

<平成23年10月20日 大和中学校>

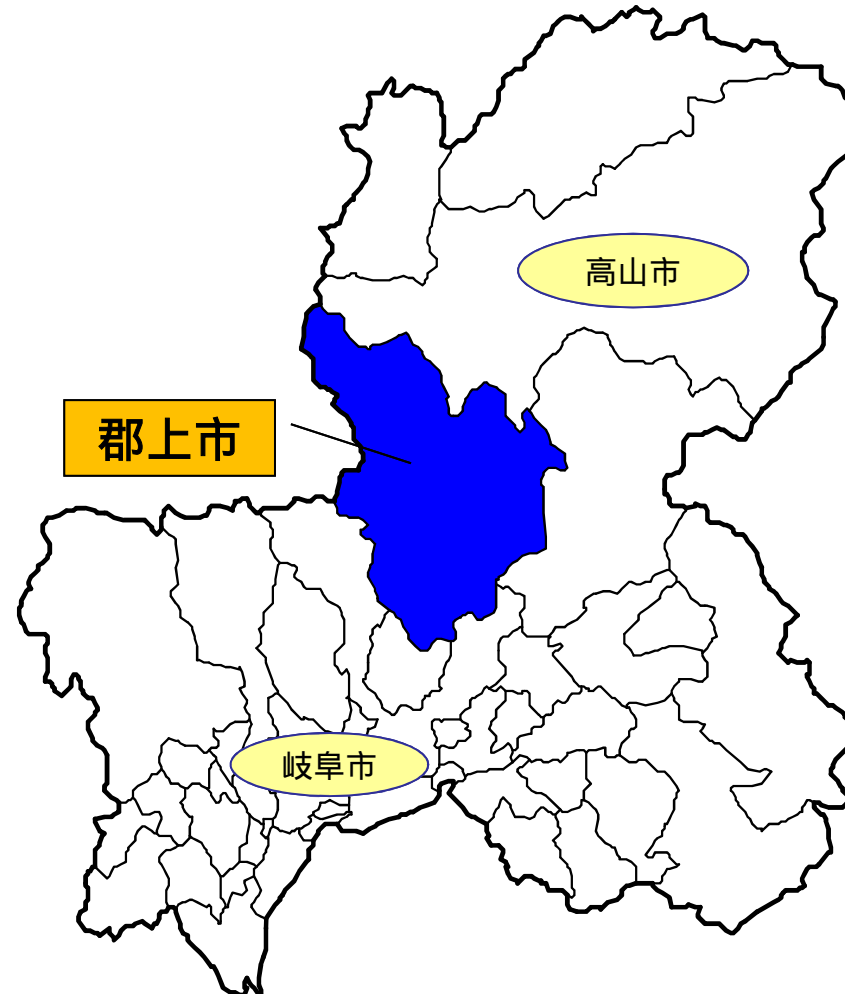
# データ分析マスター講座

～ふるさと岐阜県・郡上市(大和町)を知る～



テルミーちゃん

2011/10/18生まれ



平成23年10月  
岐阜県 総合企画部  
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

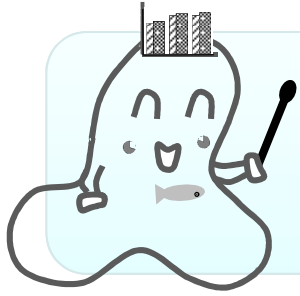
傾向をつかもう

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



# これからお話すること

集める

必要な情報を集めよう

標本調査って何？

分析する

傾向をつかもう

母集団全体の数量を推計しよう

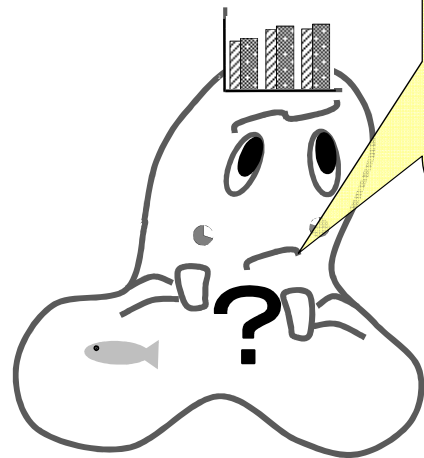
伝える

効果的なグラフを活用して発信しよう

ふるさと岐阜県・郡上市について知ろう

要介護認定者数など

