

Introduction to Statistics

統計学入門

Week 9 | June 20, 2023

Week 8 小テスト

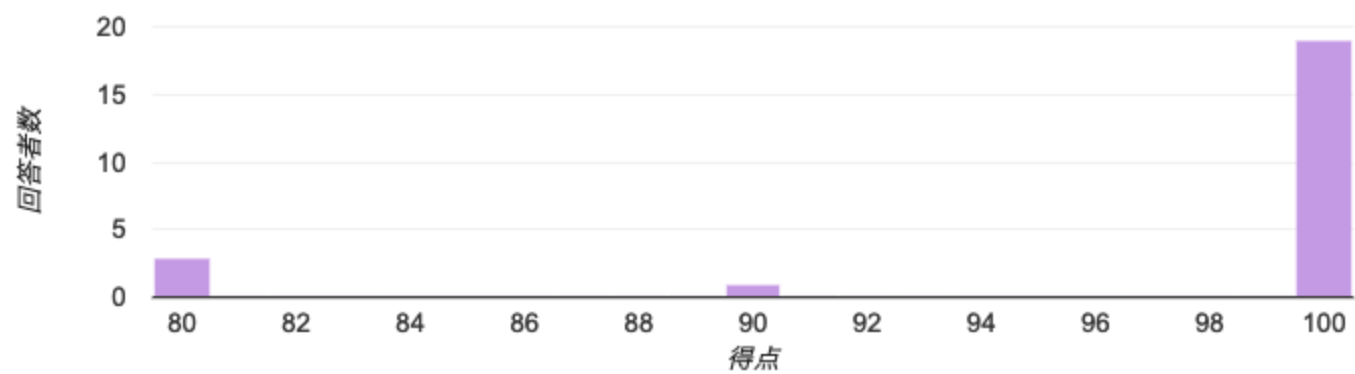


平均
96.96/100 ポイント

中央値
100/100 ポイント

範囲
80~100 ポイント

合計点の分布



Last week

標準偏差の公式

(population ・ 母集団)

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

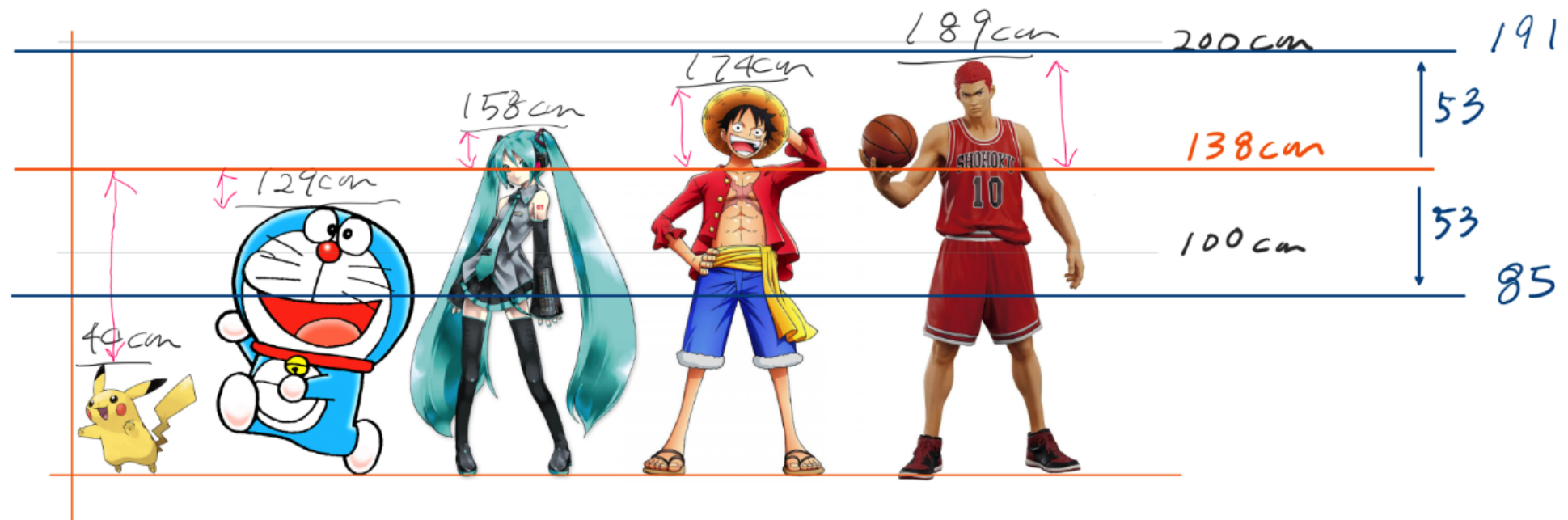
標準偏差の求め方

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

Diagram illustrating the components of the standard deviation formula with numbered labels:

- 7: σ (Standard Deviation)
- 5: $\frac{1}{N}$ (Reciprocal of sample size)
- 4: $\sum_{i=1}^N$ (Summation over all data points)
- 2: $(x_i - \mu)^2$ (Squared deviation from the mean)
- 1: μ (Mean)
- 3: x_i (Data point)
- 6: The entire expression inside the square root (Variance)

順番	統計	記号	式
1	平均値	μ	$\frac{1}{N} \sum x_i$
2	偏差	-	$x_i - \mu$
3	偏差平方	-	$(x_i - \mu)^2$
4	偏差平方の和	-	$\sum (x_i - \mu)^2$
5	分散	σ^2	$\frac{1}{N} \sum (x_i - \mu)^2$
6	標準偏差	σ	$\sqrt{\sigma^2}$



平均値の計算:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

すなわち

$$\mu = \frac{(40_{\text{ピカチュ}} + 129_{\text{ドラえもん}} + 158_{\text{初音ミク}} + 174_{\text{ルフィー}} + 189_{\text{花道}})}{5} = 138cm$$

