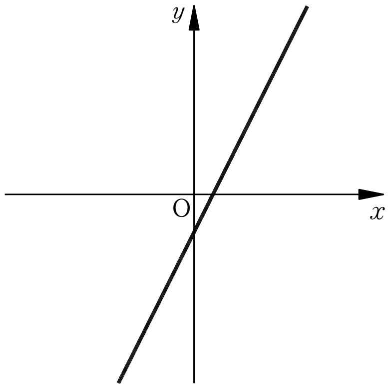
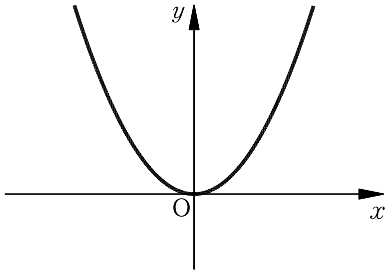
◎　今回紹介した偏差値の考え方は「一般的な統計学」（前ページの※）に基づいた場合という仮定が加わることにも注意が必要です。

「一般的な統計学」とは，「データのばらつき具合が**正規分布**になる場合」という意味です。これを理解するには，「正規分布とは何か？」を知っておく必要があります。

そこで，次は正規分布について詳しく見ていきましょう。

2．確率密度関数

正規分布を紹介する前に，どうしても越えなければならない壁があります。それが**確率密度関数**です。「関数」といえば，現時点でみなさんが知っているものでいうと

関　　数	式	グラフ
1 次関数	$y=ax+b$	
2 乗に比例する関数	$y=ax^2$	

というように，y を x で表した式と，そのグラフがセットになっていました。

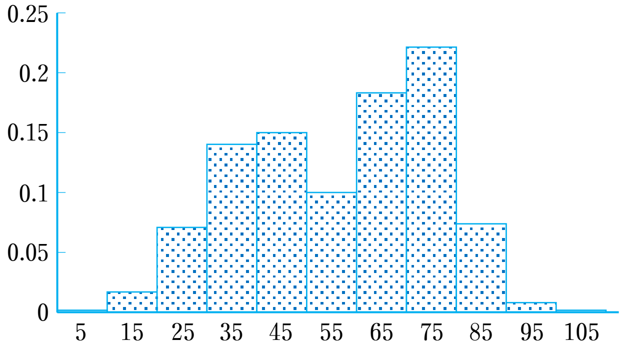
確率密度関数も，「関数」の一種なので，式とグラフがあります。その具体的な形を見る前にもう一度，度数分布表とヒストグラムの話をしておきましょう。

下の度数分布表は，ある県で中学 3 年生を対象に行った英語のテストの得点をまとめたものです。

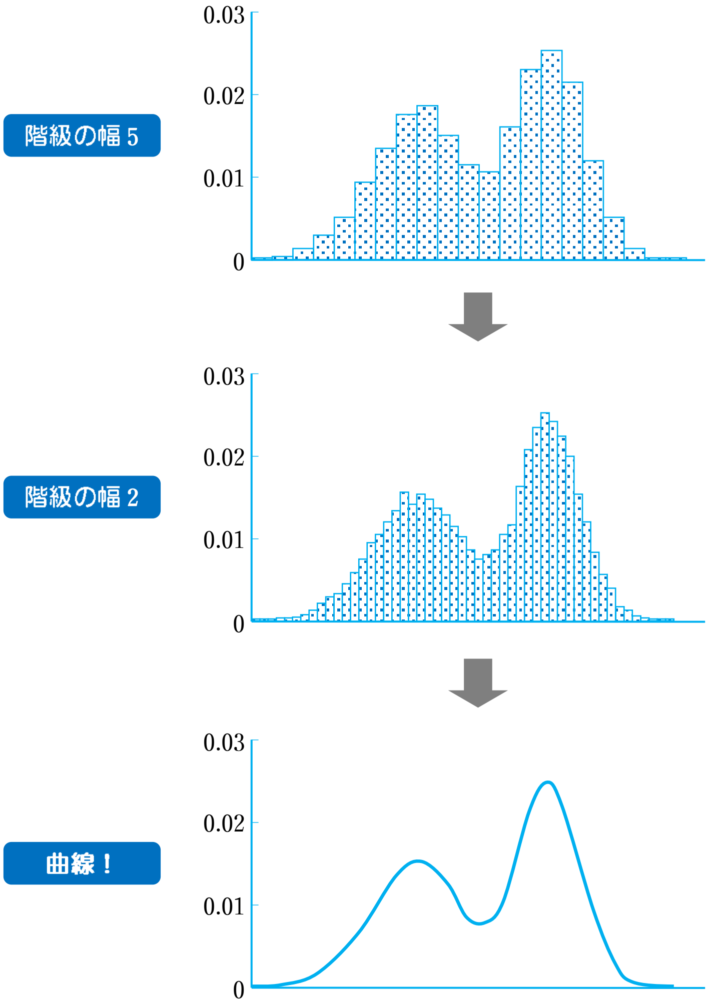
階級（点）	階級値（点）	度数（人）	相対度数	$\dfrac{\text{相対度数}}{\text{階級の幅}}$
以上　未満				
0～ 10	5	86	0.00271	0.000271
10～ 20	15	648	0.02040	0.002040
20～ 30	25	2286	0.07195	0.007195
30～ 40	35	4662	0.14673	0.014673
40～ 50	45	4922	0.15492	0.015492
50～ 60	55	3365	0.10591	0.010591
60～ 70	65	5883	0.18516	0.018516
70～ 80	75	7181	0.22602	0.022602
80～ 90	85	2535	0.07979	0.007979
90～100	95	200	0.00629	0.000629
100～110	105	4	0.00013	0.000013
合　　計		31772	1.00	0.1