# 輪ゴムはいくつ入っているのかな？

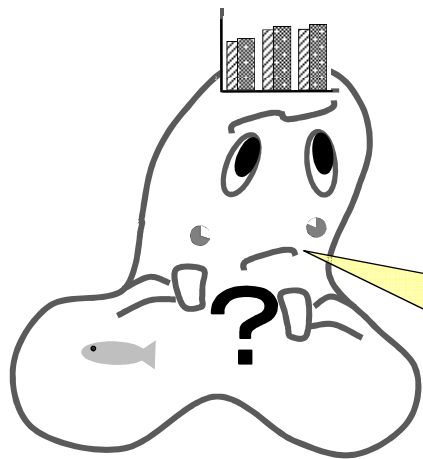


数えればわかるけど、数えるのはすごく大変だ。何かいい方法はないのかな？

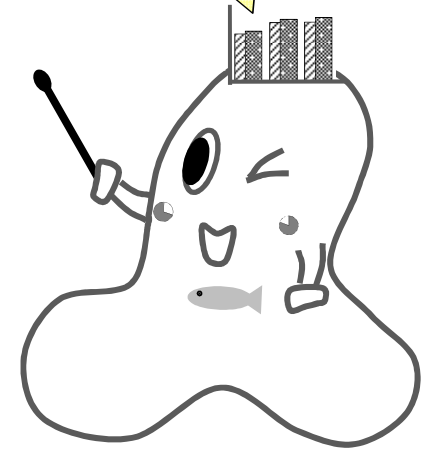
# 形も大きさも同じで色だけ違う輪ゴムを使うよ



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  
もできるよ。この青い輪ゴムを  
100個使って考えよう。



そんなことできるの？  
どうすればいいのかな？

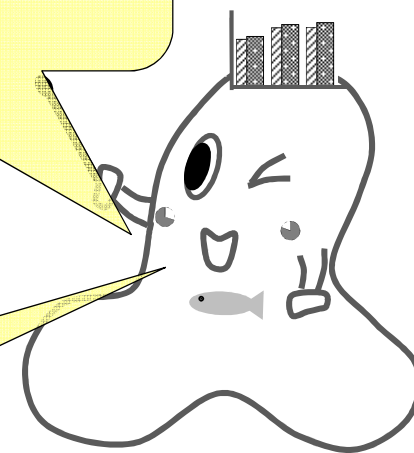


青い輪ゴムが、かたよりなく散らばるように！



青色の輪ゴム100個を  
先ほどの輪ゴムの中に入  
れましょう。

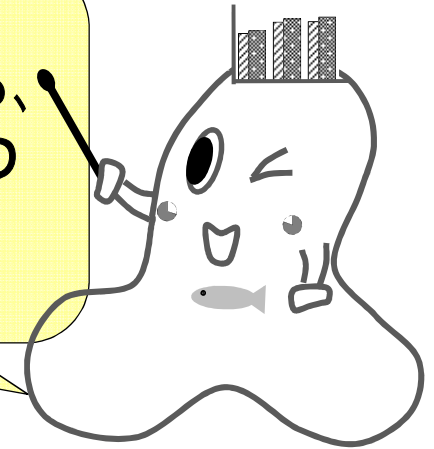
青い輪ゴムが、かたよりなく散ら  
ばるように「一握りして、パラパラ」を  
30回繰り返しましょう。