分散の計算：

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

すなわち

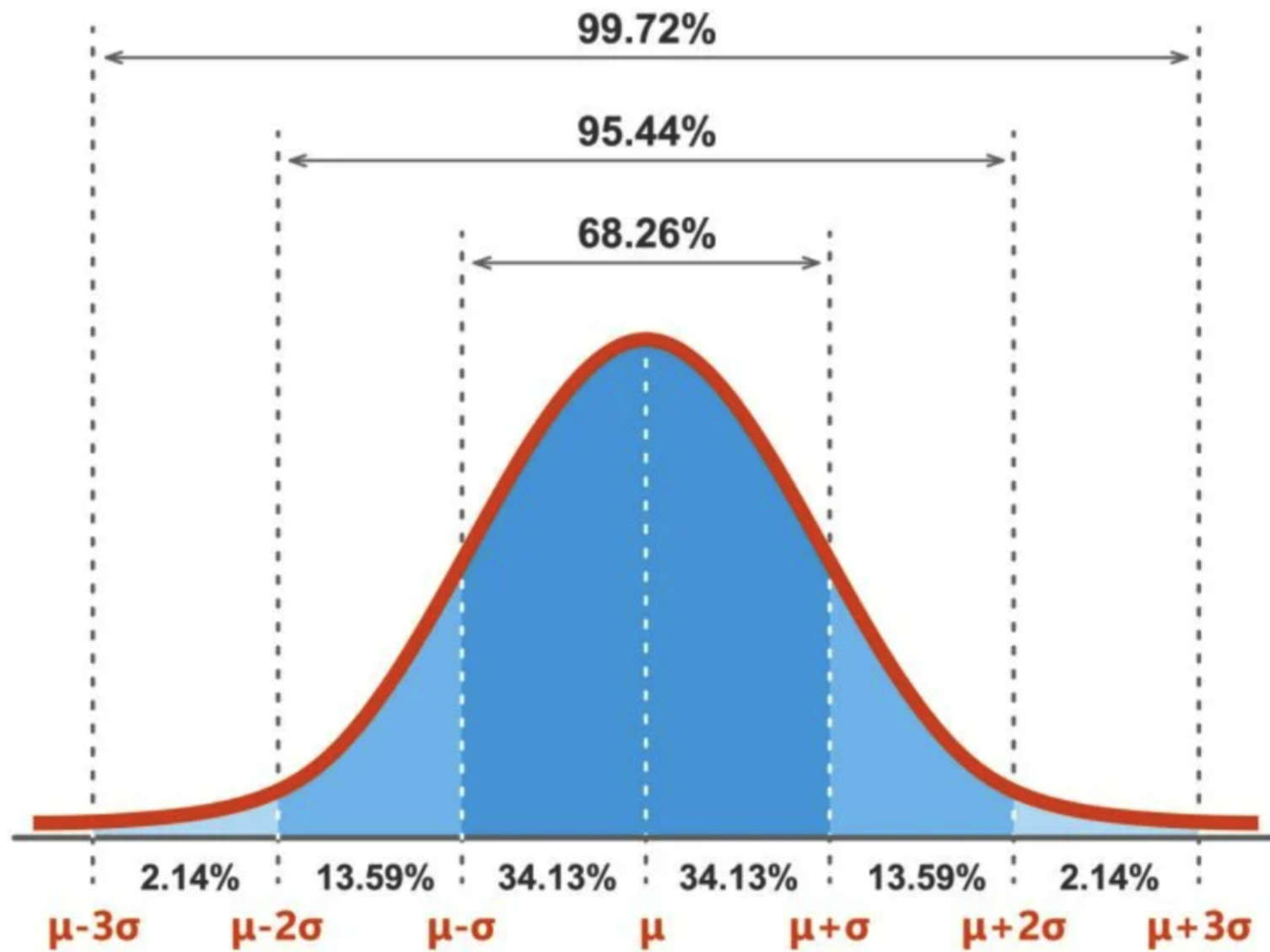
$$\begin{aligned}\sigma^2 &= \frac{(40 - 90)^2 + (129 - 90)^2 + (158 - 90)^2 + (174 - 90)^2 + (189 - 90)^2}{5} \\ &= \frac{13982}{5} \\ &= 2796.4\end{aligned}$$

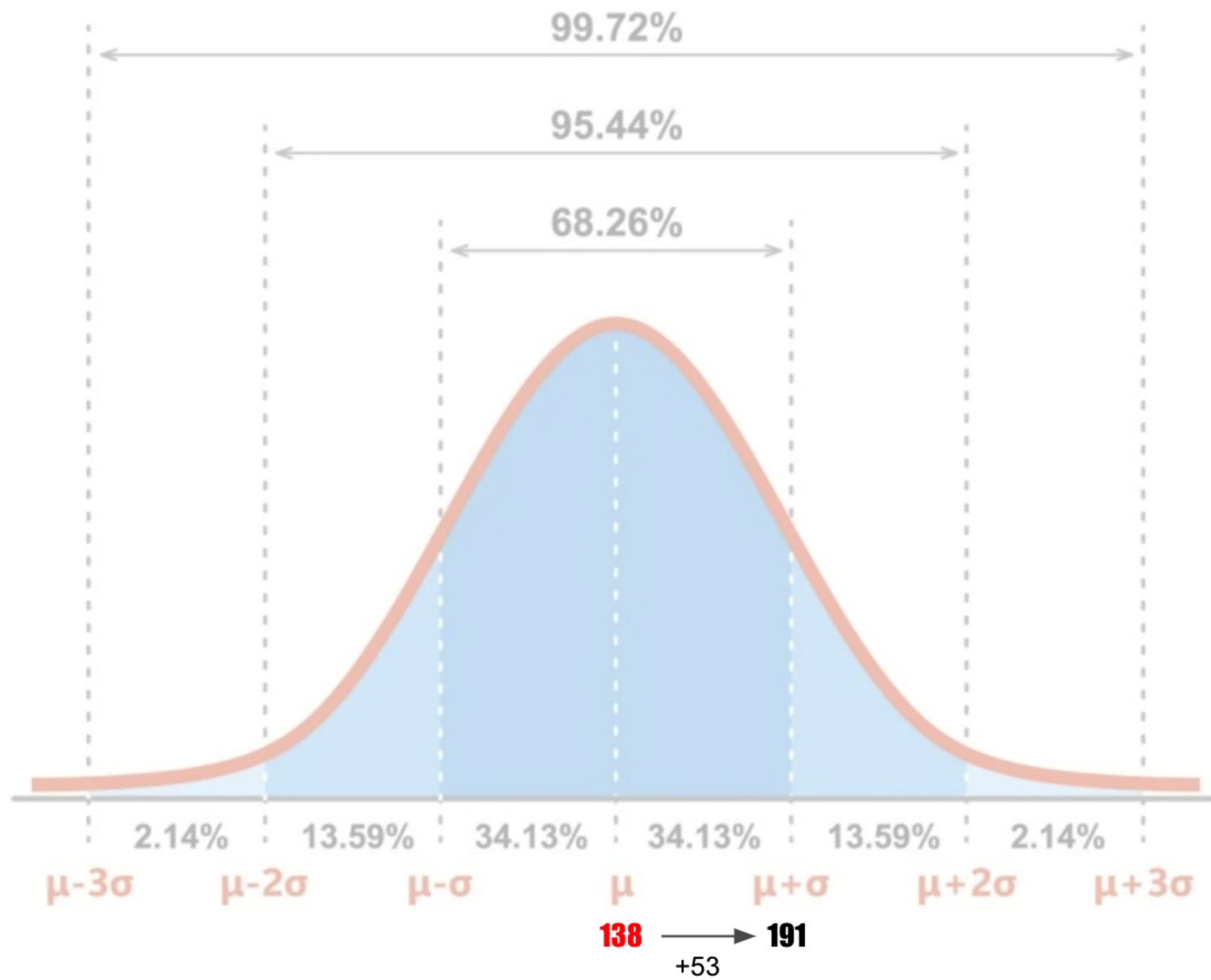
標準偏差の計算:

