

Handwritten mathematical formulas on a chalkboard:

- $f(x, \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $F(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left[\frac{\ln(x) - \mu}{\sigma \sqrt{2}}\right]$
- $\mu_1 = e^{\mu + \sigma^2/2}$, $\mu_2 = e^{\frac{1}{2}\mu + \frac{1}{4}\sigma^2}$, $\mu_3 = e^{\frac{3}{2}\mu + \frac{3}{2}\sigma^2}$, $\mu_4 = e^{\frac{5}{2}\mu + \frac{5}{2}\sigma^2}$
- $\mu_k = e^{k\mu + \frac{k^2}{2}\sigma^2}$
- $E(x) = e^{\mu + \frac{\sigma^2}{2}}$, $\mu = \ln(E(x)) - \frac{1}{2} \ln\left(1 + \frac{\operatorname{Var}(x)}{E(x)^2}\right)$
- $g(x) = \int_{-\infty}^{\infty} x f(x) dx$
- $g(k) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \phi\left(\frac{-\ln(k) + \mu + \sigma^2}{\sigma}\right)$

Variable

In Exploratory Data Analysis (EDA), understanding the types of variables is crucial as it dictates the kind of analysis and visualizations you can perform. Variables are typically classified into the following types:

1. Based on Measurement Level

a. Quantitative Variables (Numerical Variables)

- Represent numeric values and allow arithmetic operations.
- **Types:**
 - 1. Continuous Variables:**
Values can take any number within a range (e.g., height, weight, temperature).
Example: 73.5 kg, 23.8°C.
 - 2. Discrete Variables:**
Take whole number values (e.g., number of students, count of cars).
Example: 3 students, 7 cars.

b. Qualitative Variables (Categorical Variables)

- Represent categories or groups, not numbers.
 - **Types:**
 1. **Nominal Variables:**

Categories without a specific order.

Example: Colors (Red, Blue, Green), Gender (Male, Female).
 2. **Ordinal Variables:**

Categories with a meaningful order, but differences between them are not defined.

Example: Education level (High School < Bachelor's < Master's).
-

2. Based on Data Collection

a. Independent Variables (Predictor Variables)

- Variables used to predict or explain outcomes.
- Example: Age, Hours Studied in predicting exam scores.

b. Dependent Variables (Response Variables)

- The outcome or result being studied.
 - Example: Exam Score depends on Hours Studied.
-

3. Based on Behavior in Analysis

a. Time Variables

- Variables that change with time.
- Example: Year, Month, Timestamp.

b. Identifier Variables

- Unique identifiers for rows or observations.
- Example: Customer ID, Order ID.

4. Based on Encoding

a. Binary Variables

- Variables with two possible values (0/1 or True/False).
- Example: Yes/No, Has Credit Card.

b. Dummy Variables

- Categorical variables converted into binary format for modeling.
 - Example: Gender (Male=1, Female=0).
-

Summary Table

Type	Sub-Type	Examples
Quantitative	Continuous	Height, Weight, Temperature
	Discrete	Number of Students, Count of Cars
Qualitative	Nominal	Colors, Gender, Blood Group
	Ordinal	Education Level, Satisfaction Rating
Behavior in Analysis	Time Variables	Year, Month, Timestamp
	Identifier Variables	Customer ID, Product Code
Encoding	Binary	Yes/No, Has Credit Card
	Dummy Variables	Male=1, Female=0

Visualization Suggestions:

- **Numerical Variables:** Histograms, Boxplots, Scatter Plots.
- **Categorical Variables:** Bar Charts, Pie Charts.
- **Time Variables:** Line Charts.
- **Binary Variables:** Stacked Bar Charts.