

対象となる集団の一部のみを調査し、全体の様子を推測する方法がある。これを**標本調査**という。標本調査において、対象とする集団全体を**母集団**、母集団に属する個々のものを**個体**、個体の総数を**母集団の大きさ**という。

また、母集団から選び出された一部の個体の集合を**標本**、その中の個体の個数を**標本の大きさ**、標本を選び出すことを**抽出**という。そして、母集団から抽出した標本の平均を**標本平均**という。

中学校の学びを振り返ろう

果物の収穫時期を決める際、果物の糖度を調べることがある。糖度とは、果汁 100g に糖分がどのくらい含まれているかを表したものである。

サクランボ 300 個を母集団として、糖度の平均値を標本調査で推定してみよう。



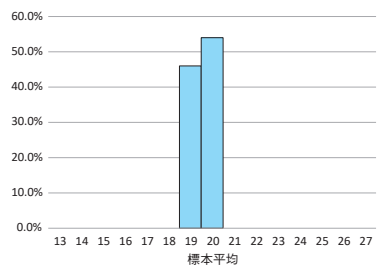
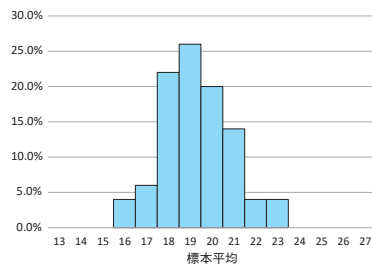
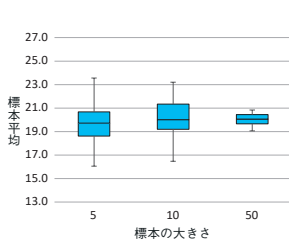
- ① 次のページの表は、サクランボ 300 個 (母集団) の糖度のデータである。この中から、10 個を無作為に抽出して、平均値を求めてみよう。さらに、300 個から 10 個を無作為に抽出することを 5 回繰り返してみよう。
- ② 標本の大きさを 5, 10, 50 にした場合のそれぞれについて、アプリケーションを利用して、無作為抽出をし、平均値を求めることを 50, 100, 300, 1000 回繰り返すシミュレーションをしてみよう。その結果の分布 (標本平均の分布) の箱ひげ図からどのようなことがわかるか。また、ヒストグラムからはどのようなことがわかるか。

● 気づいたこと

<サクランボ 300 個の糖度>

23.7	24.9	25.7	24.1	26.4	26.6	13.8	14.8	20.5	18.1
13.6	24.3	18.6	18.7	16.2	21.1	13.6	19.2	21.1	22.1
17.6	19.7	16.9	22.3	19.6	18.7	16.0	24.6	26.1	22.6
17.0	14.5	20.1	20.8	22.7	14.3	16.0	19.9	26.6	21.0
25.4	19.0	23.6	16.6	14.3	24.9	14.8	14.0	19.4	19.6
14.4	24.1	19.9	14.8	20.0	15.3	23.6	13.8	25.5	15.0
19.9	26.6	15.7	24.6	24.4	21.6	24.4	21.7	21.1	18.9
21.1	24.8	22.9	22.1	15.6	21.3	15.3	21.5	19.3	23.3
14.8	26.2	22.8	17.5	14.0	18.1	20.1	23.7	18.5	17.1
25.1	17.8	14.0	20.9	16.1	25.7	15.0	23.9	16.9	14.4
24.9	17.4	20.1	15.5	20.8	21.1	16.6	24.6	15.1	20.3
25.8	15.2	24.0	15.9	22.3	22.0	17.7	25.9	24.6	21.5
18.4	21.5	18.6	22.7	24.3	25.0	14.3	20.8	17.5	20.1
24.7	24.2	16.2	21.0	26.6	19.5	14.2	19.7	22.5	24.4
17.9	22.1	20.7	24.5	22.4	14.5	15.6	16.4	16.9	21.8
18.1	20.7	21.8	25.3	20.0	15.4	20.8	22.1	21.0	25.3
14.4	16.5	18.0	19.6	22.4	17.5	18.0	26.4	25.4	24.8
21.9	17.1	21.1	21.7	19.9	22.4	25.5	24.1	16.1	17.2
14.1	19.2	15.9	14.7	19.7	13.6	15.7	22.1	25.0	20.1
24.2	17.1	22.7	17.4	22.9	15.2	14.4	18.4	17.2	21.4
24.0	18.6	19.0	25.4	18.0	18.0	23.5	16.1	26.0	13.8
13.7	26.0	17.1	19.0	25.5	18.5	16.1	26.0	18.6	24.7
15.4	17.2	26.6	17.5	20.2	23.5	26.2	24.1	22.7	14.1
18.8	17.6	18.4	21.4	19.5	18.5	19.7	18.2	15.2	18.2
18.5	25.1	16.6	21.7	23.8	20.1	23.1	24.5	24.3	22.4
25.7	17.5	14.0	26.1	26.1	23.6	26.6	16.1	15.6	26.3
14.5	25.8	25.5	20.6	18.4	26.5	23.1	25.9	19.5	18.6
18.5	23.4	20.4	18.1	16.5	15.2	16.2	18.2	19.9	15.8
20.7	22.9	18.1	15.8	21.3	20.5	14.1	19.7	16.3	13.7
14.3	16.1	17.8	17.1	23.9	17.2	18.4	26.5	19.2	18.2

