

平均に騙されるな！

吉倉劇場(4回目)

2019年9月17日

はじめに

世の中はデータであふれています。

データが何を語っているかを知り、判断をするために指標を使います。

平均値、最大値、中央値。。。

様々な指標がありますが、各指標の意味を理解しておかないと、
誤った判断につながる可能性があります。

今回は、広く使われる「平均値」について理解を深めてみましょう。

何が読み取れますか？

テストを行いました。A組・B組・C組があり、XくんのいるA組は30人います。
平均点は40点でした。

【質問】

1. A組は40点を取った人数が多い？
2. A組は40点以上を取った人は半分以上いる？
3. A組の成績はB組・C組と比べて、良い方？

平均値の定義

- ・ n 個 ($n \geq 1$)の数値があり、合計を S とする。
- ・平均値は 以下の式で求められる。

$$\text{平均値} = S(\text{n個の数値の合計}) \div n$$

<例> 5, 10, 25の平均値は？

$$S = 5 + 10 + 25 = 45$$

$$n = 3$$

$$\text{平均値} = 45 \div 3 = 15$$

何が読み取れますか？（回答）

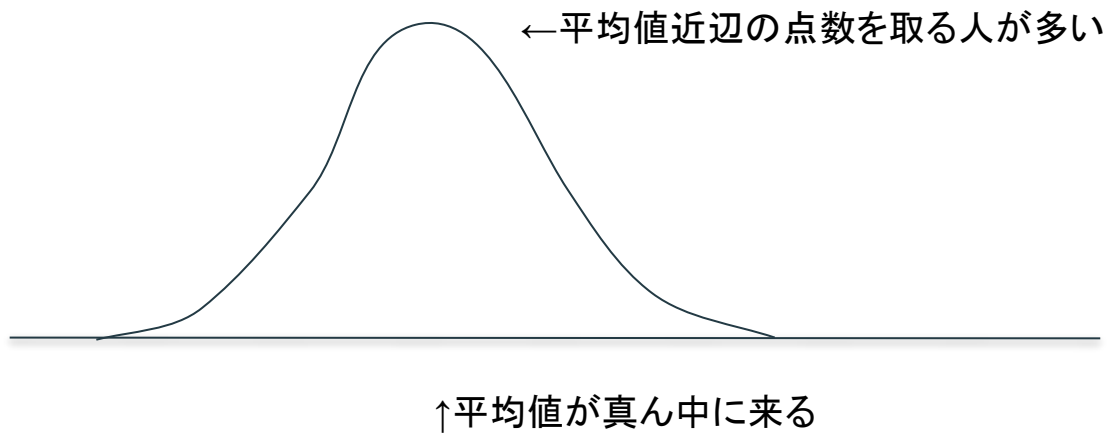
テストを行いました。A組・B組・C組があり、XくんのいるA組は30人います。
平均点は40点でした。

→30人の点数の合計が1200点であることはわかる。

【質問】

1. A組は40点を取った人数が多い？→ わからない
2. A組は40点以上を取った人は半分以上いる？→ わからない
3. A組の成績はB組・C組と比べて、良い方ですか？→ わからない

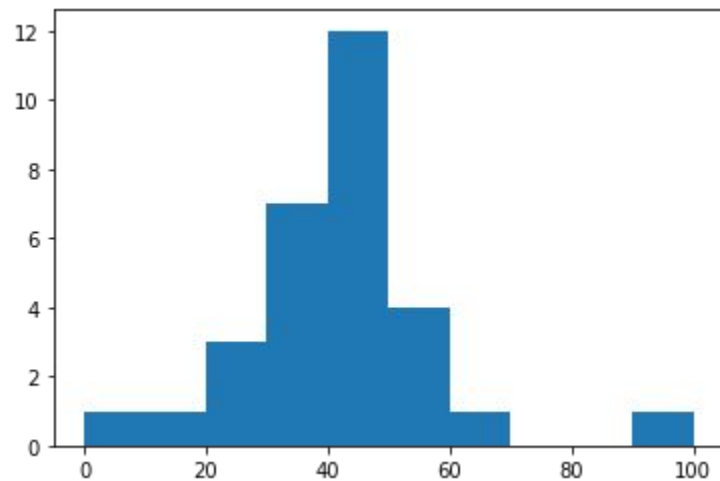
こんな分布をしていると想像しちゃう？



平均40点の分布①適度にばらけ

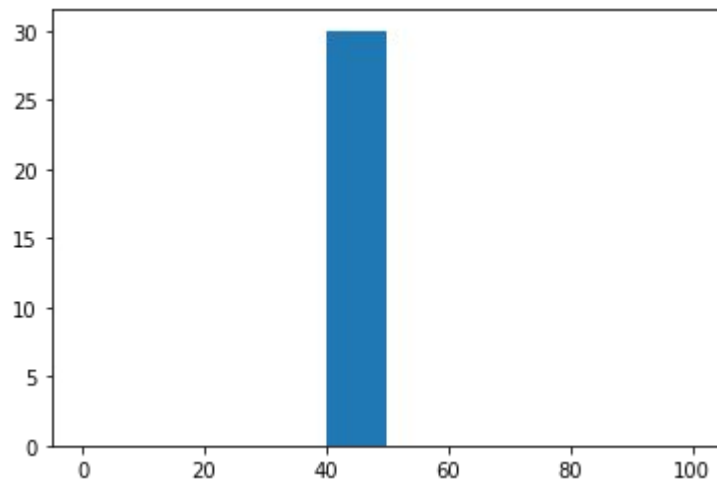
それなりにばらけている。

95,63,59,56,53,52,49,46,44,43,43,40,40,40,40,40,40,40,39,36,35,33,32,31,30,25,21,21,11,3