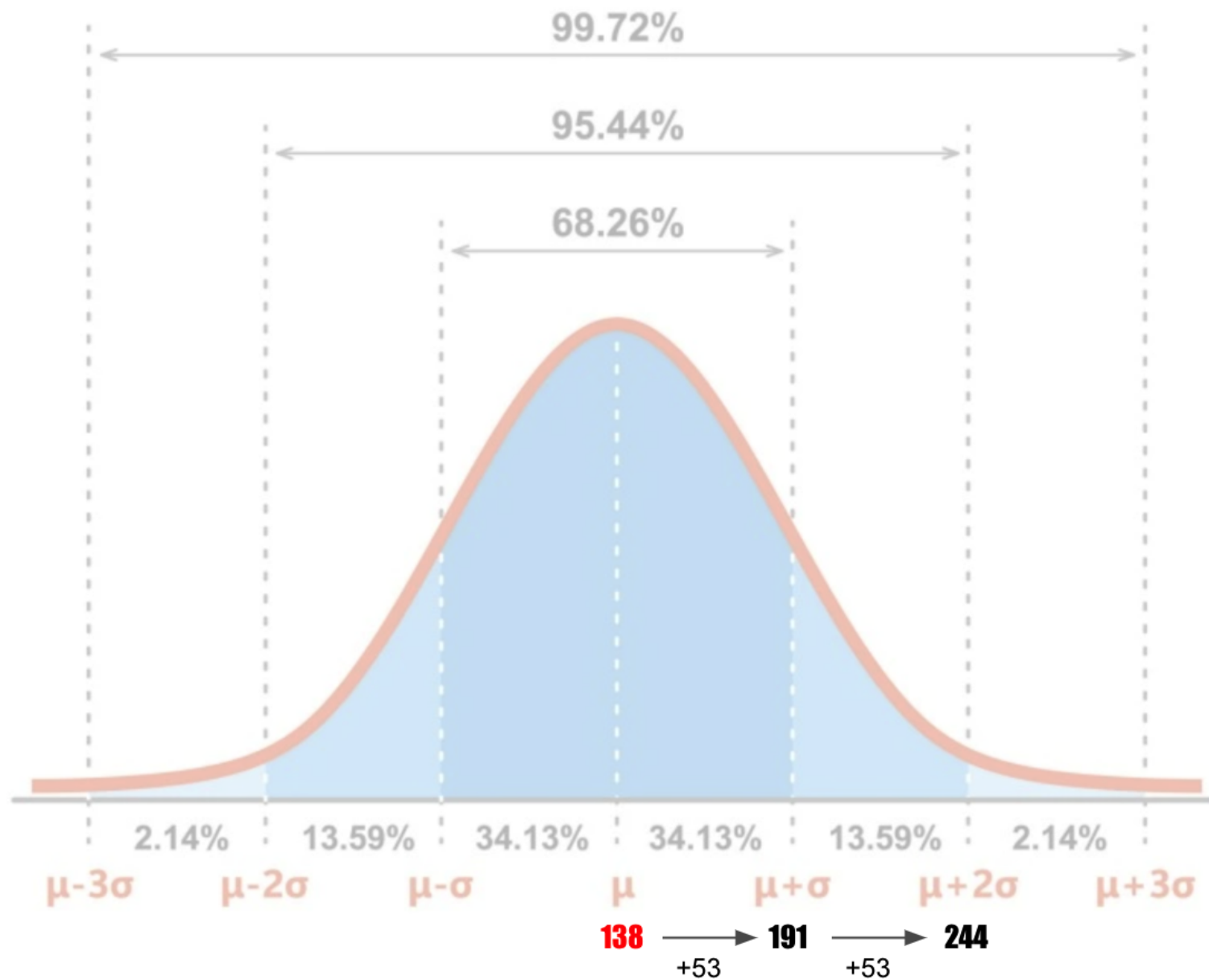
$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