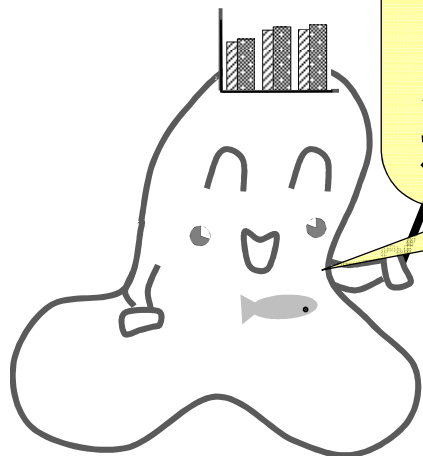
袋の中を見ないで、一握り取り出しましょう！



よくシャッフルができたなら、袋の中を見ないで一握りつかみましょう。



取り出した一握りのうち、普通と青の輪ゴムがいくつあるかを調べればいいんだ。



# 班で手分けして、取り出した輪ゴムを数えよう！



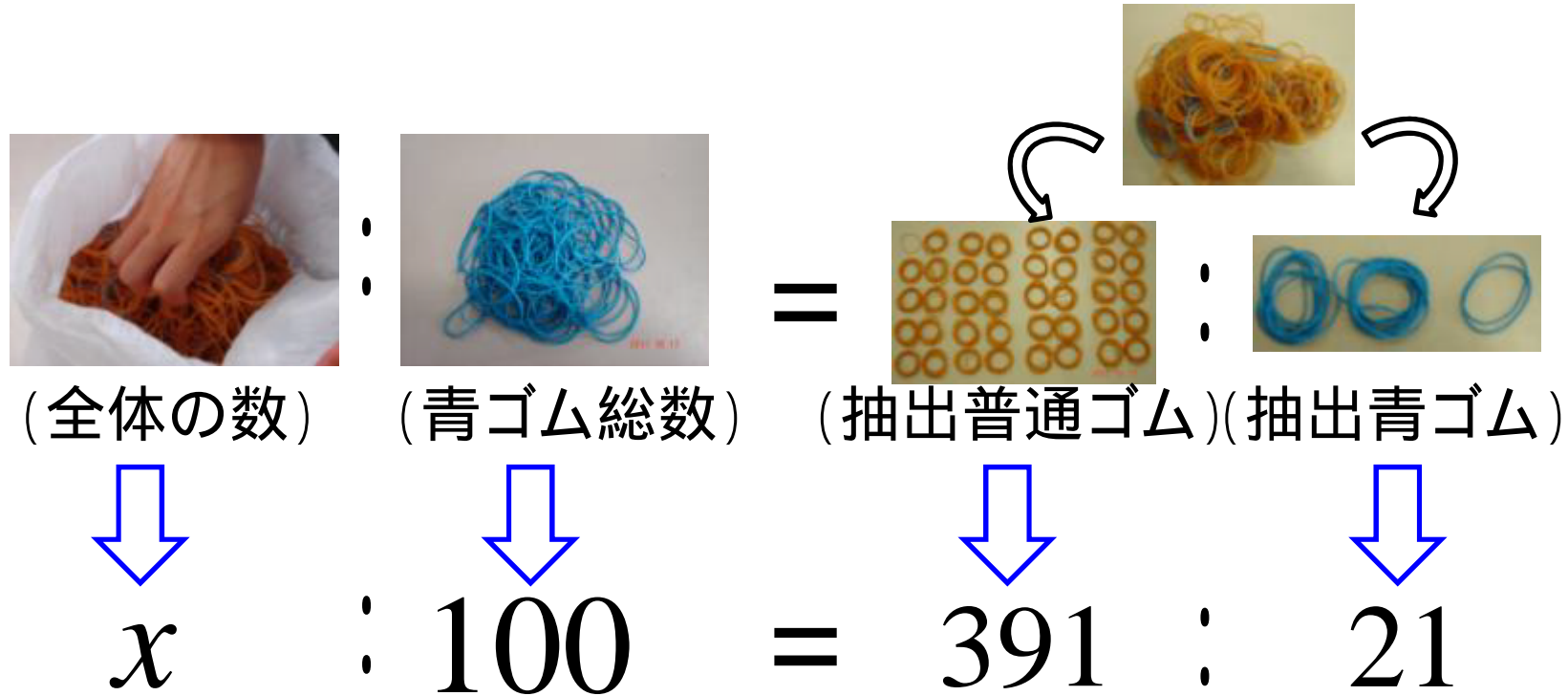
班で手分けして、青の輪ゴムと普通の輪ゴムの数を数えましょう。10のかたまりを作っていくと間違いがすくないよ。



10個のかたまりが10個で100個

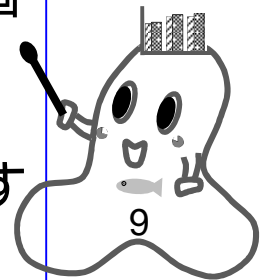


# 1年生で学習した「比」の学習を活用するよ！



$$\begin{aligned}
 x : 100 &= 391 : 22 \\
 22x &= 39100 \\
 x &= 1777. \dots
 \end{aligned}$$