平均40点の分布②全部同じ

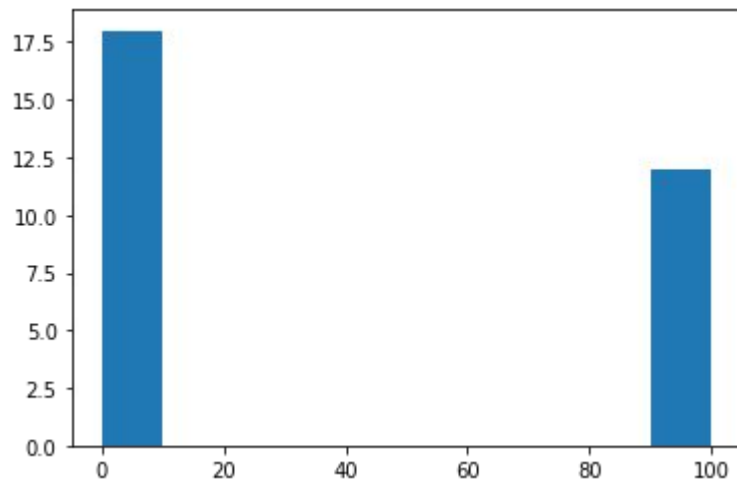
全員同じ点数。

[illegible]

平均40点の分布③両端に振れる

両端に振れている。100点と0点しかない。

100,100,100,100,100,100,100,100,100,100,100,100,100,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0



他の指標

平均点(40点)と母数($n=30$)では、判断できることは少ないです。

※なお、平均値のみで、母数も出ていない場合は、判断できることがもっと少なくなります。

なので、他の指標を出してみましょう。

最大値・最小値の定義

文字通りです。

最大値 : n 個の数値のうち、最大の値

最小値 : n 個の数値のうち、最小の値

中央値の定義

点数順に並べて、中央に来た人の値。
母数が偶数の場合は、中央の2つの値を平均する。

<例>

3,6,10,12,12の中央値

→ 母数5なので3番目の値を取る。中央値は10

3,6,10,12,12,18の中央値

→ 母数6なので3番目と4番目の平均値を取る。中央値は11

指標を並べてみます

並べてみると少し違いが見えてきますね。

	①	②	③	備考
母数	30	30	30	①～③同じ
平均値	40	40	40	①～③同じ
最大値	95	40	100	
中央値	40	40	0	
最小値	3	40	0	

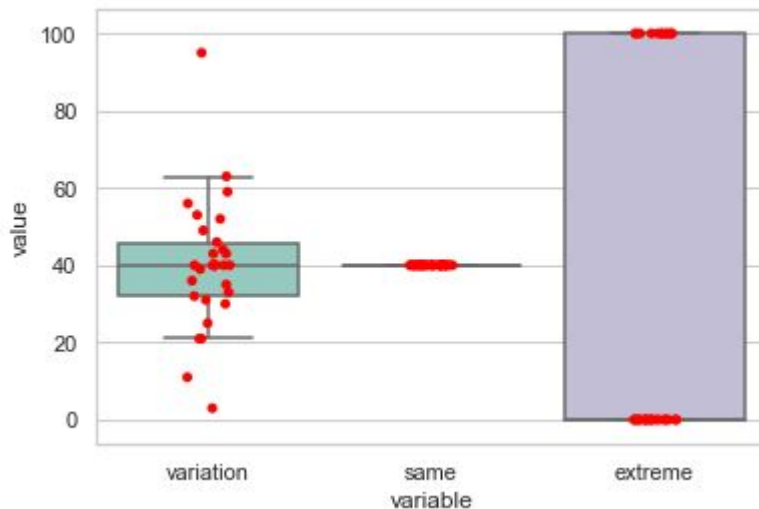
箱ひげ図 (BoxPlot)

データのばらつきをわかりやすく表現するための統計図。

最大値、中央値、最小値に加えて、第1四分位点、第3四分位点、四分位範囲を表す。

※自分で調べてみてね！

- ・第1、3四分位点
- ・四分位範囲



まとめ

- ・平均値だけだと、読み取れないことが多い
- ・せめて最大値・中央値・最小値は欲しい
- ・指標が並ぶと、見えるものは増えてくる
- ・ヒストグラムや箱ひげ図があると、なお良し

おしまい

次回(以降)予告

- ・「アナタとワタシの秘密の通信」(暗号化)
- ・「あなた。。。いったい誰？」(認証)