ここでは，階級の幅を 10 点とし，今後の説明の都合上，$\dfrac{\text{相対度数}}{\text{階級の幅}}$ も書き加えてあります。

この度数分布表をもとにつくったヒストグラムが下の図です。



このヒストグラムの縦軸は相対度数ですが，説明の都合上，$\dfrac{\text{相対度数}}{\text{階級の幅}}$ に変更します。で，階級の幅を現在の 10 から狭めていくと，ヒストグラムの形状がどうなるか確かめてみましょう。



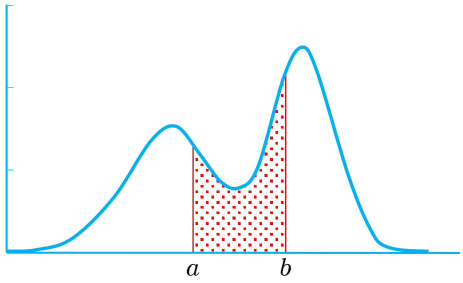
このように，ヒストグラムの**階級の幅を狭めていって最終的に到達する曲線の式**こそが確率密度関数なのです（厳密に言うと今回の例はテストの得点なので，階級の幅を 1 より小さくすることはできませんが，あくまでも感覚的に曲線になっていくだろうと理解してください）。

注意　確率密度関数の大事な性質

①　**確率密度関数のグラフと横軸とで囲まれた部分の面積は，必ず 1** になります。

②　確率密度関数のグラフと横軸とで囲まれた部分のうち，a から b までの部分の面積は，データが a 以上 b 以下となる確率を表します（右図）。つまり，①の性質は，「データが最小値から最大値の間のどこかに入る確率は 1」ということを意味しています。

③　全範囲にわたって，グラフは横軸より上にあります。



問題 2

縦軸に$\dfrac{\text{相対度数}}{\text{階級の幅}}$をとった場合のヒストグラムの柱部分の面積を考えることによって，「確率密度関数のグラフと横軸とで囲まれた部分の面積が 1 になる」理由を説明してみましょう。

3．正規分布とは？

ヒストグラムの山の形を統計学では**分布**と呼びます。そのため，確率密度関数が表す曲線を**分布曲線**ともいいます。

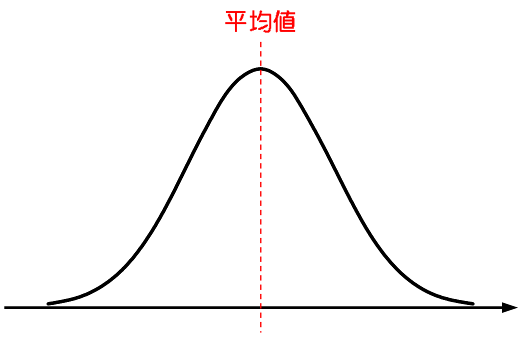
正規分布は，分布曲線の一種で，以下の数式で表される関数のことを指します。

$$y=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ （μ ：平均値，σ ：標準偏差（σ^2 ：分散），π ：円周率）

