Day 5 of #100daysofmathandstats:

Real use cases for Estimates of Variability terms and implementation in R and Python

By Harsh Kathiriya

Outline

- Importance of terms for estimates of variability
- R Implementation with output
- Python Implementation with output

Standard deviation in healthcare

Standard deviation is widely used by insurance analysts and actuaries in the healthcare industry.

For example:

- Insurance analysts often **calculate the standard deviation of the age of the individuals** they provide insurance for so they can understand how much variation exists among the age of individuals they provide insurance for.
- Actuaries calculate standard deviation of healthcare usage so **they can know how much variation in usage to expect in a given month, quarter, or year.**

Other Use cases of Standard Deviation

- Real Estate
- Weather forecasting
- Human Resources
- Marketing

Interested to know how SD is used in these domains? Let's discuss!
DM me or Comment in post.

Mean absolute deviation (L1-norm/ Manhattan norm)

The mean of the absolute values of the deviations from the mean.

$$\text{Mean absolute deviation} = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n}$$

Real life Example of MAD

Everyone enjoys posting pictures on instagram or any social media. For example, Here's how many "likes" the past 5 pictures uploaded by xyz user:

10, 15, 20, 30, 32

Now what if you wants to find how far your posts from the average likes?
On the next page, we will see how to calculate it.

Real life Example of MAD (CONTD…)

1. Calculate the mean
   a. Mean = 10 + 15 + 18 + 30 + 22 / 5 = 19
2. Calculate the distance between each data point and mean

Data point	Abs distance
10	9
15	4
18	1
30	11
22	3

Real life Example of MAD (CONTD…)

3. Adds the distances together

   a. $9 + 4 + 1 + 11 + 3 = 28$

4. MAD = 28 / 5 ~ 5.6

So, each post is about 5.6 likes away from mean.

Percentile (Quantile)

Everyone knows the most common use case of this which is in calculating the results/rank for the competitive exams like CAT, GRE, School Exams, Board Exams etc.

Interquartile range (IQR)

The main use case of this is in boxplot and to find the outliers in the data.

Please visit the following link to get more details.

https://www.statology.org/outliers-real-life-examples/

R Implementation with output

```
> data <- datasets::mtcars
> # standard deviation
> stats::sd(data$mpg)
[1] 6.026948
>
> # variance
> # variance = sd ^ 2
> (stats::sd(data$mpg))^2
[1] 36.3241
>
>
> # mean absolute deviation from mean
> stats::mad(mtcars$mpg,
+            center = mean(mtcars$mpg))
[1] 6.37518
>
> # mean absolute deviation from median
> stats::mad(mtcars$mpg,
+            center = median(mtcars$mpg))
[1] 5.41149
>
> #range
> range(mtcars$mpg)
[1] 10.4 33.9
> max(mtcars$mpg) - min(mtcars$mpg)
[1] 23.5
>
> # percentile
> stats::quantile(data$mpg)
    0%    25%    50%    75%   100%
10.400 15.425 19.200 22.800 33.900
>
> # 25th 50th and 95h percentile
> stats::quantile(data$mpg, c(.25,.5,.95))
   25%    50%    95%
15.425 19.200 31.300
>
> # IQR
> Q3 <- stats::quantile(data$mpg, c(.75))
> Q1 <- stats::quantile(data$mpg, c(.25))
> IQR <- Q3 - Q1
> print(IQR)
   75%
7.375
>
> # with in-built function
> stats::IQR(mtcars$mpg)
[1] 7.375
>
```

Python Implementation with output

```
import pandas as pd
data = pd.read_csv('/content/mtcars.csv')
# standard deviatiion
print('SD: ',data['mpg'].std())

# variance
# variance = sd ^ 2
print('Variance: ',data['mpg'].std() ** 2)

# mean absolute deviation
print('MAD:  ', data['mpg'].mad())

# Range
print('Range:   ',max(data['mpg']) - min(data['mpg']))

# Percentile
print('75th percentile:',data.mpg.quantile(0.75))

# IQR
Q3 = data.mpg.quantile(0.75)
Q1 = data.mpg.quantile(0.25)
IQR = Q3 - Q1
print('IQR: ',IQR)
```

```
SD:  6.026948052089105
Variance:  36.32410282258065
MAD:   4.714453125
Range:    23.5
75th percentile: 22.8
IQR:  7.375
```

Thank you

Github Link: https://github.com/harsh9898/100daysofstatandmath

Don’t forget to post your queries or feedbacks on the post.

Share or like for the benefit of others.