

割合の近似値がいくつもある場合に 共通分母を探索するコマンド

アンケート調査や疫学報告において簡便な
記載ミス発見とサンプルサイズ推定のために

2022年10月12日(金) 下野寿之



1. 問題提起

割合の数を表す近似が、いくつか与えられた場合に、分母を推定したいことがある。
まず、その具体例のいくつかと問題意識を提起する。

素朴な疑問(1/2)

<https://www.mhlw.go.jp/content/000940528.pdf>
東京都のワクチン接種会場で配布資料の一部

■ 接種後7日間に現れた症状

報告割合	1回目接種後	2回目接種後	3回目接種後
50%以上		圧痛 65.2%	圧痛 81.4% 疲労 63.3% 疼痛 54.6% 筋肉痛 51.0%
10 - 50%	圧痛 48.2% 疼痛 26.9% 疲労 23.1% 頭痛 21.6% 筋肉痛 20.0% 倦怠感 12.2%	疼痛 45.6% 疲労 35.6% 筋肉痛 30.8% 頭痛 29.6% 倦怠感 26.4% 関節痛 14.8%	倦怠感 46.9% 頭痛 45.9% 関節痛 28.6% 発熱 17.3% 恶心・嘔吐 13.3% 腫脹・硬結 11.3% 紅斑 10.3%
1 - 10%	関節痛 6.7% 恶心・嘔吐 5.9% 発熱 2.4%	恶心・嘔吐 7.2% 腫脹・硬結 5.6% 紅斑 4.8% 発熱 4.4%	

(注) 対象:1回目接種後253人、2回目接種後250人、3回目接種後97人

出典:審査報告書より改編

数値が10個程度以上あれば、
確認作業は容易ではなく、
よく記載ミスは発生する。

上記の割合は正しく記載されているか?
間違いがあるとしたら、 ありうる理由は?

素朴な疑問(2/2)

<https://nlab.itmedia.co.jp/nl/articles/1812/12/news047.html>
バンダイのお年玉についての記事(2018年)

小学生 TOP5			中学生 TOP5		
1位	貯金	39.9%	1位	貯金	33.7%
2位	ゲーム機・ゲームソフト	30.4%	2位	文房具、雑貨	31.8%
3位	おもちゃ、カードゲーム(ゲーム機・ゲームソフト除く)	29.5%	3位	衣服・衣類雑貨	23.5%
4位	書籍(マンガ以外)	18.0%	4位	ゲーム機・ゲームソフト	22.7%
5位	お菓子やジュースなどの飲食物	17.2%	5位	マンガ	22.0%

男子 TOP5 (小中学生)			女子 TOP5 (小中学生)		
1位	ゲーム機・ゲームソフト	37.5%	1位	貯金	39.4%
2位	貯金	36.2%	2位	文房具、雑貨	35.7%
3位	おもちゃ、カードゲーム(ゲーム機・ゲームソフト除く)	27.3%	3位	書籍(マンガ以外)	21.5%
4位	お菓子やジュースなどの飲食物	18.8%	4位	衣服・衣類雑貨	20.7%
5位	マンガ	18.5%	5位	お菓子やジュースなどの飲食物	18.4%

複数回答 有効回答726人

どれかの内訳の人数が少なすぎて、その場合に信頼性の低い統計値になっていないだろうか?

「有効回答人数は726人」というヒントあり。
小学生、中学生、男子、女子のそれぞれの人数は?

引き続く疑問

1. 複数の割合近似値から共通分母の値を求めたい。
2. 共通分母の候補値は、どんな現れ方をするか？
3. そこにある未知の数理的な性質はないか？
 - 何桁の近似値が何個揃えば、分母を推定しやすいか？
4. あらうる手入力ミス等への良い対応の仕方は？
5. 上記の解決策に対して、有用性の確認方法は？





