# Random Variables

Sepehr Akbari

2024-02-26

made with quarto :))))

## Problem 1

Identify whether the following random variables are discrete or continuous.

(a) X represents the number of centimeters of rain recorded at a weather station in a month.

(b) Y represents the number of tweets by a celebrity on a given day.

### Answer

(a) **Continuous**: as the amount of rain can take on any value within a continuous range throughout the month.

(b) **Discrete**: as the number of tweets is whole, countable values. It's impossible for a celebrity to post "half a tweet" or any fraction of a tweet. They can only post 0 tweets, 1 tweet, 2 tweets, and so on, leading to a finite set of possible outcomes.

## Problem 2

The following table shows star ratings (0-5) for a product on Amazon.

| Stars | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Count | 22 | 73 | 84 | 101 | 144 | 152 |

(a) Use this data to construct a discrete probability distribution for a random variable X representing the star rating of a randomly-selected reviewer.

(b) What is the probability that a randomly-selected reviewer gives a rating of at least 3 stars?

(c) Determine the mean and standard deviation of X.

### Answer

(a) Using R:

```
stars <- c(0,1,2,3,4,5)
counts <- c(22,73,84,101,144,152)

counts / sum(counts)
```

```
[1] 0.03819444 0.12673611 0.14583333 0.17534722 0.25000000 0.26388889
```

(b) Using R:

```
0.17534722 + 0.25000000 + 0.26388889
```

```
[1] 0.6892361
```

(c) using R:

Mean:

```
p <- counts / 576
sum(stars * p)
```

```
[1] 3.263889
```

Standard Deviation (sd):

```
mu <- sum(stars * p)
sqrt(sum((stars - mu) ^ 2 * p))
```

```
[1] 1.494137
```

## Problem 3

Determine the missing probability in the following discrete probability distribution.

| x | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|
| P(x) | .22 | .31 | ?? | .25 | .05 |

Compute the mean, variance, and standard deviation of this random variable.

### Answer

The missing data:

$$1 - (0.22 + 0.31 + 0.25 + .05) = 0.17$$

Mean:

```
x <- c(10,15,20,25,30)
p <- c(0.22, 0.31, 0.17, 0.25, 0.05)

sum(x * p)
```

```
[1] 18
```

Variance:

```
mu <- sum(x * p)
sum((x - mu) ^ 2 * p)
```

```
[1] 37
```

Standard deviation (sd):

```
sqrt(sum((x - mu) ^ 2 * p))
```

```
[1] 6.082763
```