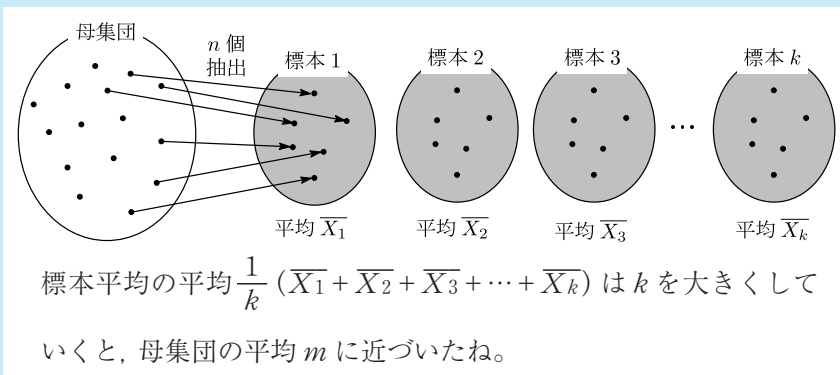
(このデータ全体の平均は 20.0 である。)



平均 m の母集団から、大きさ n の標本を無作為抽出するとき、 n を大きくしていくと、標本平均 \bar{X} は母集団の平均 m に近づく。これを **大数の法則** という。

母集団の平均、分散、標準偏差を、それぞれ **母平均**、**母分散**、**母標準偏差** という。

18 ページの ② で調べたヒストグラムは、これらの標本それぞれで平均（標本平均）を求め、その分布について考えたわけだね。



実際の標本調査では、通常、抽出を繰り返すことはしない。また、今回のサクランボの例では、母平均は 20.0 とわかっているが、本来は、その値を推測するために標本調査を行う。

そこで、1 回の抽出で得られた値をもとに、母平均がいくつからいくつの間にあるかを、正規分布を用いて推定する方法を考えよう。

方針 ① 標本平均 \bar{X} の分布（標本分布という）を正規分布とみなし、母平均 m がわかったとすると、たとえば、 \bar{X} が図 1 の $m - \alpha \leq \bar{X} \leq m + \alpha$ の区間に含まれる確率が 0.95 となる α を求めることができる。

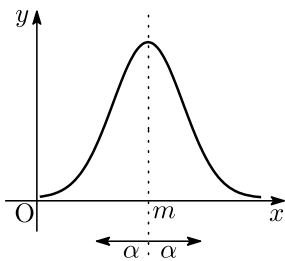


図 1

- ② 実際には m の値はわからず，わかるのは1回抽出したときの標本平均 \bar{X} である。その \bar{X} に対して， $\bar{X} - \alpha \leq m \leq \bar{X} + \alpha$ の区間を考えれば， m は0.95の確率でこの区間に含まれることになる。

この理由を図2をもとに説明してみよう。

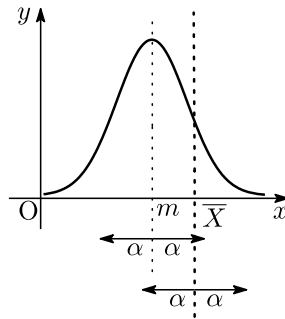


図2

● ②の説明

このような考え方を **区間推定** という。また，母平均 m をある確率で含む区間を **信頼区間** という。**方針** では，ある確率として0.95を用いていたので，この信頼区間のことを，母平均 m に対する **信頼度95%の信頼区間** という。

標本平均 \bar{X} がどのような正規分布に従うかはわからない。しかし，大数の法則より， n が大きいとき， \bar{X} は母平均 m に近づくことが知られているので，あと標準偏差がわかれば，母平均 m をある確率で含む区間を求められることになる。