上から二桁の概数で考えると1800個と見ることができます。  
 ただし、青ゴムを100個追加している  
 ので**100個分を引いて1700個と推計**することができます。



# より正確に全体の輪ゴムの数を求めるには？

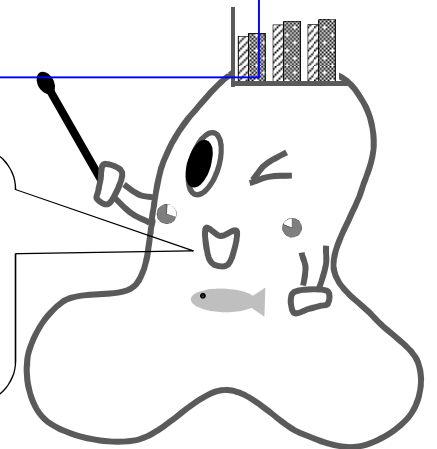
	1回目
普通輪ゴム	391
青輪ゴム	22
計算結果	1777

小数第1位を四捨五入しましょう。

5回の平均を求めましょう。と同じく小数第1位を四捨五入しましょう。

$$\begin{aligned} & (1778 + 1826 + 1859 + 1738 + 1896) \div 5 \\ & = 9097 \div 5 \\ & = 1819 \end{aligned}$$

4つの班で出てきた平均値を学級で平均すればもっと正確な値に近づきますね。

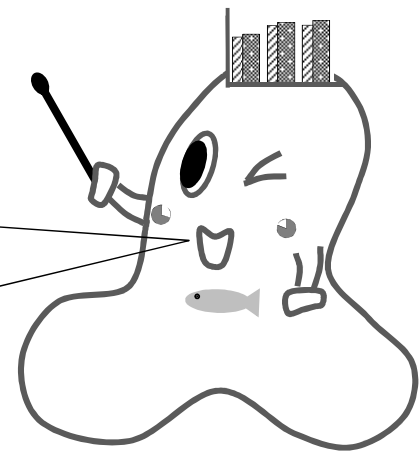


# 班での役割分担をして最低3回実験するよ！

## 【役割分担】

- ・ゴッドハンド役(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
- ・数え役青(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
- ・数え役(4人)・・・普通の輪ゴムを素早く10個数え固まりにできる人
- ・応援役(他)・・・残念ながらどの役にも入れない人は、応援係

記録を取るのは、電卓係にお願いして、他の生徒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！



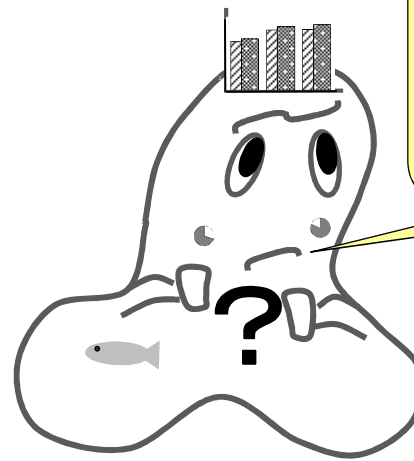
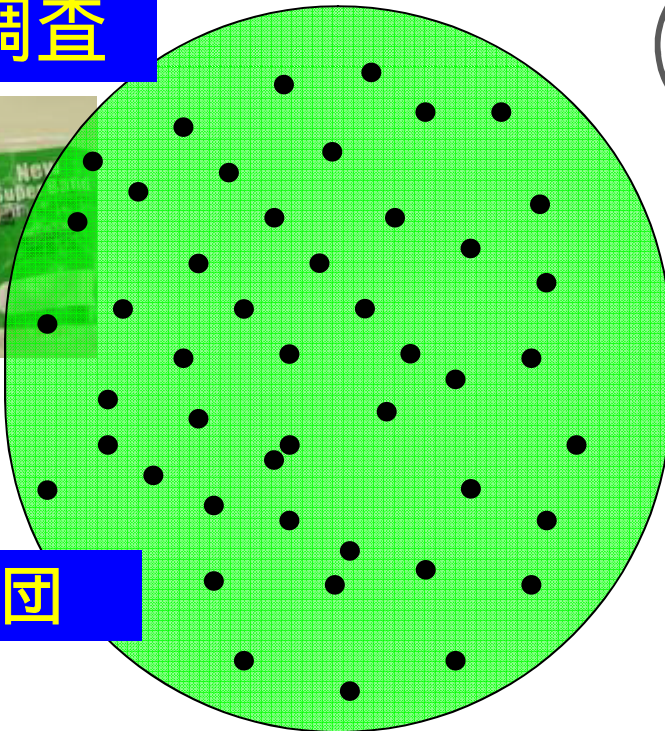
# 全数調査は、時間と手間がかかるよ！

