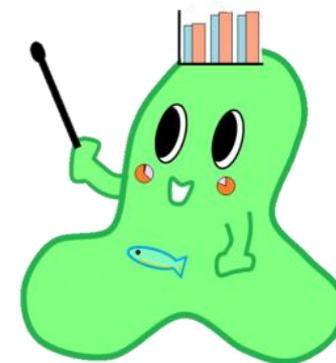
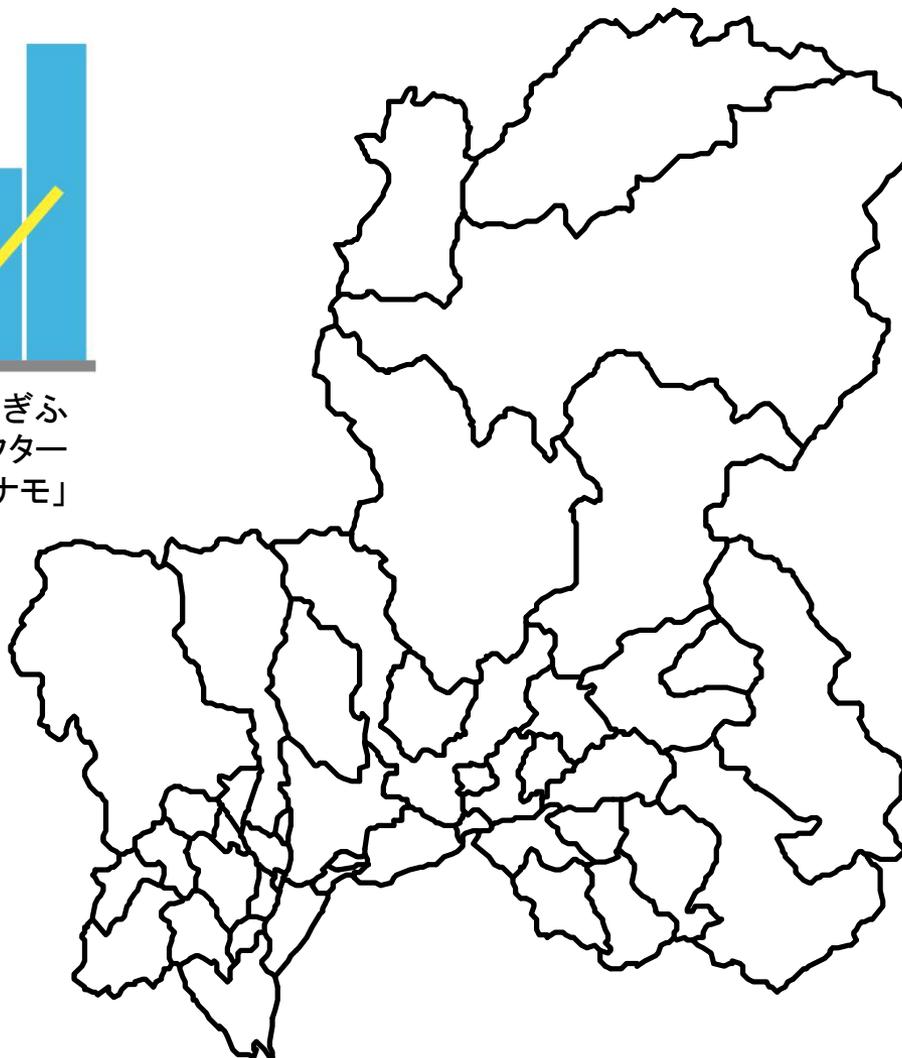
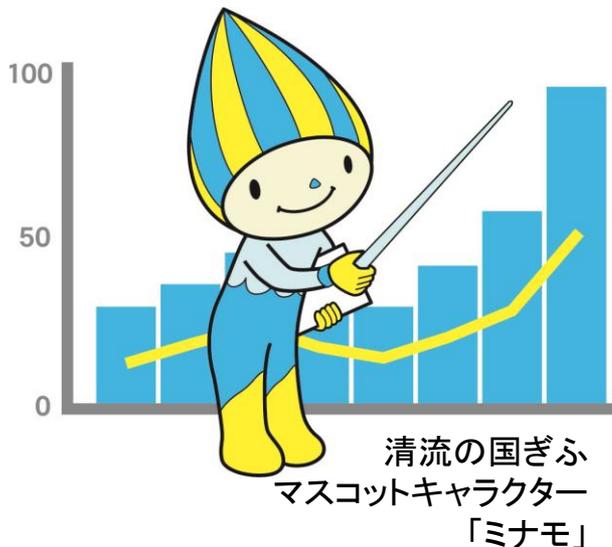