……が，こんな式はまったく覚える必要ありませんのでご安心ください。

それより正規分布の形と特徴を覚えておいてください。

上の数式をグラフ化すると，下のようになります。



正規分布の特徴は，次の 2 つです。

① **平均値**を中心に，データが**左右対称**に分布している

② 平均値から離れるほど，データの頻度が低くなっていく

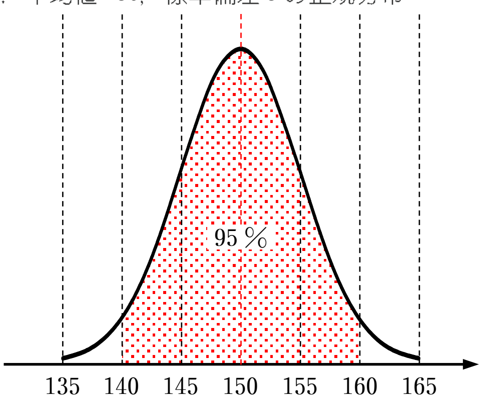
統計処理を行う際は，この 2 つの特徴を理解しておくことが重要です。逆に言うと，これらの特徴さえ理解していれば，さっきの式の意味がまったくわからなくても，統計学的に正しい処理を行えます。

正規分布は，p.96 で示した法則が常に成立します。つまり，

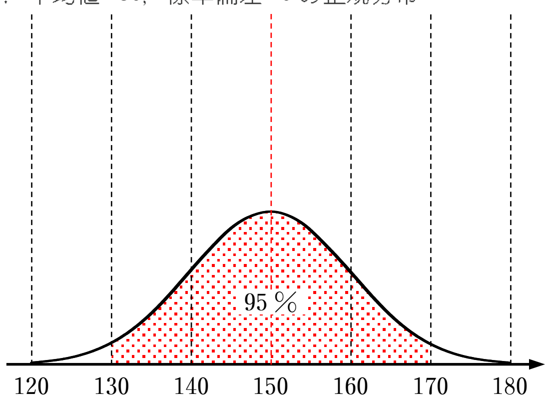
- □ **(平均値)±(標準偏差×1)**……約 68 ％（正しくは約 68.27 ％）のデータが含まれる
- □ **(平均値)±(標準偏差×2)**……約 95 ％（正しくは約 95.45 ％）のデータが含まれる
- □ **(平均値)±(標準偏差×3)**……約 100 ％（正しくは約 99.73 ％）のデータが含まれる

したがって，**標準偏差**の値に応じて，“尖った形状” や “平たい形状” のようにグラフの形が変わります。

ア．平均値 150，標準偏差 5 の正規分布



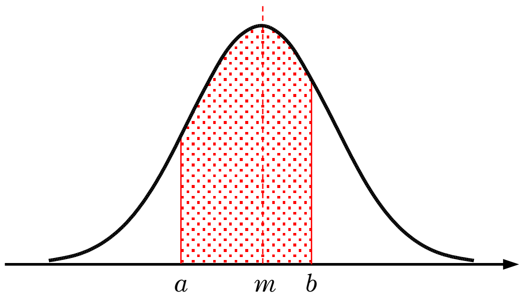
イ．平均値 150，標準偏差 10 の正規分布



正規分布は，自然界に見られる多くの分布に近く，重要なものとして知られています。

4．標準正規分布

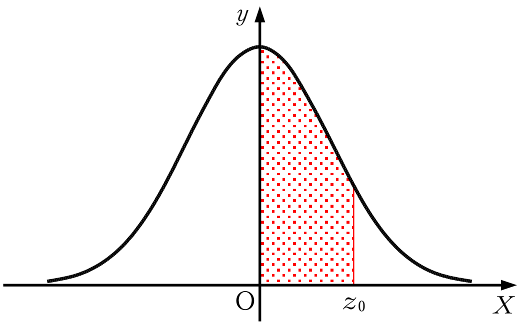
平均値が m，標準偏差が σ（分散が σ^2）である正規分布を，記号で [　　　　] と表します。あるデータ X の分布曲線が $N(m,\ \sigma^2)$ のとき，X は正規分布 $N(m,\ \sigma^2)$ に**従う**といいます。