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

全数調査

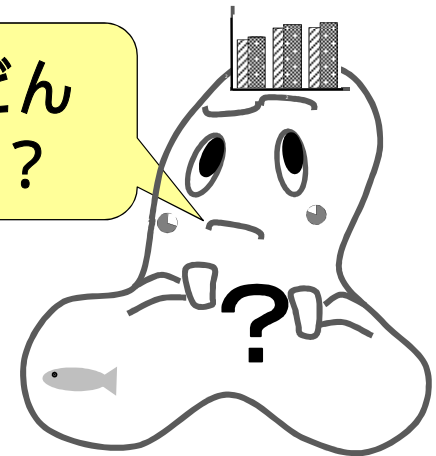


母集団



一部だけではなく  
全数を調べないと  
いけないんだよね。

全数調査にはどん  
なものがあるの？



# 今の日本の姿を知る大切な調査

**国勢調査**は、5年毎の10月1日現在、日本国内にふだん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**を考えていきます。

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ます。

基礎統計調査 国勢調査調査票 平成22年10月1日 総務省統計局

記入は  
黒の鉛筆で  
数字の  
記入欄  
1 2 3 4 5 6 7 8 9 0

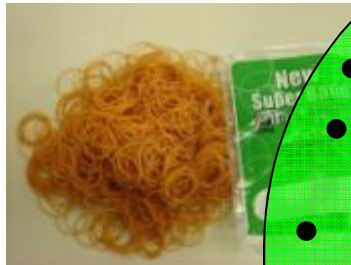
氏名及び男女の別	1		2		3		4	
	男	女	男	女	男	女	男	女
2 世帯主との続き柄	世帯主 本人	世帯主 本人	世帯主 本人	世帯主 本人	世帯主 本人	世帯主 本人	世帯主 本人	世帯主 本人
3 出生の年月	昭和	大正	昭和	大正	昭和	大正	昭和	大正
4 配偶者の有無	あり	なし	あり	なし	あり	なし	あり	なし
5 国籍	日本	外国	日本	外国	日本	外国	日本	外国

(平成22年10月1日に行われた国勢調査の調査票)