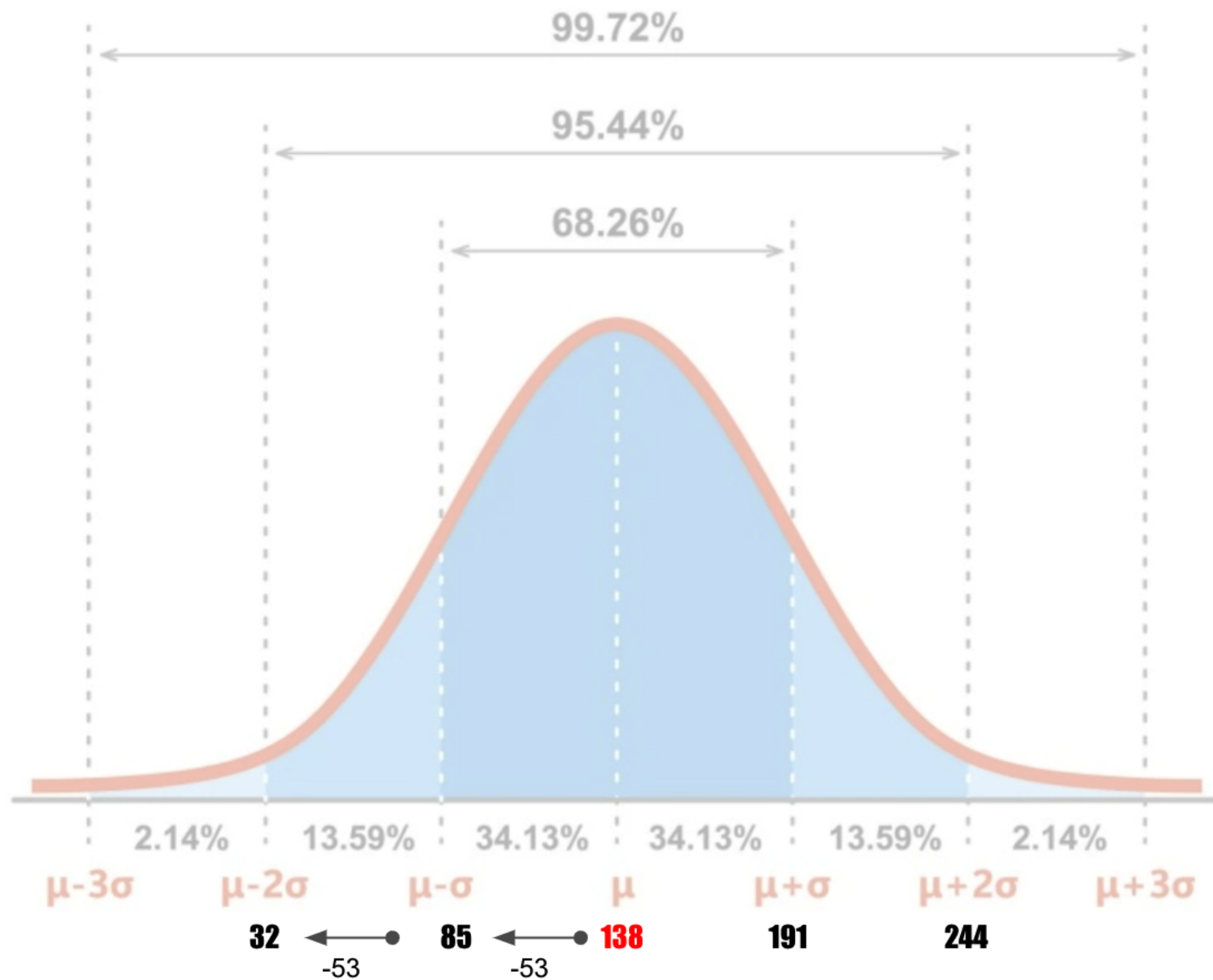
$$\sigma = \sqrt{2796.4}$$

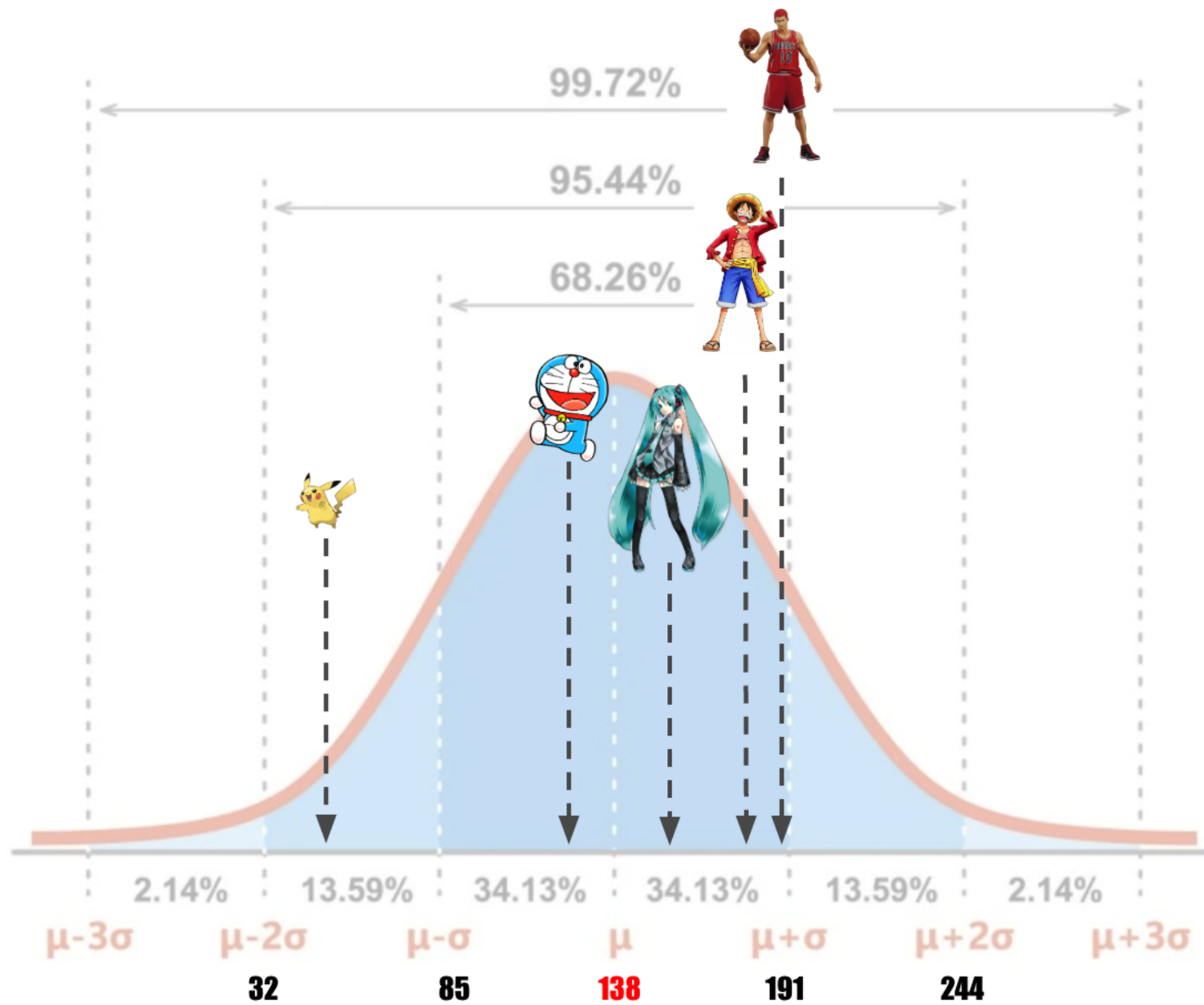
$$\sigma \approx 53cm$$











Topic #1: 変動係数

Coefficient of Variation

データの相対的な **ばらつき** を表す統計量

$$CV = \sigma / \mu$$

変動係数 = 標準偏差 / 平均値



そもそも変動係数っていつ使うの？



ある測定をカテゴリー別で **比べたい** 時！

例えば：50m走



6歳の時と19歳の時とどっちの方が
ばらつきある？

男と女のばらつきはどう？

	50m走					(秒)		
年齢	男 子				女 子			
	標本数	平均値	標準偏差	変動係数	標本数	平均値	標準偏差	変動係数
6	1104	11.40	1.01	0.09	1098	11.72	1.00	0.09
7	1099	10.55	0.81	0.08	1103	10.90	0.85	0.08
8	1104	10.00	0.81	0.08	1105	10.34	0.79	0.08
9	1107	9.61	0.76	0.08	1116	9.85	0.73	0.07
10	1097	9.17	0.73	0.08	1101	9.50	0.70	0.07
11	1102	8.78	0.71	0.08	1112	9.12	0.68	0.07
12	1355	8.42	0.75	0.09	1371	8.93	0.70	0.08
13	1359	7.77	0.63	0.08	1373	8.59	0.67	0.08
14	1370	7.42	0.55	0.07	1384	8.53	0.69	0.08
15	1386	7.43	0.56	0.08	1393	8.85	0.75	0.08
16	1386	7.26	0.53	0.07	1394	8.84	0.81	0.09
17	1370	7.16	0.53	0.07	1406	8.84	0.84	0.10
18	891	7.33	0.53	0.07	991	9.13	0.80	0.09
19	741	7.32	0.53	0.07	670	9.02	0.78	0.09

	50m走					(秒)		
年齢	男 子							
	標本数	平均値	標準偏差	変動係数	標本			
6	1104	11.40	1.01	0.09				
7	1099	10.55	0.81	0.08	1			
8	1104	10.00	0.81	0.08	1105	10.34	0.79	0.08
9	1107	9.61	0.76	0.08	1116	9.85	0.73	0.07
10	1097	9.17	0.73	0.08	1101	9.50	0.70	0.07
11	1102	8.78	0.71	0.08	1112	9.12	0.68	0.07
12	1355	8.42	0.75	0.09	1371	8.93	0.70	0.08
13	1359	7.77	0.63	0.08	1373	8.59	0.67	0.08
14	1370	7.42	0.55	0.07	1384	8.53	0.69	0.08
15	1386	7.43	0.56	0.08	1393	8.85	0.75	0.08
16	1386	7.26	0.53	0.07	1394	8.84	0.81	0.09
17	1370	7.16	0.53	0.07	1406	8.84	0.84	0.10
18	891	7.33	0.53	0.07	991	9.13	0.80	0.09
19	741	7.32	0.53	0.07	670	9.02	0.78	0.09