X の値が a 以上 b 以下である確率は，右の図の赤い部分の面積なのですが，これを求めるのは至難の業です。

そこで，正規分布のある範囲の曲線と横軸で囲まれた部分の面積をあらかじめ一覧にした便利な表があるのです。それが巻末に掲載した**正規分布表**です。

ただし，この正規分布表は，平均が 0，標準偏差が 1 である正規分布 $N(0,\ 1)$ について，右の図の赤い部分の面積を小数で示したものとなっています（正規分布 $N(0,\ 1)$ を [　　　　　　　　] といいます）。



問題３　正規分布表の使い方

標準正規分布 $N(0,\ 1)$ に従うデータ Z について，次の確率を求めてみましょう。

(1)　$0 \leqq Z \leqq 0.65$ である確率 $P(0 \leqq Z \leqq 0.65)$

(2)　$P(0.6 \leqq Z \leqq 1.13)$

(3)　$P(-0.6 \leqq Z \leqq 1.13)$

(4)　$P(Z \leqq 0.6)$

5．標準化

一般の正規分布 $N(m,\ \sigma^2)$ に従うデータ X について

$$Z = \boxed{}$$

とおくと，Z は標準正規分布 $N(0,\ 1)$ に従うことが知られています。このように，X を Z に変換する作業を**標準化**といいます。

自然界に登場する正規分布は，いつも標準正規分布であるとは限らないので，確率を求めたいときは正規分布表を使えるようにするために，その都度標準化しなければなりません。

練習

正規分布 $N(2,\ 3^2)$ に従うデータ X について

$5 \leqq X \leqq 6.2$ となるような確率　$P(5 \leqq X \leqq 6.2)$

を求めてみましょう。

問題４

ある高校の3年生男子150人の身長 X cm が，

平均 170.6 cm，　標準偏差 5.9 cm

の正規分布に従っているものとします。

このとき，身長が 165 cm から 175 cm までの生徒は，およそ何人いるでしょうか。

15 推測統計学に挑戦

無作為に抽出した標本のデータから全体のようすを推定しよう。

1．標本平均の平均値，標準偏差

母集団におけるある量 x の平均値を [　　　　] といい，標準偏差を [　　　　] といいます。

母集団から大きさ n の標本を無作為に抽出し，それらに対する x の値を X_1，X_2，……，X_n とするとき

$$\overline{X}=\frac{X_1+X_2+\cdots\cdots+X_n}{n}$$

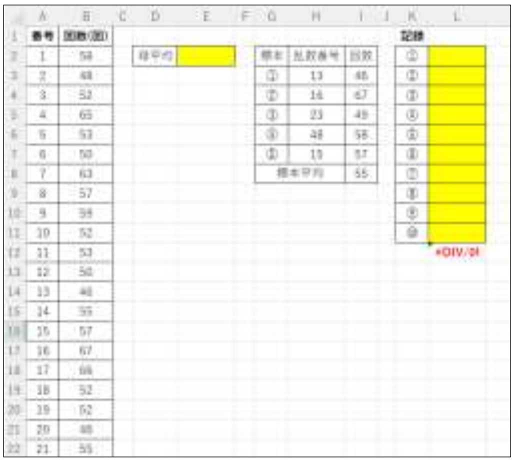
を [　　　　] といいます。

実験

① Teams 上にある「母平均の推定.xlsx」を自分のタブレットにダウンロードする。

② シートの左にある表は，ある高校の男子生徒 50 人の反復横跳びの記録です。これらを母集団として，母平均を求める。適切な関数を使って，E2 セルに母平均を表示させてみよう。

③ L2 セルから L11 セルに，I8 セルに表示されている数値を順に入力していく。ただし，「乱数番号」の欄の 5 つの数字に同じものがある場合は，キーボードの **F9** をおして数値を更新する。



④ 10 個の数値を入力できたら，L12 セルに表示されている数値と②で求めた母平均を比較してみよう。

◎ わかったこと

次の事実が知られています。

> 母平均 m，母標準偏差 σ の母集団から無作為に抽出した大きさ n の標本の標本平均について
>
> 標本平均の平均値は母平均 m に等しい。
>
> 標本平均の標準偏差は母標準偏差 σ を $\sqrt{n}$ で割った $\dfrac{\sigma}{\sqrt{n}}$ に等しい。