2. プログラムの設計

割合の数を表す近似が複数ある場合に、
その分母の候補値を算出するプログラムを作ったので、
そのプログラムの設計について解説する。

計算原理

1. 近似値xを、実数上の半区間と見なす。
 - xは小数点以下d桁まで記載されていたとする。
 - その半区間の幅は 0.1^d (0.1のd乗)となる。
2. 元の数は
 - $[x, x + 0.1^d)$: 切り捨ての場合
 - $(x - 0.1^d, x]$: 切り上げの場合
 - $[x - 0.5 \times 0.1^d, x + 0.5 \times 0.1^d)$: 四捨五入の場合
 - $[x - 0.55 \times 0.1^d, x + 0.5 \times 0.1^d)$: 四捨五入を2回行った場合。
 - 10進数を意識して計算する(2進小数で計算するとまずい)。
3. 候補となる分母の整数Dに対し:
 - その半区間にDを乗算する。
 - その半区間に含まれる整数Nが**分子の候補**。
4. $D = 1, 2, 3, \dots$ において、全ての近似値において:
 - Dに対するNの候補が存在するかどうかを判定。
 - 途中の計算結果も分かりやすく表示する。

参考用; 知っている各種ソフトを比較するつもりでないと、このページはあまり役に立たないであろう。

設計方針

- 汎用性(古めの環境でも、50年後でも。)
- コマンド入力時に使えるオプション
 - 的確なオプションを設計。結果的に多数。
- 出力結果:
 - 色(ANSIカラー)を多用している。
- コマンドラインを使うことを前提とする
 - なお、JavaScript + HTMLも検討したが、未着手。
- 機能の複雑さに応じた内部設計
 - 必ずしも、平易なプログラムでは無い。
- 切れ味の良さ ← 現状曖昧; 素朴な疑問に的確に応じているかetc.

コマンドのインストール方法

- インストール: **cpanm App::denomfind**
- アンインストール: **cpanm -U App::defnomfind**

cpanmが使えない場合は、**cpan**でも良いが、モジュールのアンインストールが難しくなる。

上記は Perlのモジュールのインストールでよく用いられる方法である。

cpanmや**cpan**による一般的なインストールの方法は下記で網羅的に説明をしている。

1. <https://metacpan.org/release/TULAMILI/Table-Hack-0.12/source/Table-Hack.pdf>
2. <https://www.ite.or.jp/wp/wp-content/uploads/opendata/tool201807.pdf>

上記2個のURLにある同じPDF形式ファイルは、2018年に下野寿之が用意している。

コマンドの用途と有用性：

- **割合の近似値**から、分母となる整数の候補を見つける。
- 割合の近似値がいくつもあれば、分母が同じ場合は、分母の値の候補を少なくできる。
- **割合の数値を記載した文書**に対して、下記が分かる：
 1. (サンプルサイズNが明示されて無くても) **Nが十分大きいか**
 2. 数値の記載ミスの少ない、**完成度が高い文書**と言えるか
 1. Nの値の記載ミス、記載した割合の数値の記載ミス
 2. そもそも共通の分母の値を持っていない可能性
 - (数値記載ミスがある場合は、それを正しく復元できる。)
 3. **でたらめな数**を記載していないか

参考用；コマンドを何度も使わない無いと、このページはあまり役に立たないであろう。

コマンド denomfind の引数

- **denomfind**
 - g [最小分母,]探索個数
 - y 「整数分子を持ちうる割合の個数」をフィルタリング指定
 - D num ← シアン色で、割り算の式と小数点以下num桁表示
 - I ← 分子の候補となる値の区間範囲を出力する指定
 - Q ← 分子の候補をコンマ区切りで並べることで出力が簡便に。
 - % ← 残りの引数の値が百分率であることの指定
 - num [num] .. [num] ← 割合の近似値(主要な入力値)
- **その他のオプション：**
 - -f により**切り捨て**による近似であることを指定
 - -c により**切り上げ**による近似であることを指定
 - 上記のような -f も -c も無ければ**四捨五入**による近似を指定
- (出力に関わるオプションは英語大文字を採用。)

コマンドdenomfindの出力方法

- 出力はANSIカラー付きのテキスト形式
- 出力各行について：
 - 1列目：候補となる分母の値(整数) D
 - 2列目：引数の割合近似値(n個)の内の何個に対して、Dを分母として整数の分子を持つことが、可能か。**0からnの値**となる。n以外の値は候補外と見なす。
 - 3列目以降：補助情報
 - 候補となる分数を、**割り算の式**で示す。
 - 分子のとりうる値を、**整数に限らない実数の区間**で示す。

denomfind 以外に使うプログラム

- ある分母の値が候補なら、
その値の倍数も全て候補になるので、
それらの出力を最初だけに抑制する
コマンド **primefind** も作成した。
- App::expandtab に含まれる **expandtab** も作成。
タブ文字区切り入力をタブで立て揃え出力する。
- 出力の色を無くす colorplus -0 (App::colorplus)
- なお、古くからあるテキストビューアの less は
使いこなしていることを前提とする。