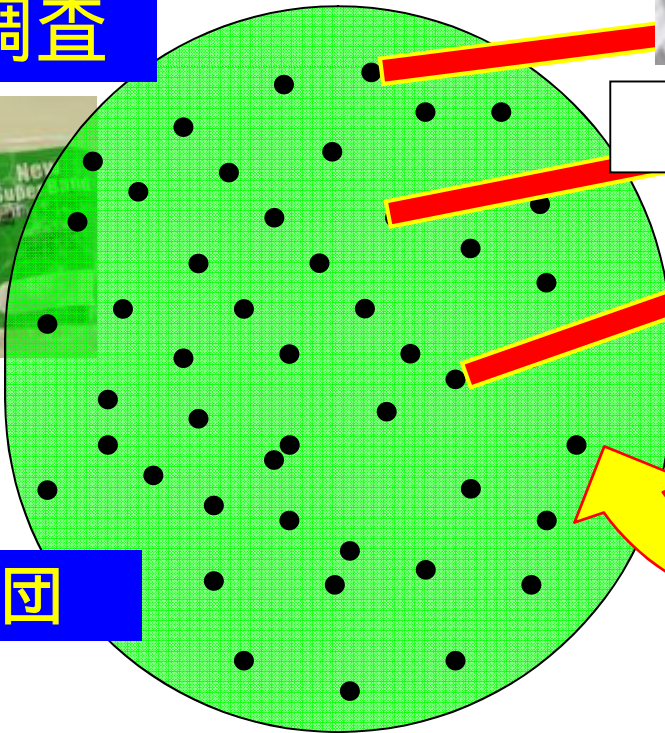
# 全数調査は、時間と手間がかかるよ！

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

全数調査



母集団

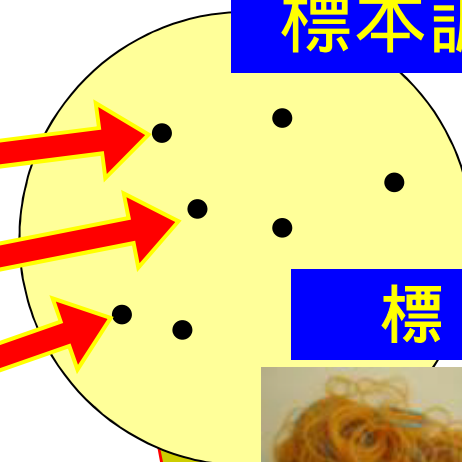


集団の一部分について調べて、その結果から**元の集団の性質を推定する**調査



抽出