母平均 m，母標準偏差 σ の母集団から無作為に抽出した大きさ n の標本の標本平均 $\overline{X}$ は，n が大きいとき，正規分布 $N\left(m,\ \left(\dfrac{\sigma}{\sqrt{n}}\right)^2\right)$ に従います。

問題 1

母平均 50，母標準偏差 18 の母集団から，大きさ 81 の標本を無作為に抽出するとき，標本平均 $\overline{X}$ が 46 以下になる確率を求めてみましょう。

2．母平均の区間推定

母平均を標本平均から推定する場合，母平均を正確に求めることはできませんが，母平均をほぼ確実に含む範囲を求めることはできます。

ここでは，「ほぼ確実に含む」ということを

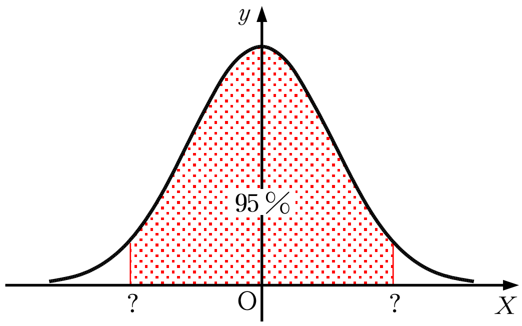
その範囲が母平均を含む確率が 95 ％である

と定めることにします。

この範囲のことを，$\boxed{\qquad\qquad\qquad\qquad}$といいます。

⑴　**準備**

標準正規分布 $N(0,\ 1)$ で，右の図の？の値がいくつになるか，正規分布表を用いて求めてみましょう。



このことにより

$$P(-\boxed{\qquad}\leqq Z\leqq\boxed{\qquad})=0.95$$

であることがわかります。

⑵　**信頼度 95 ％の信頼区間の求め方**

母標準偏差を σ，標本の大きさ n，標本平均を $\overline{X}$ とします。n が大きいとき，$\overline{X}$ は

正規分布 $N\left(m,\ \left(\dfrac{\sigma}{\sqrt{n}}\right)^2\right)$ に従う

と考えてよいので

$$Z=\boxed{\qquad\qquad}$$

とおくと，Z は標準正規分布 $N(0,\ 1)$ に従います。

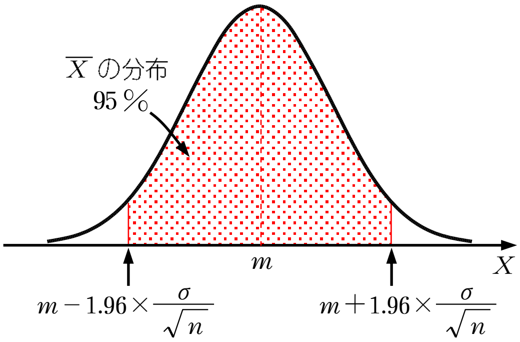
よって，信頼度 95 ％の信頼区間は

$$\boxed{\qquad\qquad\qquad\qquad\qquad\qquad}$$

のように求めることができます。

まとめておきましょう。

母平均 m，母標準偏差 σ の母集団から無作為に抽出した大きさ n の標本の標本平均 $\overline{X}$ は，n が大きいとき，$N\left(m,\ \left(\dfrac{\sigma}{\sqrt{n}}\right)^2\right)$ に従うと考えてよい。



第 1 章
第 2 章
第 3 章
第 4 章
第 5 章

問題 2

体力・運動能力調査

と検索し，e-Stat に公開されている集計データから，令和 3 年度の 14 歳の男子，女子について 50 m 走の「標本数」，「平均値」，「標準偏差」を調べ，この年の 14 歳男女それぞれの 50 m 走の記録の全国平均 m 秒を，信頼度 95 %で推定しましょう。