# データ分析のコツを知る～標本調査とは～



岐阜県統計課  
マスコットキャラクター  
「テルミー」  
2011/10/18生まれ

平成30年11月版  
岐阜県 環境生活部  
統計課 企画分析係

# 「集める」「分析する」「伝える」3つの力を 総合的な学習の時間で

## 集める

必要な情報を集めよう

HP・文献・聞き取り・標本調査など

## 分析する

傾向をつかもう

度数分布表・ヒストグラム・平均値・

偏差値・中央値・最頻値・ちらばりなど

## 伝える

効果的なグラフを活用して発信しよう

棒グラフ・円グラフ・折れ線グラフ・

レーダーチャート・散布図など

# 本日お話すること



## 集める

必要な情報を集めよう

① 標本調査って何？

## 分析する

傾向をつかもう

② 母集団全体の数量を推計しよう

## 伝える

効果的なグラフを活用して発信しよう

③ ふるさと岐阜県について知ろう

# 輪ゴムはいくつ入っているのかな？

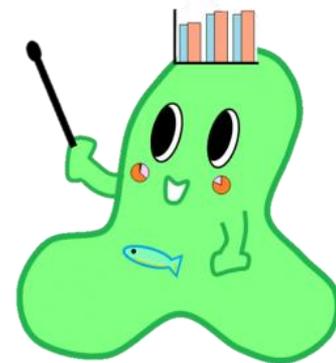


全部数えればわかるけど、  
数えるのはすごく大変だね。  
何かいい方法はないのかな？

# 形も大きさも同じで色だけ違う輪ゴムを使うよ



いいアイデアがあるよ！  
形も大きさも同じで色だけ違う  
ゴムを使えば、全部を数えなくて  
もわかるよ。この青い輪ゴムを  
100個使って考えよう。

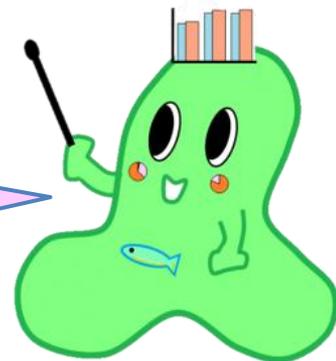


# 青い輪ゴムが、かたよりなく散らばるように混ぜる！



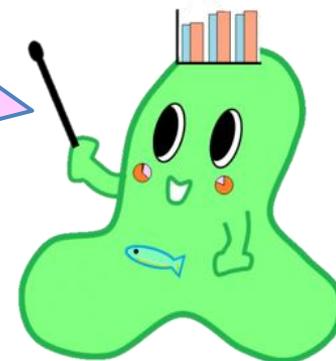
① 青色の輪ゴム100個を最初の輪ゴムの中に入れてよう。

② 青い輪ゴムがかたよりなく散らばるよう、「一握りして、パラパラ」を30回繰り返そう。



# 袋の中を見ないで、一握り取り出す！

③ よくシャッフルができれば、袋の中を見ないで一握り取り出そう。

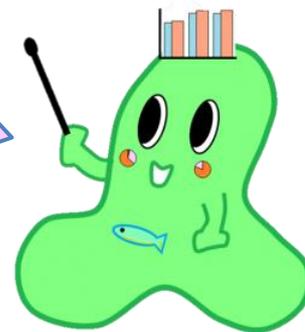


取り出した一握りのうち、普通の輪ゴムと青の輪ゴムがいくつあるかを調べればいいんだね。

# 班で手分けして、取り出した輪ゴムを数えよう！



④ 班で手分けして、普通の輪ゴムと青の輪ゴムの数を数えよう。



10個のかたまりが10個で100個

10のかたまりを作っていけば正確に数えられそうだね。

【今回の結果】

普通の輪ゴム	391 個
青の輪ゴム	22 個
合計	413 個



# 中学1年生で学習した「比」を活用するよ！

<知りたい数>

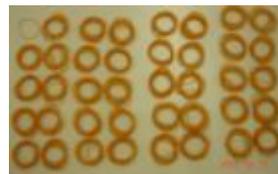


(普通のゴムの数)