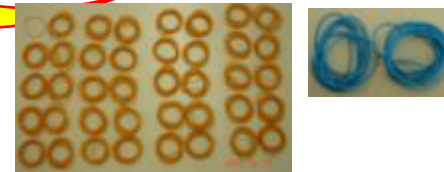
標本調査



標本



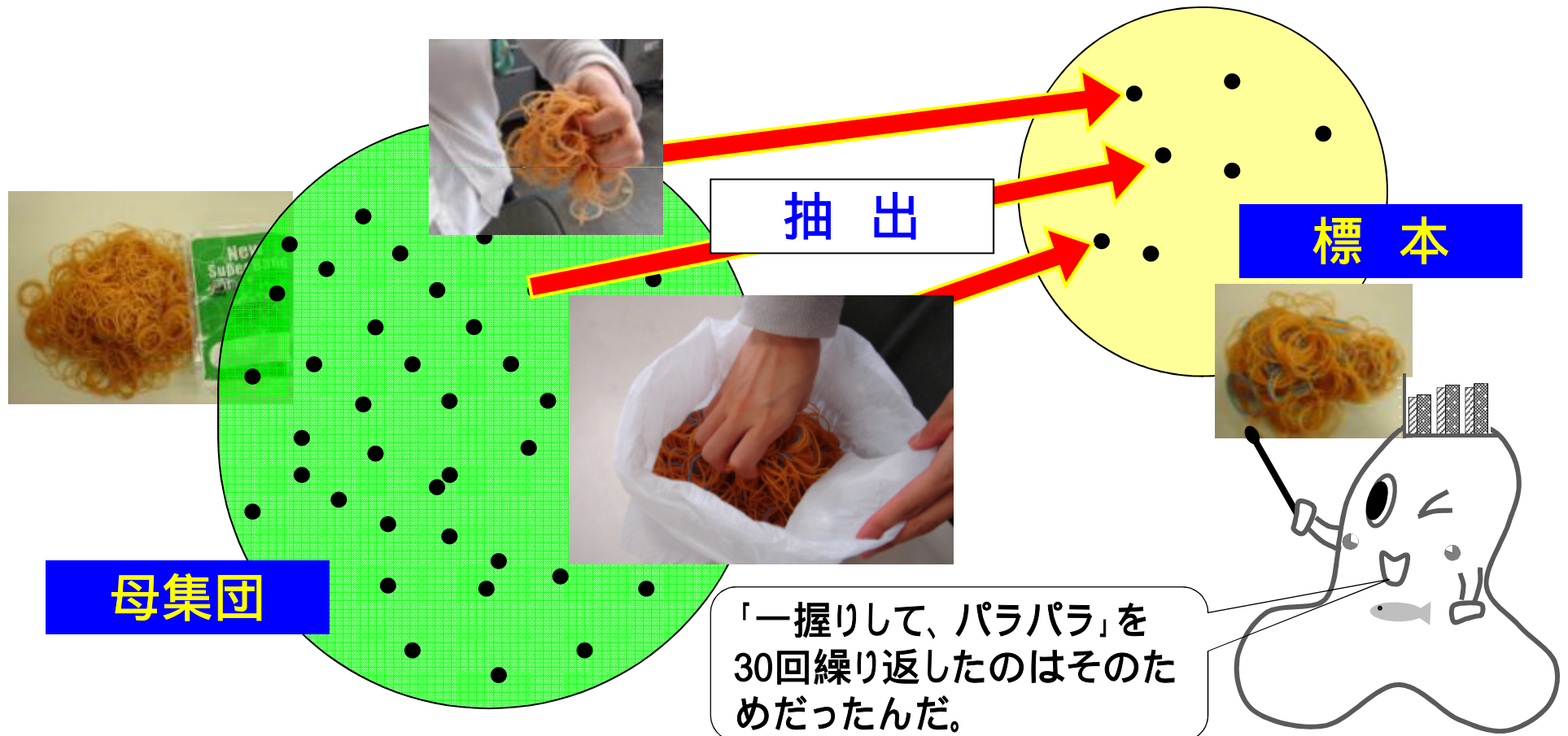
推定



# 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。

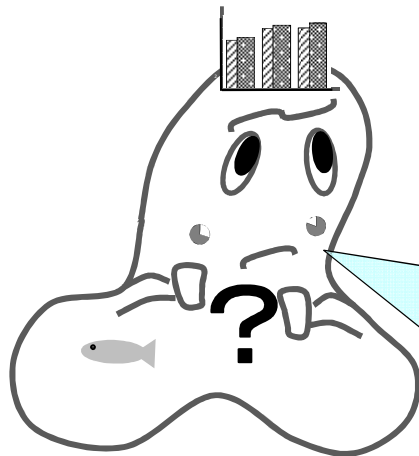
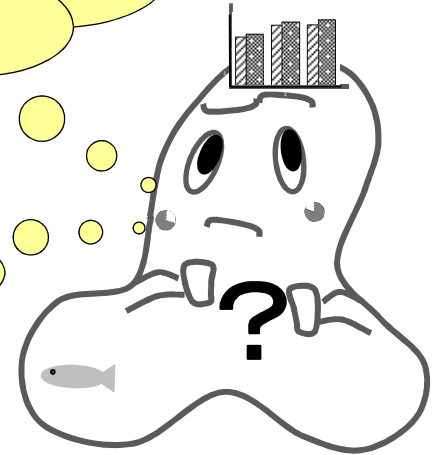