コマンドdenomfindの完成度について

- 4年前に作ったプログラムについて、最近、重大な見つけた。
- 結局、200行弱のPerlプログラムをほぼ全部リファクタをしながら点検をした。
- さらに新たな機能をオプションの形で追加したし、出力の表の形も見やすいようにと、何度も形を変えた。
- 最近手をえた部分について、再びほぼ全部リファクタして点検する必要はありそう。機能を加えると、関係する既に実装した機能に対して、バグを見つけることがある。
- 日本語の部分を英語化したい。
- 他の言語に移植したい。そして、もっと数理的な思いつきを容易に検証できるような状態にしたい。



3. コマンドの実行例

実際の実行画面で、どのような操作画面になるか、
実感に近いイメージを、一旦、提供する。

コマンドの使用例

異なる5個の条件に「**2.0% 50.7% 13.5% 21.6% 6.8%**」が該当した時に、サンプルサイズを**148人**が正しいかどうか確かめたい。

denom	fit	f1:0.020	f2:0.507	f3:0.135	f4:0.216	f5:0.068
141	1	-	-	19/141=.13475	-	-
142	1	-	72/142=.50704	-	-	-
143	0	-	-	-	-	-
144	1	-	73/144=.50694	-	-	-
145	0	-	-	-	-	-
146	2	-	74/146=.50685	-	-	10/146=.06849
147	2	3/147=.02041	-	-	-	10/147=.06803
148	5	3/148=.02027	75/148=.50676	20/148=.13514	32/148=.21622	10/148=.06757
149	1	3/149=.02013	-	-	-	-
150	2	3/150=.02000	76/150=.50667	-	-	-
151	1	3/151=.01987	-	-	-	-
152	2	3/152=.01974	77/152=.50658	-	-	-
153	2	3/153=.01961	-	-	33/153=.21569	-
154	0	-	-	-	-	-
155	1	-	-	21/155=.13548	-	-
156	1	-	-	21/156=.13462	-	-
157	0	-	-	-	-	-
158	0	-	-	-	-	-
159	0	-	-	-	-	-
160	0	-	-	-	-	-

分母の値を141から初めて、20個の場合について結果を表示した。

コマンドの使用例

5個の条件に「**2.0% 50.7% 13.5% 21.6% 6.8%**」が該当した場合に、サンプルサイズを推定(5個中2個以内の記載ミスの場合も考えた)。

```
> denomfind -y 3..5 -g141,20 -D5 -% 2.0 50.7 13.5 21.6 6.8 | expandtab
denom fit f1:0.020 f2:0.507 f3:0.135 f4:0.216 f5:0.068
148 5 3/148=.02027 75/148=.50676 20/148=.13514 32/148=.21622 10/148=.06757
205 3 4/205=.01951 104/205=.50732 - - 14/205=.06829
207 3 - 105/207=.50725 28/207=.13527 - 14/207=.06763
222 3 - - 30/222=.13514 48/222=.21622 15/222=.06757
245 3 5/245=.02041 - 33/245=.13469 53/245=.21633 -
250 3 5/250=.02000 - - 54/250=.21600 17/250=.06800
251 3 5/251=.01992 - 34/251=.13546 - 17/251=.06773
278 3 - 141/278=.50719 - 60/278=.21583 19/278=.06835
282 3 - 143/282=.50709 38/282=.13475 61/282=.21631 -
292 3 - 148/292=.50685 - 63/292=.21575 20/292=.06849
294 3 6/294=.02041 149/294=.50680 - - 20/294=.06803
296 5 6/296=.02027 150/296=.50676 40/296=.13514 64/296=.21622 20/296=.06757
304 3 6/304=.01974 154/304=.50658 41/304=.13487 - -
306 3 6/306=.01961 155/306=.50654 - 66/306=.21569 -
310 3 - - 42/310=.13548 67/310=.21613 21/310=.06774
342 3 7/342=.02047 - 46/342=.13450 74/342=.21637 -
343 3 7/343=.02041 174/343=.50729 - 74/343=.21574 -
347 4 7/347=.02017 176/347=.50720 47/347=.13545 75/347=.21614 -
348 3 7/348=.02011 - 47/348=.13506 75/348=.21552 -
349 3 7/349=.02006 177/349=.50716 47/349=.13467 -
```