(青ゴム総数)

=



(抽出普通ゴム)



(抽出青ゴム)



X

:



100

=



391

:

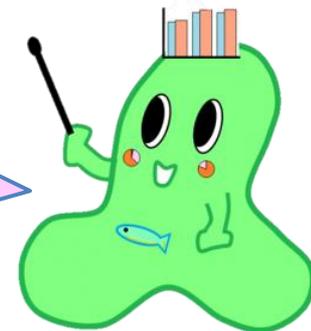


22

【Xの値を求めよう】

$$\begin{aligned}
 x : 100 &= 391 : 22 \\
 22x &= 39100 \\
 x &= 1777.27\cdots
 \end{aligned}$$

上から二桁の概数で考えると、約1800個とすることができるね。



# より正確に全体の輪ゴムの数を求めるには？

【ある班で 5 回実施したときの結果】

	1回目	2回目	3回目	4回目	5回目	平均
結果 (X)	1777	1740	1872	1868	1822	1816

平均を求めて小数点以下を四捨五入しよう。

$$\begin{aligned} & (1777 + 1740 + 1872 + 1868 + 1822) \div 5 \\ & = 9079 \div 5 \\ & = 1815.8 \end{aligned}$$



何回か試して、  
その結果を平均してみたよ。

それぞれの班で出てきた値を平均すれば、  
もっと正確な値に近づきそうだね。

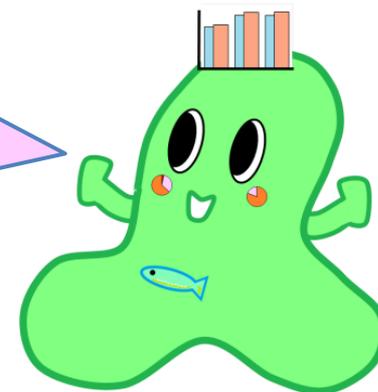


# 班での役割分担をして、最低3回実験するよ！

## 【役割分担】

ゴッドハンド係	1人	一握りの輪ゴムを握れる神の手をもつ人
シャッフル係	1人	青ゴムがかたよりなく散らばるようにさばける人
電卓係	1人	比の計算で素早く全体の数を算出できる頭脳をもつ人
数え係（青ゴム）	1人	青の輪ゴムを素早く見つけ出し数を数えることのできる人
数え係（普通ゴム）	4人	普通の輪ゴムを素早く10個数え、まとめることができる人
応援係	他の人	残念ながらどの役にも入れないけれど、熱い心のある人