# ゴッドハンド役さん、なぜ一握りだったの？



もし抽出数を  
10にしたら？

逆に抽出数を  
1500にしたら？



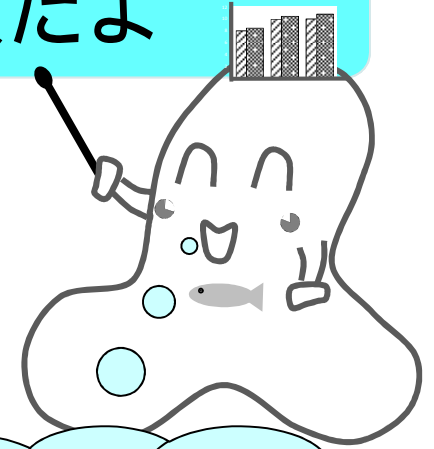
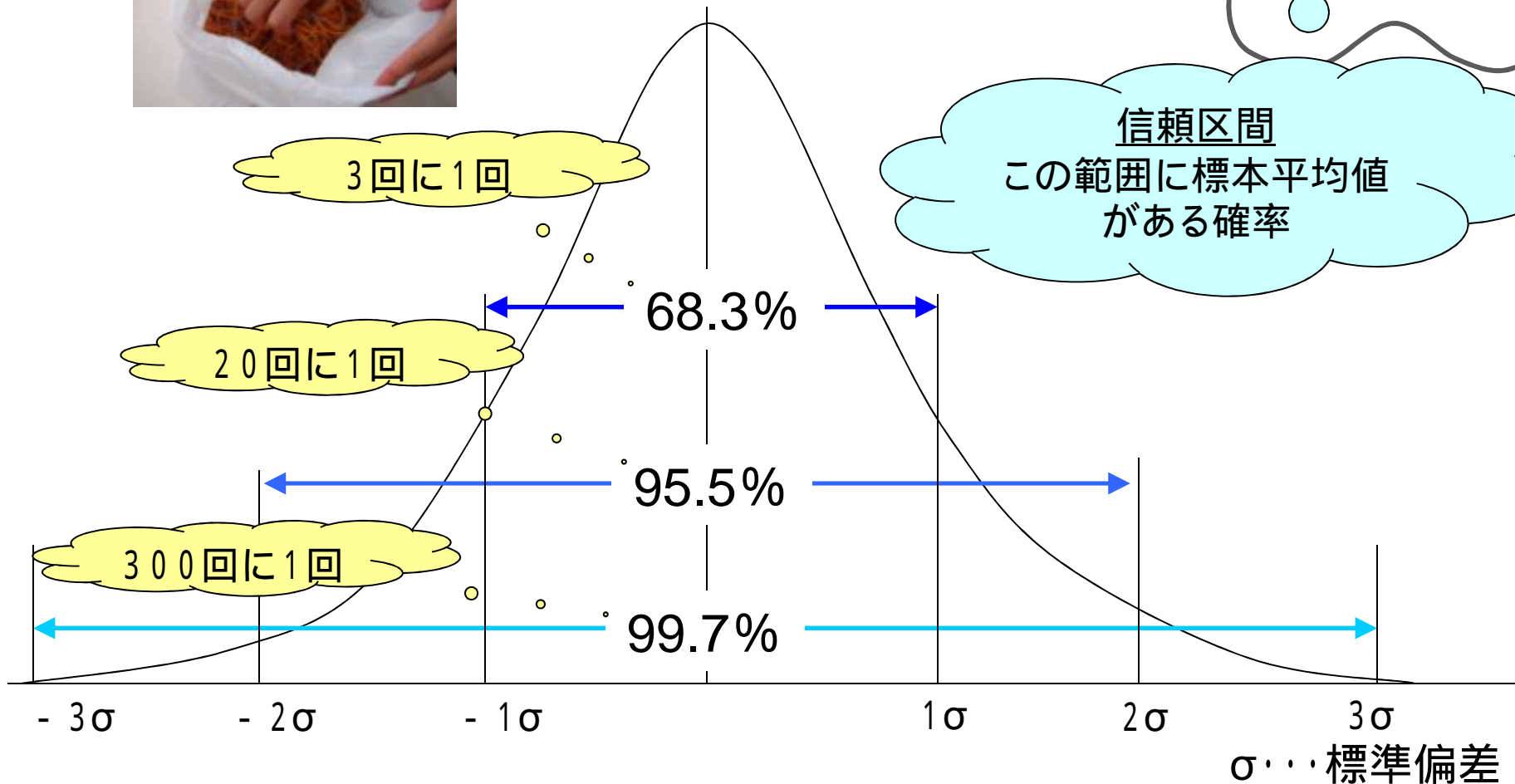
抽出数を多くしていけば、全数調査に近づいていくのでより正確な値を推計できるよね。  
だけど、多すぎても…。



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~

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

(nは標本数)

$\lambda$ : 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

過去の調査から、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~

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。

信頼水準から、標本誤差dに対する標準誤差の倍率λが決まる。

信頼水準	68.3%	95.5%	99.7%
λ	1倍	2倍	3倍

95%	99%
1.96倍	2.58倍

なお、95%のとき、大雑把にλ=2とすることが多い。

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

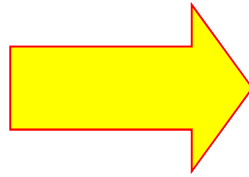
$$d = \lambda \sigma_{\bar{x}} \rightarrow \sigma_{\bar{x}} = \frac{d}{\lambda}$$

$$n = \frac{\sigma^2}{\sigma_{\bar{x}}^2}$$

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

実は、400は理論式から求めた数だよ

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



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

比率が分からない場合は、 $p=0.5$ として計算する。

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、母集団の比率は $P = N_1 / N$ と表せる。

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

実は、400は理論式から求めた数だよ

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。

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dについては、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。

もっと簡単に考えると、信頼水準95%で $\lambda = 2$ とし、さらに $p = 0.5$ として考えると、

$$n = \frac{\lambda^2 p(1-p)}{d^2} = \frac{2^2 \times 0.5(1-0.5)}{d^2} = \frac{4 \times 0.25}{d^2} = \frac{1}{d^2}$$

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？

標本誤差を5%、 $\lambda = 2$ と考えると、分母に0.05を代入すると、

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

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。)