各分母に対して、整数同士の割り算になる場合が3~5個の場合に、限定した。

コマンドの使用例

表示内容は、割り算式ではなくて、分子の値の範囲を表示。
緑色の場合は、その区間に整数が1個以上含まれることを示す。

```
> denomfind -y 0,3,4,5 -g290,20 -I -% 2.0 50.7 13.5 21.6 6.8 | expandtab
denom fit numerator_1 numerator_2 numerator_3 numerator_4 numerator_5
292 3 [5.694 5.986) [147.898 148.19) [39.274 39.566) [62.926 63.218) [19.71 20.002)
294 3 [5.733 6.027) [148.911 149.205) [39.543 39.837) [63.357 63.651) [19.845 20.139)
296 5 [5.772 6.068) [149.924 150.22) [39.812 40.108) [63.788 64.084) [19.98 20.276)
304 3 [5.928 6.232) [153.976 154.28) [40.888 41.192) [65.512 65.816) [20.52 20.824)
306 3 [5.967 6.273) [154.989 155.295) [41.157 41.463) [65.943 66.249) [20.655 20.961)
310 3 [6.045 6.355) [157.015 157.325) [41.695 42.005) [66.805 67.115) [20.925 21.235)
313 0 [6.1035 6.4165) [158.5345 158.8475) [42.0985 42.4115) [67.4515 67.7645) [21.1275 21.4405)
314 0 [6.123 6.437) [159.041 159.355) [42.233 42.547) [67.667 67.981) [21.195 21.509)
316 0 [6.162 6.478) [160.054 160.37) [42.502 42.818) [68.098 68.414) [21.33 21.646)
317 0 [6.1815 6.4985) [160.5605 160.8775) [42.6365 42.9535) [68.3135 68.6305) [21.3975 21.7145)
321 0 [6.2595 6.5805) [162.5865 162.9075) [43.1745 43.4955) [69.1755 69.4965) [21.6675 21.9885)
330 0 [6.435 6.765) [167.145 167.475) [44.385 44.715) [71.115 71.445) [22.275 22.605)
331 0 [6.4545 6.7855) [167.6515 167.9825) [44.5195 44.8505) [71.3305 71.6615) [22.3425 22.6735)
332 0 [6.474 6.806) [168.158 168.49) [44.654 44.986) [71.546 71.878) [22.41 22.742)
342 3 [6.669 7.011) [173.223 173.565) [45.999 46.341) [73.701 74.043) [23.085 23.427)
343 3 [6.6885 7.0315) [173.7295 174.0725) [46.1335 46.4765) [73.9165 74.2595) [23.1525 23.4955)
347 4 [6.7665 7.1135) [175.7555 176.1025) [46.6715 47.0185) [74.7785 75.1255) [23.4225 23.7695)
348 3 [6.786 7.134) [176.262 176.61) [46.806 47.154) [74.994 75.342) [23.49 23.838)
349 3 [6.8055 7.1545) [176.7685 177.1175) [46.9405 47.2895) [75.2095 75.5585) [23.5575 23.9065)
351 3 [6.8445 7.1955) [177.7815 178.1325) [47.2095 47.5605) [75.6405 75.9915) [23.6925 24.0435)
> █
```