記録を取るのは、電卓係にお願いして、他の皆さんは机の上は何も置かないようにしましょう。さあ、準備ができたところから、スタート！



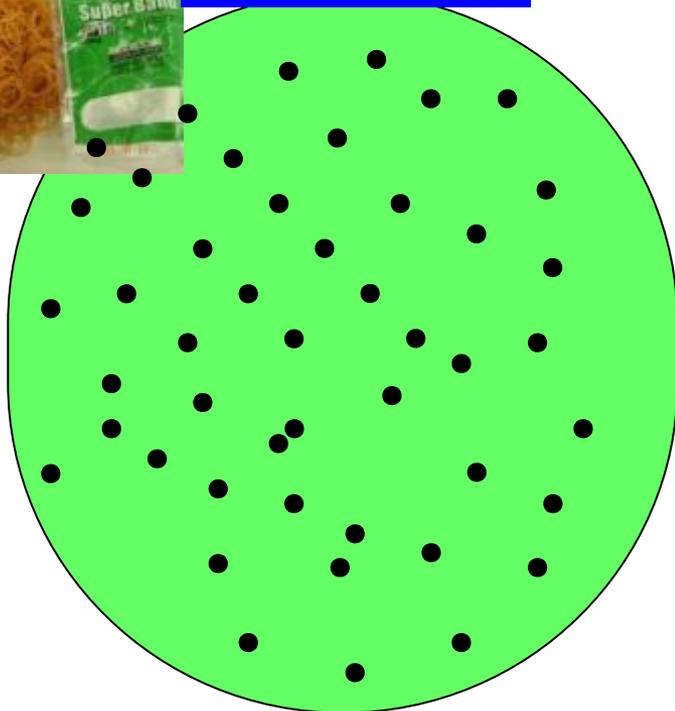
# 全数調査は、時間と手間がかかるよ！

## 全数調査

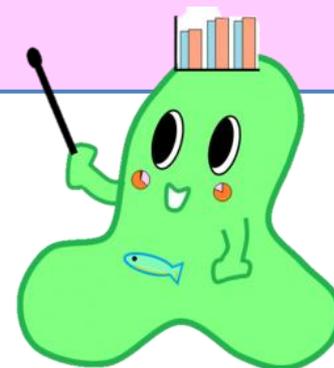
集団のもっている性質を調べるために、その**集団をつくっているもの全部**について行う調査



## 母集団



一部だけではなく、全部を調べる調査を、「**全数調査**」というよ。  
全て調べないといけないから、大変な調査になるね。



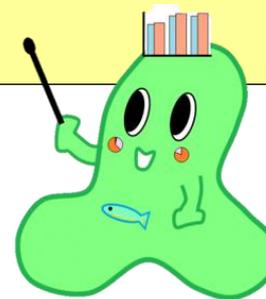
統計調査には、  
どんな全数調査があるのかな？

# 「国勢調査」は、今の日本の姿を知る大切な全数調査

## 国勢調査

5年ごとの10月1日現在、日本国内に住んでいるすべての人及び世帯について調査を行います。

国や県、市町村は、この調査結果をもとに**みなさんの生活をよりよくしていく政策**を考えていくよ。



この調査をするために、全国で**約70万人**の国勢調査員、さらに、調査員を指導する**約10万人**の国勢調査指導員が携わっているよ。

前回調査は平成27年。みんなのお家にも調査票が届いているよ。

秘 基幹統計調査 国勢調査調査票

平成27年10月1日 総務省統計局

数字は右づめに記入は黒の鉛筆で

数字の記入例 1 2 3 4 5 6 7 8 9 0

○黒の鉛筆で記入し、間違えた場合は、消しゴムできれいに消してください。  
○記入欄が○の場合は、当てはまる○を●のようにぬりつぶしてください。  
○数字を記入する場合は、下の例のようにわくの中に右づめで書いてください。たて線1本、すきまをあける、としる、はねない、上につきぬける、角をつける

世帯について (調査票が2枚以上にわたる場合は1枚目のみに記入してください)

1 世帯員の数  
ふだん住んでいる人 総数 男 女  
全員の人数を書いてください

2 住居の種類  
持ち家 都道府県、市区町村等の職員住宅 都市再生機構、公社等の職員住宅 民営の賃貸住宅 給与住宅(社員住宅) 住宅に隣接する独立型・寄居舎 会社等の寄居舎 その他

世帯員全員について (世帯員ごとに記入してください)

3 氏名及び男女の別  
ふだん住んでいる人を、もれなく書いてください

4 世帯主との続き柄  
世帯主の配偶者(妻又は夫)の祖父母・兄弟姉妹はそれぞれ祖父母・兄弟姉妹に含めます  
孫の配偶者は孫に、兄弟姉妹の配偶者は兄弟姉妹に含めます

5 出生の年月  
該当する元号又は西暦に記入したうえで、年及び月を書いてください  
年を西暦で記入する場合は西暦年の4桁を書いて

「調査票の記入のしかたを参照して、黒い太わくの」

(平成27年10月1日に行われた国勢調査の調査票)