=1.01/11.40

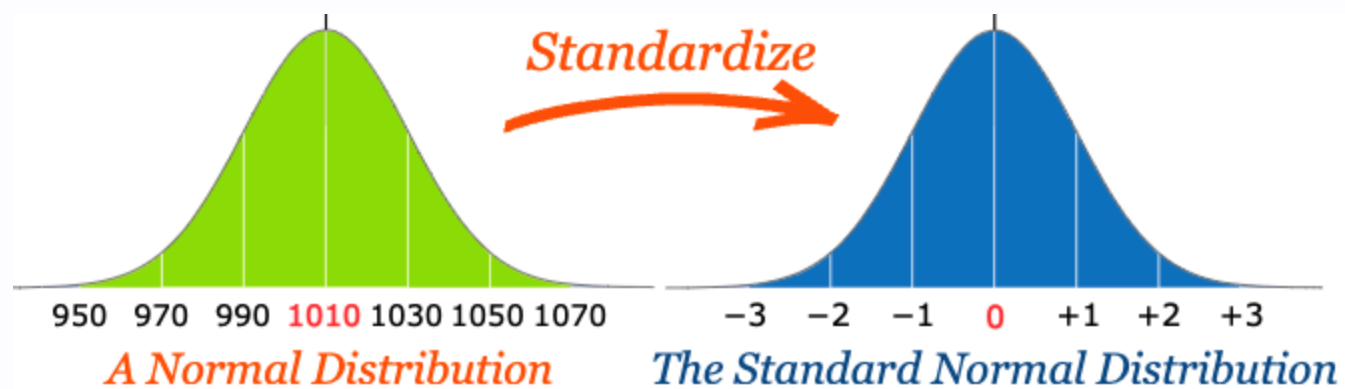
	50m走					(秒)		
年齢	男 子				女 子			
	標本数	平均値	標準偏差	変動係数	標本数	平均値	標準偏差	変動係数
6	1104	11.40	1.01	0.09	1098	11.72	1.00	0.09
7	1099	10.55	0.81	0.08	1103	10.90	0.85	0.08
8	1104	10.00	0.81	0.08	1105	10.34	0.79	0.08
9	1107	9.61	0.76	0.08	1116	9.85	0.73	0.07
10	1097	9.17	0.73	0.08	1101	9.50	0.70	0.07
11	1102	8.78	0.71	0.08	1112	9.12	0.68	0.07
12	1355	8.42	0.75	0.09	1371	8.93	0.70	0.08
13	1359	7.77	0.63	0.08	1373	8.59	0.67	0.08
14	1370	7.42	0.55	0.07	1384	8.53	0.69	0.08
15	1386	7.43	0.56	0.08	1393	8.85	0.75	0.08
16	1386	7.26	0.53	0.07	1394	8.84	0.81	0.09
17	1370	7.16	0.53	0.07	1406	8.84	0.84	0.10
18	891	7.33	0.53	0.07	991	9.13	0.80	0.09
19	741	7.32	0.53	0.07	670	9.02	0.78	0.09

ではやってみよう！

1. グループに分かれる
2. [このページ](#)から好きな項目を選ぶ
3. EXCELファイルをダウンロードして開く
4. 新しいカラムを作って、変動係数を計算する
5. 変動係数は： $\text{標準偏差} / \text{平均値}$
6. 面白く結果発表！

Topic #2: 標準化とZ得点

Z-Score



$$Z\text{得点} = \frac{x_i - \mu}{\sigma}$$

このクラスの平均身長

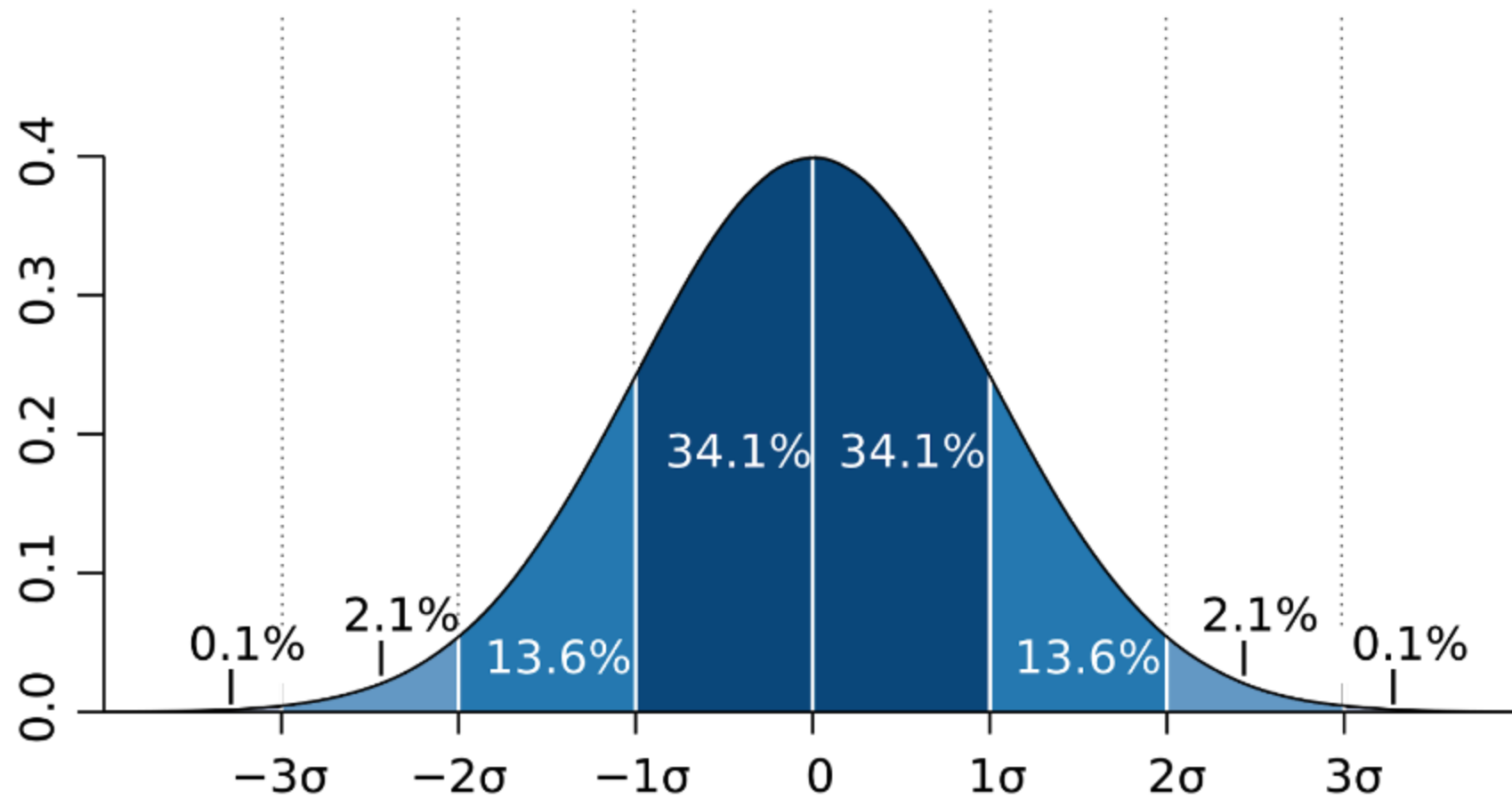
165cm(μ)