4. 冒頭2個の題材への適用

■ 接種後7日間に現れた症状

2回目接種の250人という数については、正しそうだ。

報告割合	1回目接種後	2回目接種後	3回目接種後
50%以上		圧痛 65.2%	圧痛 81.4% 疲労 63.3% 疼痛 54.6% 筋肉痛 51.0%
10 - 50%	圧痛 48.2% 疼痛 26.9% 疲労 23.1% 頭痛 21.6% 筋肉痛 20.0% 倦怠感 12.2%	疼痛 45.6% 疲労 35.6% 筋肉痛 30.8% 頭痛 29.6% 倦怠感 26.4% 関節痛 14.8%	倦怠感 46.9% 頭痛 45.9% 関節痛 28.6% 発熱 17.3% 恶心・嘔吐 13.3% 腫脹・硬結 11.3% 紅斑 10.3%
1 - 10%	関節痛 6.7% 恶心・嘔吐 5.9% 発熱 2.4%	恶心・嘔吐 7.2% 腫脹・硬結 5.6% 紅斑 4.8% 発熱 4.4%	

(注)対象:1回目接種後253人、2回目接種後250人、3回目接種後97人

出典:審査報告書より改編

```
> denomfind -y5..9 -D0 -g1,10 -% 48.2 26.9 23.1 21.6 20.0 12.2 6.7 5.9 2.4 | expandtab
Given rates are 9 pieces.
denom fit f1:0.482 f2:0.269 f3:0.231 f4:0.216 f5:0.200 f6:0.122 f7:0.067 f8:0.059 f9:0.024
208 5 - 56/208= 48/208= 45/208= - - 14/208= - 5/208=
238 5 - 64/238= 55/238= - - 29/238= 16/238= 14/238= -
245 6 118/245= 66/245= - 53/245= 49/245= 30/245= - - 6/245=
253 5 122/253= 68/253= - - - 17/253= 15/253= 6/253=
255 8 123/255= - 59/255= 55/255= 51/255= 31/255= 17/255= 15/255= 6/255=
286 5 - 77/286= 66/286= - - 35/286= - 17/286= 7/286=
290 5 - 78/290= 67/290= - 58/290= - - 17/290= 7/290=
305 5 147/305= 82/305= - 66/305= 61/305= - - 18/305= -
320 6 - 86/320= 74/320= 69/320= 64/320= 39/320= - 19/320= -
328 5 158/328= - - 71/328= - 40/328= 22/328= - 8/328=
```

```
> denomfind -y4..11 -D0 -g1,10 -% 81.4 63.3 54.6 51.0 46.9 45.9 28.6 17.3 13.3 11.3 10.3 | expandtab
Given rates are 11 pieces.
denom fit f1:0.814 f2:0.633 f3:0.546 f4:0.510 f5:0.469 f6:0.459 f7:0.286 f8:0.173 f9:0.133 f10:0.113 f11:0.103
49 4 - 31/49=1 - 25/49=1 23/49= - 14/49= - - - - -
97 4 79/97=1 - 53/97=1 - - - - - 11/97= 10/97= -
98 7 - 62/98=1 - 50/98=1 46/98= 45/98= 28/98= 17/98= 13/98= - - -
133 4 - - - - 61/133= 38/133= 23/133= - 15/133= -
145 4 118/145=1 - - 74/145=1 68/145= - - - - 15/145= -
147 4 - 93/147=1 - 75/147=1 69/147= - 42/147= - - - -
150 4 - 95/150=1 - - - - 26/150= 20/150= 17/150= -
177 4 144/177=1 112/177=1 - - 83/177= - - - - 20/177= -
185 5 - - 101/185=1 - - 85/185= 53/185= 32/185= - - 19/185= -
194 7 158/194=1 - 106/194=1 99/194=1 91/194= 89/194= - - - 22/194= 20/194=
```

