Statistics and Mathematics Notes

Abdullah Al Mahmud

2025-08-13

Contents

About

This websites contains notes on statistics and mathematics. These are chiefly meant for being used as reference materials, which, together with the lectures slides, make for a more matured system of learning.

- See the lecture presentations here

This website is created with the help of `Rstudio IDE` using the `R`package `bookdown`.

**If you find any mistakes or have any suggestions, please let me know

You can learn more about me here and about my writings on statistics, data science, and linux here.

Part I

Statistics

Statistics

Scales of Measurements

Interval Scales

Examples

- Temperature (Celsius scale)
- Dates (AD)
- Location in Cartesian coordinates
- Direction measured in degrees

Why Locations in Cartesian coordinates are interval data

Locations in Cartesian coordinates are **interval data** because:

In short, — $x = 10$ ***is not "twice as far to the right" as*** $x = 5$

Detailed

1. **Equal intervals have equal meaning**

   - In the Cartesian system, the difference between two x-coordinates (or two y-coordinates) represents the same physical distance, regardless of where it is measured.
   - For example, moving from $x = 2$ to $x = 5$ is a change of 3 units, which is the same "amount of movement" as going from $x = 20$ to $x = 23$.

2. **Arbitrary origin**

   - The $(0, 0)$ origin is chosen for convenience, not because it represents an absolute "zero" location.
   - A point at $x = 0$ does not mean "no position" — it's just the reference point we picked.

- Because the zero is arbitrary, coordinates behave like interval data, not ratio data.

3. **Meaningful differences, not meaningful ratios**

   - You can meaningfully talk about differences (Δx or Δy) to measure displacement.
   - But a ratio like "this point's x-coordinate is twice that of another" is meaningless — $x = 10$ is not "twice as far to the right" as $x = 5$ in an absolute sense, because the zero point is arbitrary.

So, **Cartesian coordinates** $\rightarrow$ **interval scale Distances between points** (via Euclidean distance) $\rightarrow$ **ratio scale** (because distance has a true zero).

Why Date is an interval scale

In short, ***A person born in 1940 doesn't always have double age of a person born in 1950. It's so only in 1960.***

Detailed

Date is an **interval scale** because it has a meaningful order and equal intervals between measurements, but it lacks a true zero point.

- **Order:** Dates can be placed in a specific order (e.g., January 1, 2024, comes before January 2, 2024).
- **Equal Intervals:** The duration between any two dates is consistent and measurable. The interval between May 1st and May 10th is exactly 9 days, just as the interval between November 1st and November 10th is 9 days. This allows for meaningful subtraction (e.g., "how many days have passed?").
- **No True Zero:** The starting point of a calendar system (e.g., the year 0 in the Gregorian calendar) is an arbitrary convention, not an absolute absence of time. You can't say that a year in the 2nd century A.D. is "twice as old" as a year in the 1st century A.D. in a meaningful ratio sense, because the scale doesn't start from a true zero.
- **Meaningful Subtraction, Not Division:** You can meaningfully subtract dates. For example, a person born in 1940 is 10 years older than a person born in 1950 in any given year ($1950 - 1940 = 10$). However, you cannot meaningfully divide them. The statement "1940 is double 1950" is only true in an arbitrary mathematical sense on a single date, not as a fundamental property of the ages.

Part II

Probability

Chapter 1

Random variable

1.1 List of probability distributions

1.2 Discrete distributions

Probability Mass Functions (PMF)

1. $P(x) = \frac{1}{14}(a + 2x); x = -3, -2, -1, 0, 1, 2, 3$
2. $P(x) = k(x - 2); x = 3, 4, 5, 6, 7, 8$
3. $P(x) = \frac{x-1}{k}; x = 2, 3, 4, 5$
4. $P(x) = \frac{3-|4-x|}{k}; x = 2, 3, 4, 5, 6$
5. $p(x) = \frac{x+4}{30}; x = 0, 1, 2, 3, 4$
6. $P(x) = \frac{2x+k}{56}; x = -3, -2, -1, 0, 1, 2, 3$
7. $P(x) = \frac{x+1}{k}; x = 1, 2, 3, 4$

1.2.1 Continuous

Probability Density Functions (PDF)

1. $f(x) = 2x; 0 < x < 1$
2. $f(x) = \frac{1}{30}(3 + 2x); 2 < x < 5$
3. $f(x) = ax^2; 0 < x < 4$
4. $f(x) = kx^2 + kx + \frac{1}{8}; 0 < x < 8$
5. $f(x) = kx; 0 < x < 4$
6. $f(x) = 3x^2; 0 \leq x \leq 1$
7. $f(y) = k(3y + 5); 1 < y < 5$
8. $f(z) = \frac{2}{9}(3z - z^2); 0 \leq x \leq 3$

1.2.1.1 Joint PDF

1. $$f(x, y) = 8xy; 0 < x, y < 1$$

2. $$f(x, y) = \frac{3}{2}(x + y); 0 < x, y < 1$$

3. $$f(x, y) = 4x(1 - y); 0 < x, y < 1$$

4. $$f(x, y) = 6xy^2); 0 < x, y < 1$$