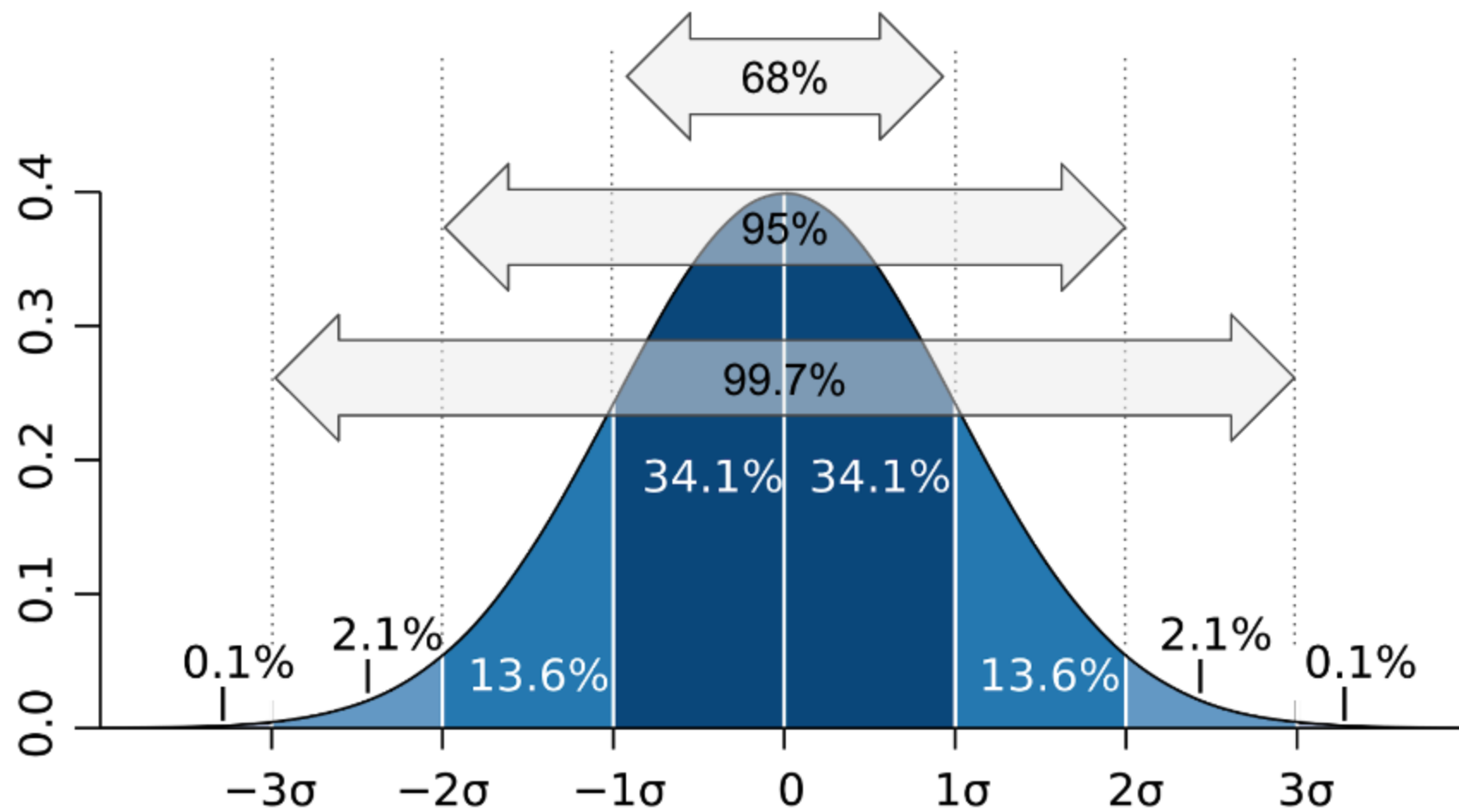
標準偏差

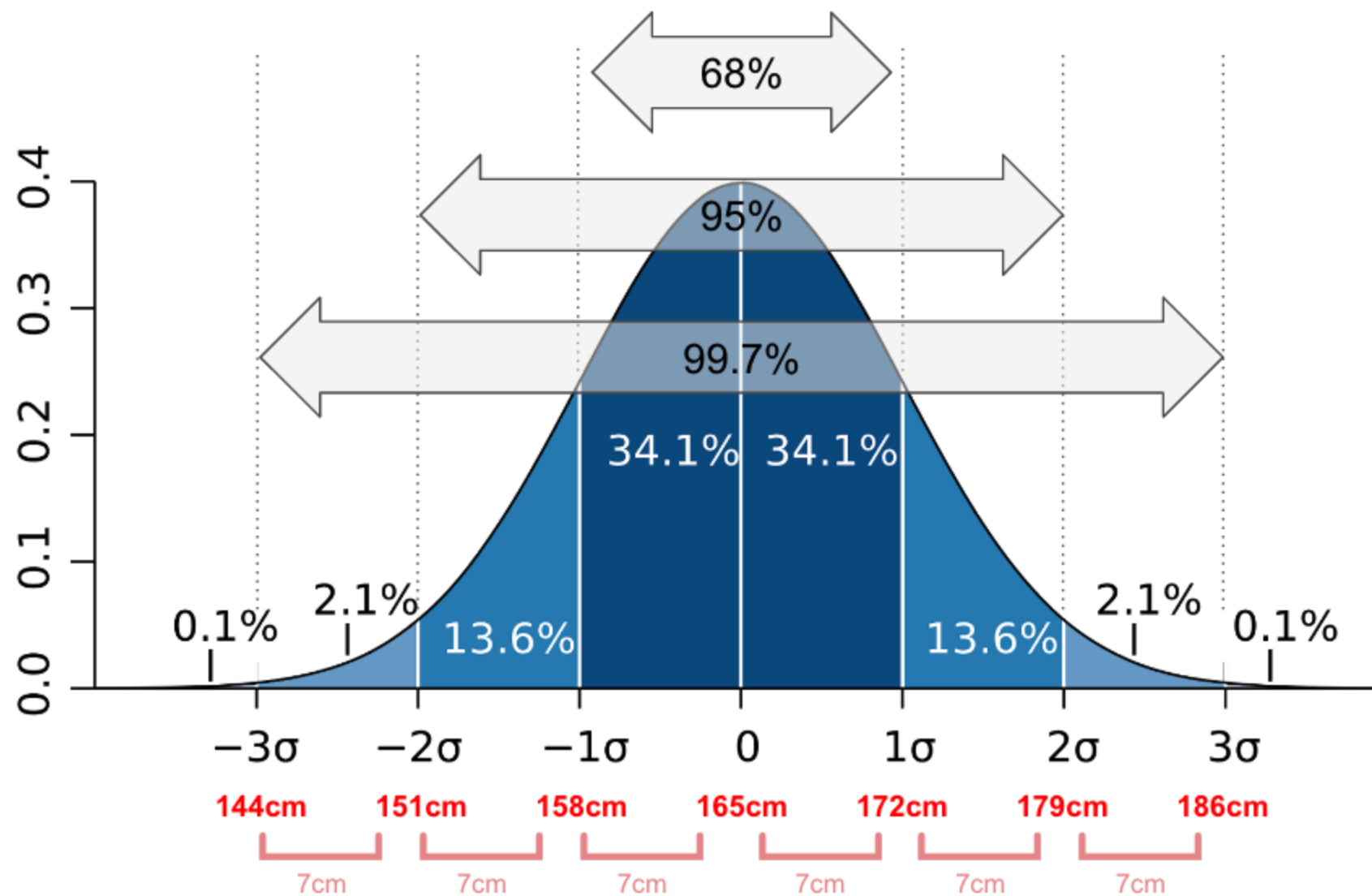
7cm(σ)