ここに書かれた説明は、1回目接種後と3回目接種後のことについて、誤りであると分かる。

1回目が253人としても、～%表示の9個の内、5個しか分子が整数になり得ない。255の書き間違いの可能性もあるが、それでも、8個だけが整数分子。

3回目が97人としても、～%の表示の内、4個しか分子が整数になり得ない。

3回目接種後97人

3回目接種後

圧痛 81.4% 疲労 63.3%
疼痛 54.6% 筋肉痛 51.0%

倦怠感 46.9% 頭痛 45.9%
関節痛 28.6% 発熱 17.3%
恶心・嘔吐 13.3%
腫脹・硬結 11.3% 紅斑 10.3%

denomfind で計算してみると、
97人のデータと98人のデータが
混在しているようにも見える。

```
> denomfind -5 -D5 -Q -%g 85,30 81.4 63.3 54.6 51.0 46.9 45.9 28.6 17.3 13.3 11.3 10.3 | expandtab
Given rates are 11 pieces.
denom fit f1:0.814 f2:0.633 f3:0.546 f4:0.510 f5:0.469 f6:0.459 f7:0.286 f8:0.173 f9:0.133 f10:0.113 f11:0.103 dividends
85 1 - - - - 39/85=.45882 - - - - 39
86 1 70/86=.81395 - - - - - - - - 70
87 1 - - - - - - - - - 9/87=.10345 9
88 1 - - 48/88=.54545 - - - - - - 48
89 0 - - - - - - - - - -
90 2 - 57/90=.63333 - - - - - 12/90=.13333 - - 57,12 (45q+12)
91 1 - - - - - 26/91=.28571 - - - - 26
92 0 - - - - - - - - - -
93 0 - - - - - - - - - -
94 0 - - - - - - - - - -
95 0 - - - - - - - - - -
96 2 - - - 49/96=.51042 45/96=.46875 - - - - - 49,45 (4q+1)
97 4 79/97=.81443 - 53/97=.54639 - - - - 11/97=.11340 10/97=.10309 79,53,11,10
98 7 - 62/98=.63265 - 50/98=.51020 46/98=.46939 45/98=.45918 28/98=.28571 17/98=.17347 13/98=.13265 - 62,50,46,45,28,17,13
99 1 - - 54/99=.54545 - - - - - - 54
100 1 - - - 51/100=.51000 - - - - - 51
101 0 - - - - - - - - - -
102 2 83/102=.81373 - 52/102=.50980 - - - - - - 83,52 (31q+21)
103 0 - - - - - - - - - -
104 2 - - - 53/104=.50962 - - - 18/104=.17308 - - 53,18 (35q+18)
105 2 - - - - - - 30/105=.28571 - 14/105=.13333 - 30,14 (16q+14)
106 1 - - - - - - - - - 12/106=.11321 - 12
107 1 - - - - - - - - - 11/107=.10280 11
108 1 - - 59/108=.54630 - - - - - - 59
109 2 - 69/109=.63303 - - 50/109=.45872 - - - - - 69,50 (19q+12)
110 2 - - 60/110=.54545 - - - 19/110=.17273 - - - 60,19 (41q+19)
111 2 - - - 52/111=.46847 51/111=.45946 - - - - - 52,51
112 1 - - - - - 32/112=.28571 - - - - 32
113 3 92/113=.81416 - - - 53/113=.46903 - - 15/113=.13274 - - 92,53,15
114 0 - - - - - - - - - -
```

“小学生”で
あり得る数:
471,516,529..

小学生 TOP5					中学生 TOP5				
1位	▷ denomfind -y5 -Qg3 -%20 \	39.9%	1位	▷ denomfind -y5 -Qg3 -%20 \	33.7%				
2位	39.9 30.4 29.5 18.0 17.2	30.4%	2位	33.7 31.8 23.5 22.7 22.0	31.8%				
3位	denom fit dividends	29.5%	3位	denom fit dividends	23.5%				
4位	471 5 188,143,139,85,81	18.0%	4位	255 5 86,81,60,58,56	22.7%				
5位	516 5 206,157,152,93,89	17.2%	5位	264 5 89,84,62,60,58	22.0%				
	▷			▷					

“男子”で
あり得る数:
373,384,520..

男子 TOP5 (小中学生)					女子 TOP5 (小中学生)				
1位	▷ denomfind -y5 -Qg3 -%20 \	37.5%	1位	▷ denomfind -y5 -Qg3 -%20 \	39.4%				
2位	37.5 36.2 27.3 18.8 18.5	36.2%	2位	39.4 35.7 21.5 20.7 18.4	35.7%				
3位	denom fit dividends	27.3%	3位	denom fit dividends	21.5%				
4位	373 5 140,135,102,70,69	18.8%	4位	353 5 139,126,76,73,65	20.7%				
5位	384 5 144,139,105,72,71	18.5%	5位	381 5 150,136,82,79,70	18.4%				
	▷			▷					

複数回答 有効回答726人

“中学生”で
あり得る数:
255,264,409..