「母平均の推定.xlsx」の「母平均の推定」シートを利用して，Excel に計算させてみましょう。

3．仮説検定

花子さんは，コインを使ったゲームをしています。コインを何回か投げたとき，

「どうもこのコインは表が出やすいな」

と感じました。そこで，このコインを 100 回投げて表が出る回数を調べたところ，表が 60 回出たことがわかりました。

この結果から，このコインは，かたよりなく作られていない，つまり

『このコインは表が出る確率を裏が出る確率は等しくない』 ……①

と判断してよいのでしょうか。

このことを判断するために，①に反する次のような仮説を立てます。

『このコインの表と裏の出る確率は等しい』 ……②

ある統計的な手法を用いて，②の仮説のもとで

『100 回中 60 回以上表が出る確率は 0.0228』

と求めることができました。

ここで，「めったに起こらないと判断する基準となる確率」を 0.05 と決めると，0.0228 は 0.05 よりも小さいので，

60 回以上表が出るのは，めったに起こらない

と判断することができます。つまり，②の仮説が正しくないと判断されます。

よって，①は正しい，つまり，このコインは表が出る確率と裏が出る確率は等しくないと判断できます。

このように，得られたデータをもとに，あることがらが正しいかどうかを判断する方法を［　　　　］といいます。

仮説検定では，ある仮説のもとで，それが「めったに起こらないと判断する基準となる確率」をあらかじめ決めておき，実際に行った結果，そのめったに起こらない事態が起きたときは，その仮説を否定しようと考えます。この判断となる基準のことを［　　　　］（または**危険率**）といい，仮説を否定することを，その仮説を［　　　　］といいます。また，めったに起こらない事態が起きる範囲を［　　　　］といいます。

仮説検定をするが支持する仮説（正しいことを主張したい仮説）のことを［　　　　］，それに対して「おかしい」と疑っている仮説を［　　　　］といいます。

通常，仮説検定の手順は以下の通りです。

① 帰無仮説と対立仮説を設定する。

② 帰無仮説のもとで，対象となる統計量の分布を定める。

③ 有意水準を決め，②の分布において対立仮説に有利となる棄却域を設定する。

④ 標本を抽出し，②の統計量の値（実測値）が棄却域にあるかどうかを調べる。棄却域にあれば帰無仮説を棄却する。棄却域になければ帰無仮説を受容する。

帰無仮説が棄却されないとき，帰無仮説は「受容」（または採択）されたといいます。しかし，受容や採択といっても，以上の論調からわかるように，積極的な意味で肯定したわけではありません。

4．仮説検定

データ X が正規分布 $N(m,\ \sigma^2)$ に従うとき，

$$P(m-1.96\sigma \leqq m \leqq m+1.96\sigma)=0.95$$

でした。

したがって

$$P(X \leqq m-1.96\sigma,\ \ m+1.96\sigma \leqq X)=0.05$$

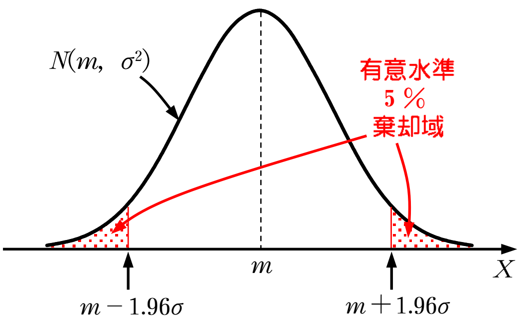
が成り立ちます。

まとめると

帰無仮説にもとづいたデータ X の母集団分布が正規分布 $N(m,\ \sigma^2)$ に従うとき，有意水準 5 %の棄却域は

$$X \leqq m-1.96\sigma,\ \ m+1.96\sigma \leqq X$$

となります！



問題３

ある工場で製造される製品の重さは，平均 500 g，標準偏差 16 g の正規分布に従うといいます。ある日，この製品 100 個を無作為に抽出して重さを調べたところ，平均値は 495 g でした。

この日の製品は異常であるといえますか。有意水準 5 %で仮説検定してみましょう。

Column 5．台風の予報円

日本列島に台風が近づくと，右の図のように，台風の予報円がテレビや新聞，インターネットの天気予報で示されます。**予報円**とは，ある時刻に，その中に台風の中心が入る確率が 70 ％であると予想される円ですが，なぜ予報円の中に台風の中心が入る確率は 70 ％となっているのでしょうか。

予報円は，数直線上で考える信頼度 70 ％の信頼区間を，平面上に広げたものと考えられます。信頼度を大きくすると信頼区間の幅が広がりますが，これと同じように，中心が入る確率を 70 ％より大きくすると，予報円はもっと大きくなります。

仮に，中心が入る確率を 90 ％にした予報円を考えると，その大きさは 70 ％のときの予報円の約 1.6 倍となり，上の図の予報円は本州の大部分と四国，九州のすべてを覆う大きな円となります。これでは，警戒するべき地域があまりに広くて不便です。

そこで，予報円があまり大きくならないようにするために，確率を 70 ％にしているのです。