# 今回、輪ゴムで行った調査は、標本調査。

## 全数調査

集団のもっている性質を調べるために、その**集団をつくっているもの全部**について行う調査



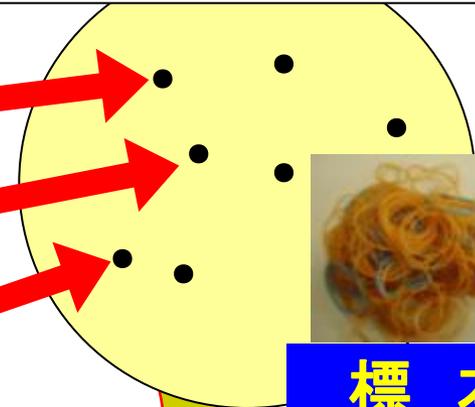
母集団

## 標本調査

集団の一部分について調べて、その結果から**元の集団の性質を推定する**調査

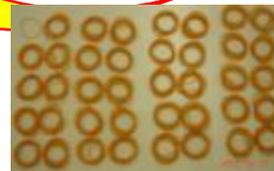


抽出



標本

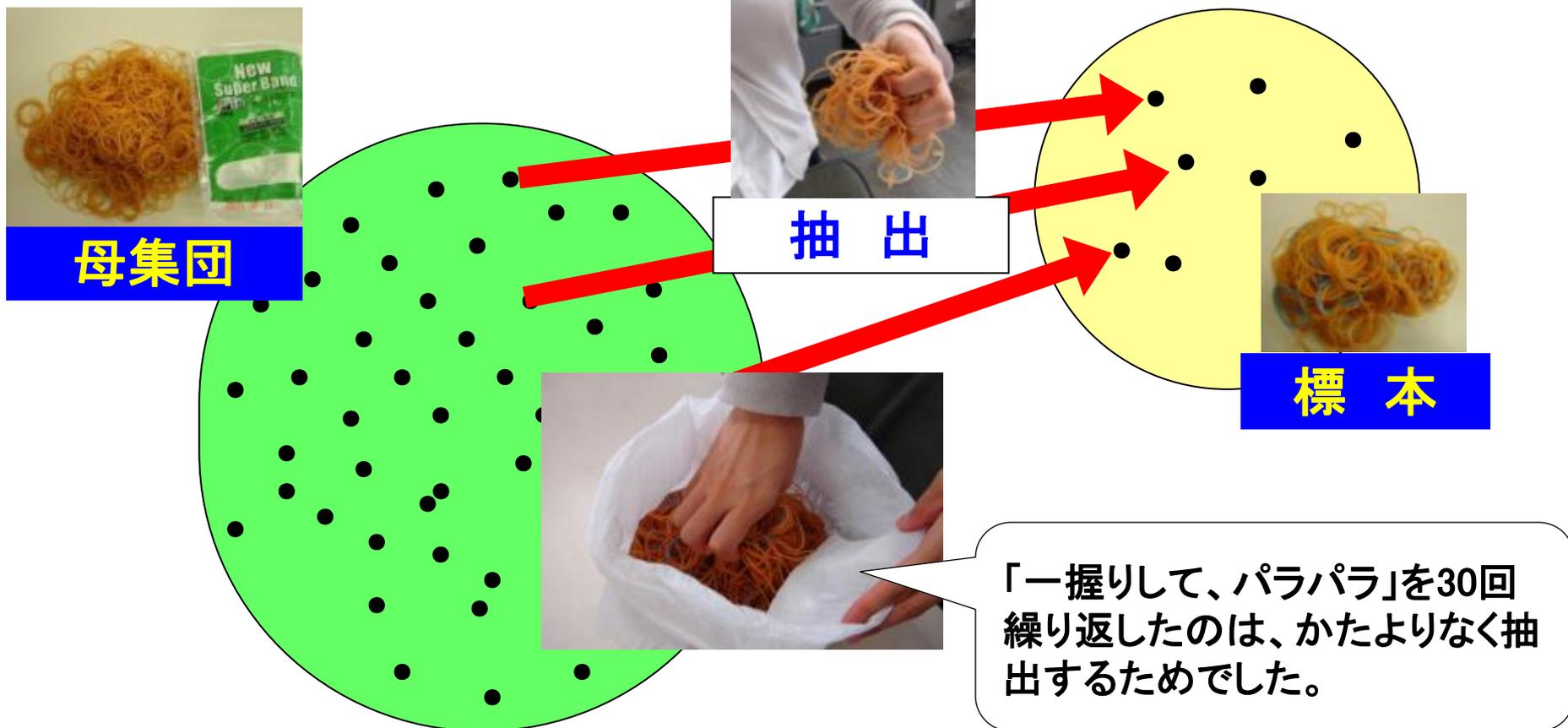