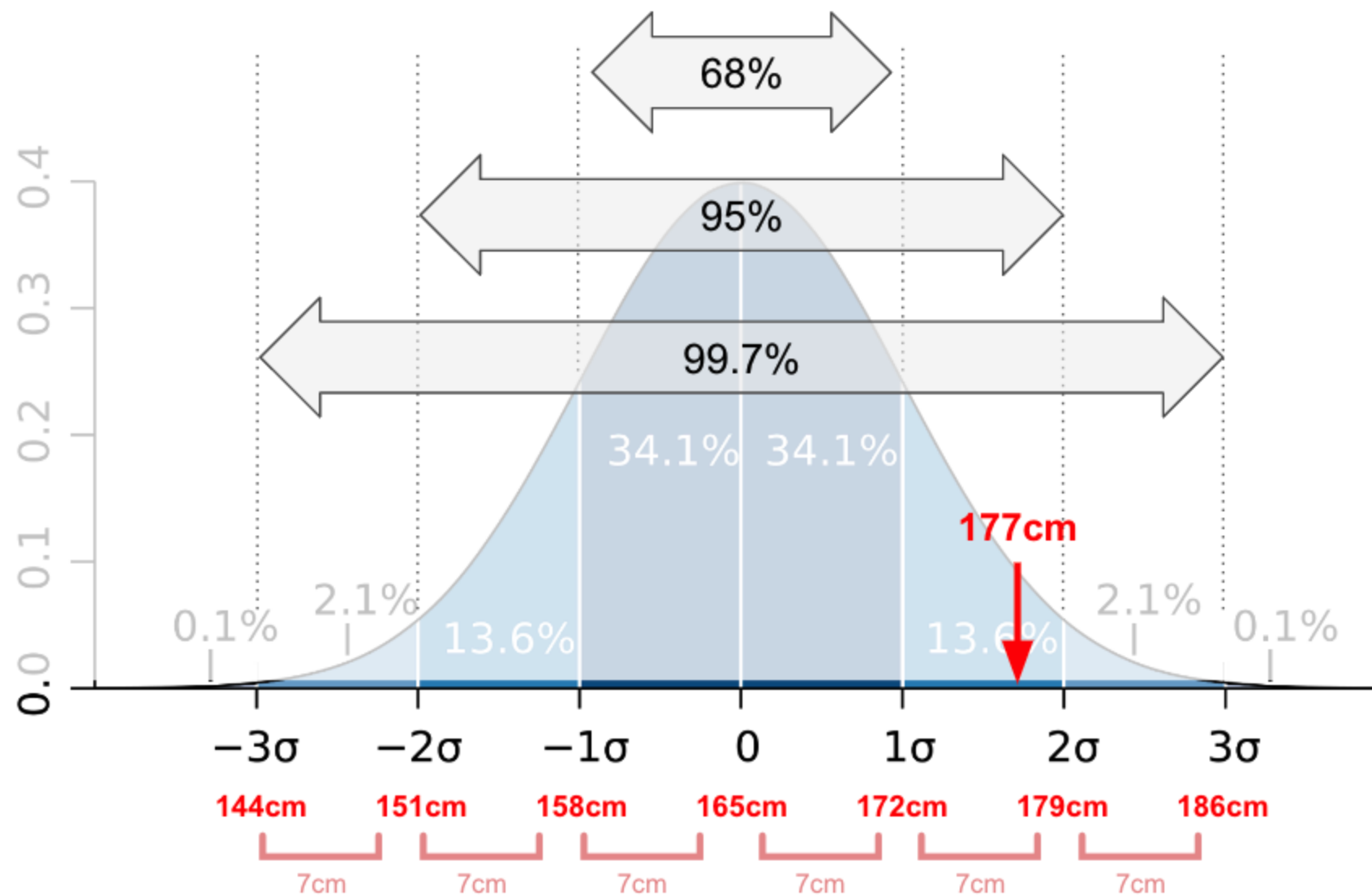
先生の身長は177cm ➡ 標準化すると？

$$\begin{aligned}\text{先生の } Z \text{ 得点} &= \frac{177 - 165}{7} \\ &= 1.71\end{aligned}$$