“女子”で
あり得る数:
353,381,526..

- 有効回答数726人とされている小中学生のお年玉について、小学生/中学生/男子/女子の内訳でそれぞれの使い道のトップ5項目の割合が表示されていた。
- その割合の数を四捨五入と見なし、5個の数を整数比の分数である場合に、**各内訳の5個に共通する分母としてあり得る数**を、小さい順に算出した。
- **denomfind -y5 -Q -g3 -% -20 39.9 30.4 29.5 18.0 17.2** のようなコマンドを用いた。
- 39.9などの数を-%で%表記と見なし、-y5により5個の割合数値が全て5個とも整数分子が見つかる整数分母を探索した。
- なお、-Qで分子の値を表示して、-20により冗長な2次情報の抑制を指定している。
- それらのあり得る数の内、小学生と中学生で足して726、男子と女子で足しても726となる数は、上記の丸で囲った数のみである。
- **各内訳(小/中/男/女)のどれもが、255人以上の少なくない人数と判明。**
- 候補として算出した数の、各4個の最小値だけが解となつた不思議さがある。 22

“割合の数値の不具合”がある場合の類型

1. 単純な記載ミス：

- 例. 十数個もあったので、どこか間違う。

2. 数値丸めの際の不整合：

- 切り捨てと四捨五入の混在。
- 四捨五入を2回(以上)行った(例. 0.45 → 0.5 → 1)

3. データの作られ方の理解の誤り：

- 例. 異なるサイズの集団の結果を併合している。

4. そもそもデータが出鱈目な場合

なお、ある資料で不具合が1つも発見できないなら
その資料で、さらに奥深く記載数値について
何か研究が出来るであろう。

5. 補足

謎な現象とその考察

- ある場合に7個もの分子の数が全て3の倍数であった。
- 珍しくて、特記すべき事象と言えるか？
- 全ての分子が、特殊な性質を満たす場合、
 p 値のような考え方で、その場合の分母は、特殊だと言えるだろうか？— 他、整数 q に対して $5q+3$ のような整数が7個もある場合

```
> denomfind -20 -g1,1000 -y7..11 -Q -% 65.2 45.6 35.6 30.8 29.6 26.4 14.8 | grep q
500 7 326,228,178,154,148,132,74 (2q)
675 7 440,308,240,208,200,178,100 (2q)
750 7 489,342,267,231,222,198,111 (3q)
825 7 538,376,294,254,244,218,122 (2q)
953 7 621,435,339,294,282,252,141 (3q)
1000 7 652,456,356,308,296,264,148 (4q)
1175 7 766,536,418,362,348,310,174 (2q)
1250 7 815,570,445,385,370,330,185 (5q)
1325 7 864,604,472,408,392,350,196 (2q)
1500 7 978,684,534,462,444,396,222 (6q)
1750 7 1141,798,623,539,518,462,259 (7q)
> █
```

コマンドprimefindについて

1. 入力の各行の最初の数値を読み取る。
2. 逐次、入力の各行を読む段階で読み取った数値が、既に読み取った数値の倍数かどうかを判定。そうであれば出力しないし、そうで無ければ、入力行をそのまま出力する。
3. **seq 2,100 | primefind** とすると100以下の素数。
4. このことにより、denomfindで算出した分母の候補について、既に候補となった値の倍数が、全て出力される冗長さを抑制できる。

denomfindの作成時の難点

1. 多数の試行錯誤: 出力の形式とオプションの設計
2. 重大なバグを見つけた:
 1. 10%の指定を、近似値0.1と見なし0.05超0.15以下と解釈。
 2. 1.0%の指定を 1%と解釈して区間が**10倍**にしてしまう。
 - 2008年頃に初めてRubyで作ったので慣れていたと思ったが、2018年に作ったPerlのプログラムにあった。
3. 作り込むべき機能が、今後使っているとまだ見つかりそう。



他のプログラムについて

- 2008年頃に一度Rubyで作った。
- primefindはdenomfindと共にApp::denomfindに同梱。
- App::colorplus のcolorplus で colorplus -0により着色を消すことができる。
- App::expandtab のexpandtab でタブ文字の縦揃えが出来る。