推定



# 標本調査では、かたよりなく抽出することが大切！

## 無作為抽出

母集団から標本を抽出する時、有意の方法で抽出を行わないという意味。簡単に言えば、**くじ引きと同じ**です。



# ゴッドハンド役さん、なぜ「一握り」だったの？



もし抽出数を  
10に減らしたら？

逆に抽出数を  
1500に増やしたら？

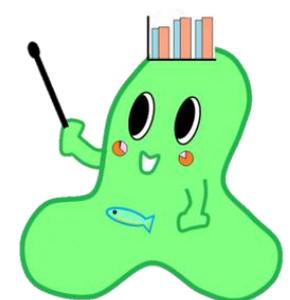
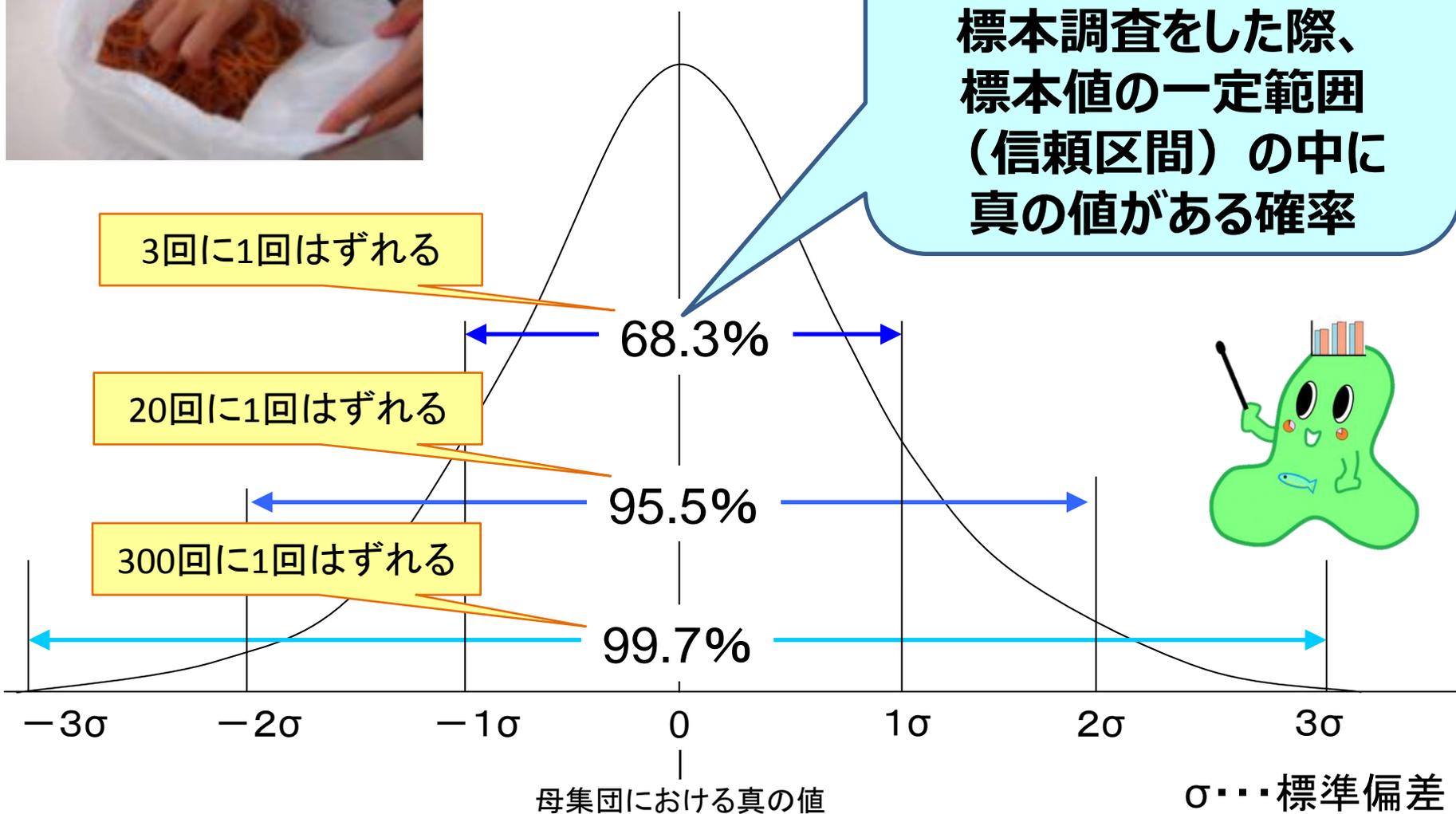