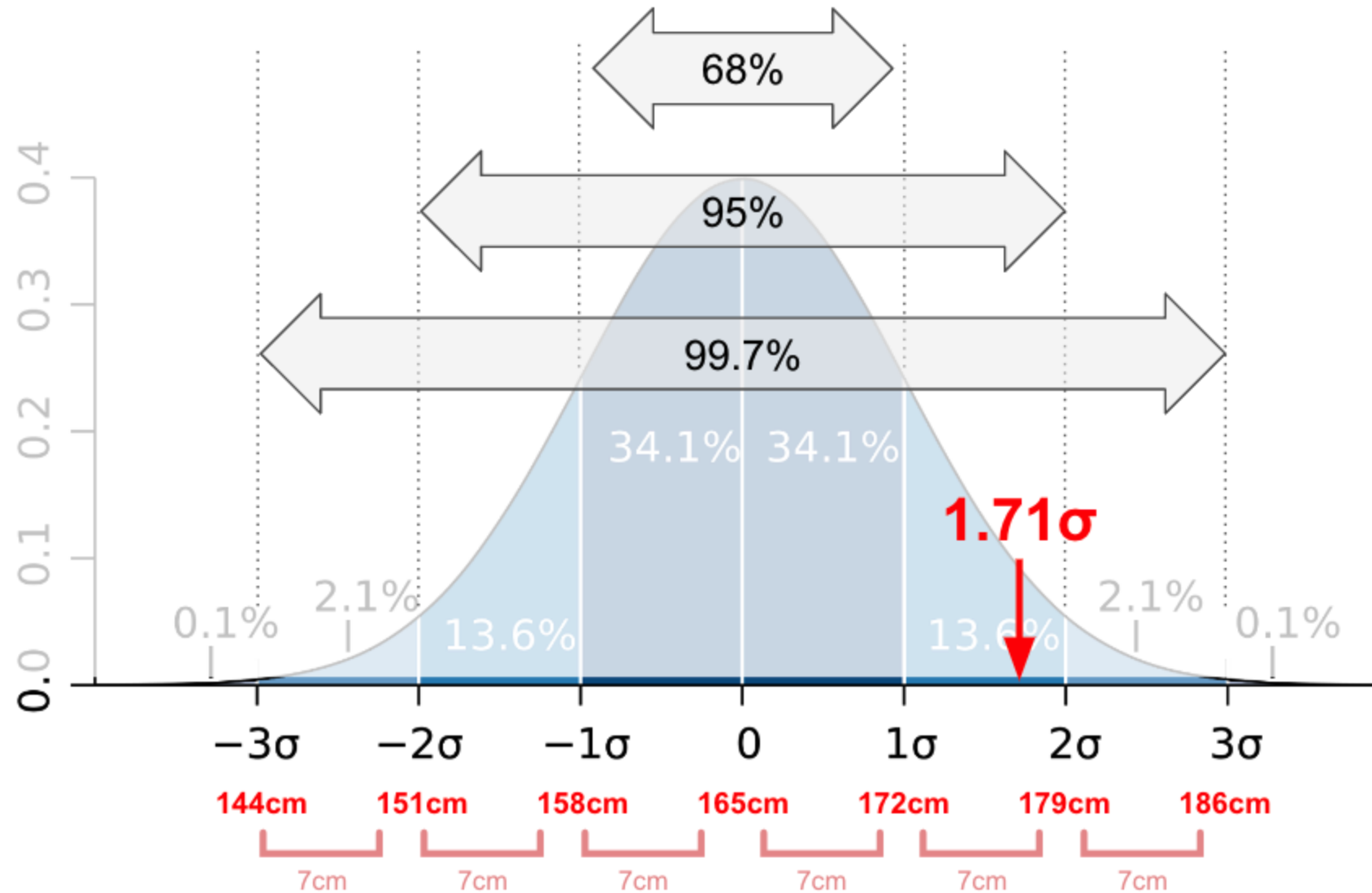


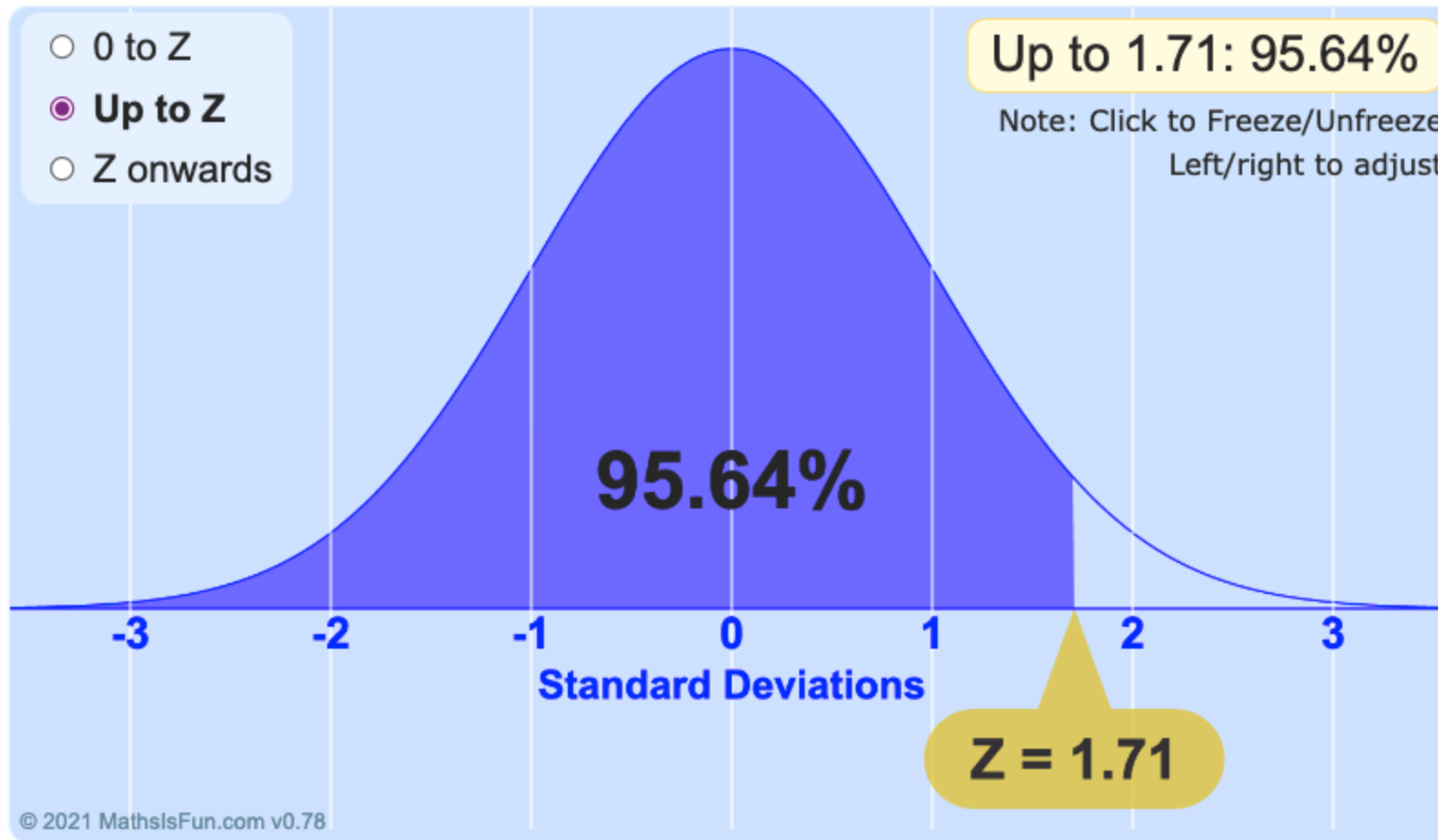






すなわち





<https://www.mathsisfun.com/data/standard-normal-distribution-table.html>

Topic #3: 偏差値



偏差値は、テストや試験の結果を分かりやすく比較するための指標です。



すなわち

偏差値は、あなたのスコアが他の人々と比べてどれくらいの位置にあるかを示す数値です。



それって変動係数と一緒にじゃないの？



You are right! But the difference is...

偏差値のスケールでは、**50**が平均値とされ、標準偏差が**10**とされています。

偏差値の計算:

$$\text{偏差値} = Z\text{得点} \times 10 + 50$$

先ほどの先生の身長は177cmだと、偏差値は？

$$\text{偏差値} = 1.71 \times 10 + 50 = 67.1$$

Excel playground

名前	性別	身長 (cm)	② 偏差	③ 偏差二乗	⑨ Z値 (身長- μ)/ σ			
Ran Komaba	Male	165						
Towa Tanahashi	Male	178						
Tomoki Wada	Male	175						
Yusuke Aono	Male	174						
Kensuke Wakabayashi	Male	167						
Masakazu Tamagawa	Male	175						
Makoto Kishino	Male	165						
Kuuta Uchida	Male	171						
Jo Yamaguchi	Male	165						
Sota Kobori	Male	170						
	① 平均($\bar{x} \cdot \mu$)→		④ 合計→					

⑤ Nは? →

⑥ 分散 (④/⑤) →

⑦ 標準偏差 σ (分散の平方根) →

⑧ 変動係数 ($\sigma \div \mu$) →

EXCEL Hints

(カッコの中の英数字はデータによって異なる)

平均 = AVERAGE(D4:D30)

合計 = SUM(E4:E30)

偏差 = C4-\$C\$31 (ドル縛り)

偏差二乗 = D4^2

分散の平方根 = SQRT(J5)

Nは? = COUNTA(B4:B30)