一握りで約400個の輪ゴムがとれたね。  
抽出数を多くしていけば全数調査に近づいていくから、より正確な値を推計できるけど、多すぎても調査が大変だ・・・。

# 実は、400は理論式から求めた数だよ



正規分布のグラフ



# 実は、400は理論式から求めた数だよ

標本数はどれだけ必要なの？ ～そもそも**標本調査**とは～

標本数を求める式（ $n$  = 標本数）

$$n = \frac{\lambda^2 \sigma^2}{d^2}$$

$\lambda$ : 信頼水準によって定まる値

何%の信頼水準(確からしさ)を求めるか(設計者が決める)によって、 $\lambda$ の値が決まる。

$\sigma$ : 母集団の標準偏差

過去の調査または類似調査から決まる値

$d$ : 標本誤差

どの程度の誤差まで許容するか、設計者が決める値

# 実は、400は理論式から求めた数だよ

## 標本数はどれだけ必要なの？ ～そもそも**標本調査**とは～

- ①どの程度の信頼水準で、どの程度の標本誤差dを許容する（又は目標とするか）を決める。
- ②信頼水準から、標本誤差dに対する標準誤差の倍率 $\lambda$ が決まる。

信頼水準	68.3%	95.5%	99.7%
$\lambda$	1倍	2倍	3倍

95%	99%
1.96倍	2.58倍

なお、**95%のとき、大雑把に $\lambda=2$** とすることが多い。

- ③標本誤差dと $\lambda$ から標準誤差 $\sigma_{\bar{x}}$ が次の式で決まるので、

$$d = \lambda \sigma_{\bar{x}} \rightarrow \sigma_{\bar{x}} = \frac{d}{\lambda}$$

$$n = \frac{\sigma^2}{\sigma_{\bar{x}}^2} \text{に代入すると、}$$

$$n = \frac{\lambda^2 \sigma^2}{d^2}$$

# 実は、400は理論式から求めた数だよ

簡単に言うと、

信頼水準と標本誤差を設定すると、標本数が決定される。

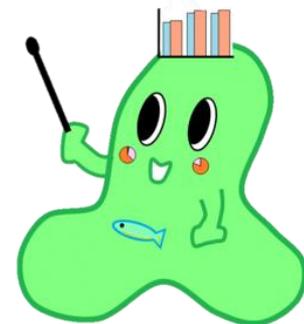
<信頼水準>

一般の統計調査では95%レベル、医薬品の効果の検定などでは99%レベルが用いられることが多い。

<標本誤差>

調査の目的、費用等に基づき、過去の同様の調査を参考として調査企画者が設定する。

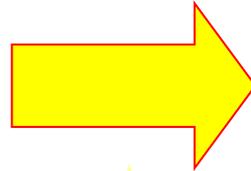
$$n = \frac{\lambda^2 \sigma^2}{d^2}$$



# 実は、400は理論式から求めた数だよ

この式を、さらに分解すると・・・

$$n = \frac{\lambda^2 \sigma^2}{d^2}$$



$$n = \frac{\lambda^2 p(1-p)}{d^2}$$

※p: 母比率

## 比率で考えてみると

母集団の大きさをN、条件に合う調査単位を $N_1$ とすると、条件に合わない調査単位数は、 $N_0 = N - N_1$ となり、母集団の比率Pは、 $N_1 / N$ と表せる。

$$\begin{aligned}\sigma^2 &= \frac{\sum_{i=1}^N (x_i - p)^2}{N} \\ &= \frac{1}{N} \left\{ N_1(1-p)^2 + N_0(0-p)^2 \right\} \\ &= p(1-p)^2 + (1-p)p^2 \\ &= p(1-p)\end{aligned}$$

## 実は、400は理論式から求めた数だよ

標本の数が400あればと言われるのはどうして？

標本誤差dを5%、 $\lambda=2$ 、 $p=0.5$ と考えて、代入すると、

$$\begin{aligned}n &= \frac{\lambda^2 p(1-p)}{d^2} = \frac{2^2 \times 0.5(1-0.5)}{d^2} = \frac{4 \times 0.25}{d^2} = \frac{1}{d^2} \\ &= \frac{1}{0.05^2} = \frac{1}{0.0025} = 400\end{aligned}$$

標本誤差を設定することによって、標本数を決定することができる。  
(ただし、信頼水準95%で $\lambda=2$ 、 $p=0.5$ とする。)

標本調査をするときや、標本調査による統計データをみるときは、標本数や抽出の方法についても注意してみてね。

