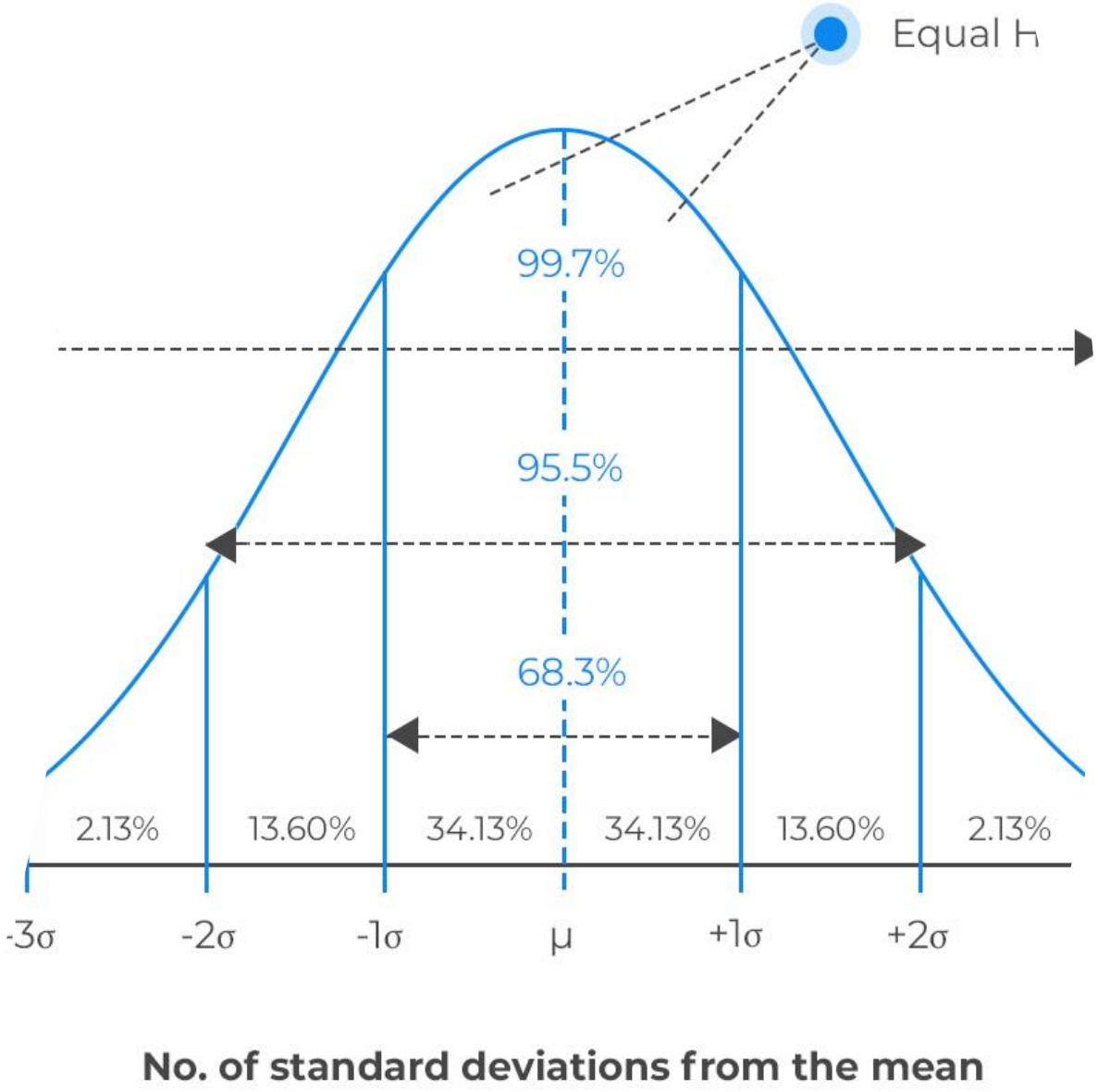

Probability &
Statistics

Gaussian Distribution



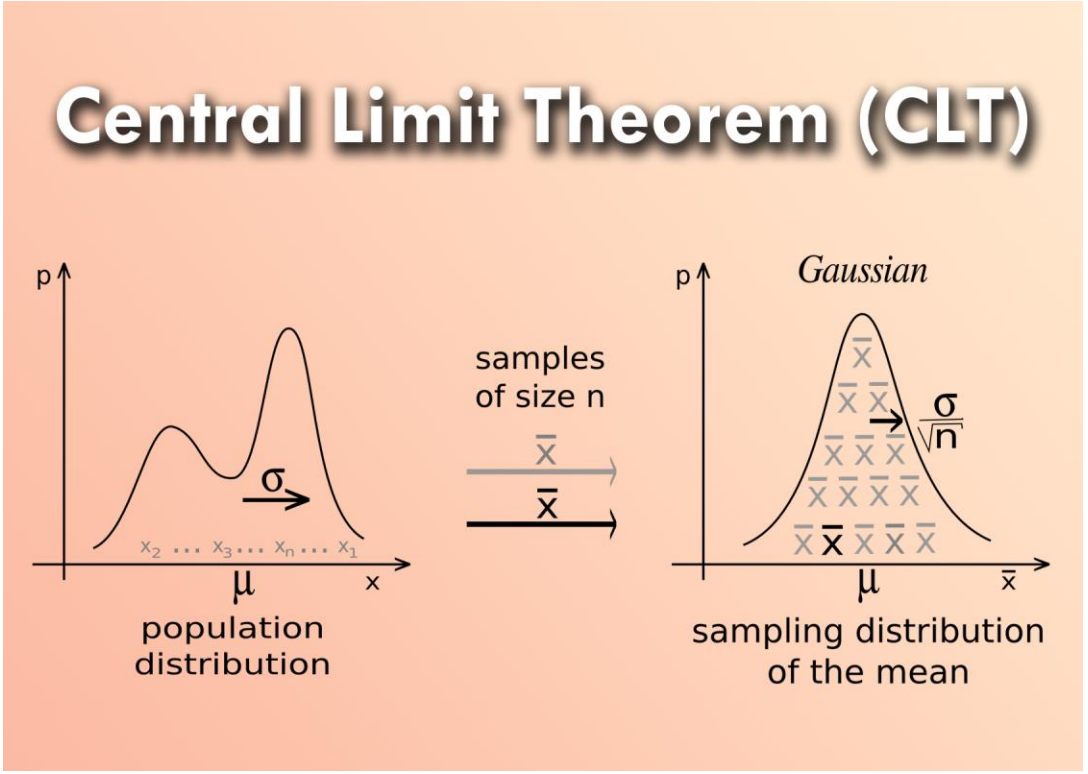
- **Gaussian (Normal) Distribution**
- Describes distribution of a **continuous variable** around a **mean**
- **Symmetric, bell-shaped** curve
- **Mean = Median = Mode**
- Defined by:
- **Mean (μ):** center
- **Standard deviation (σ):** spread
- **Empirical rule:**
- 68% within ±1σ
- 95% within ±2σ
- 99.7% within ±3σ
- Common examples:
- Heights, exam scores
- Measurement errors, noise
- Widely used in **statistics, ML, and data analysis**

Standard Normal Variate

- Standard Normal Variate & Standardization
- **Standard Normal Variate (Z):**
- A value showing how far a data point is from the mean
- Mean = **0**, Standard deviation = **1**
- **$Z = (x - \mu) / \sigma$**
- **Standardization of Data:**
- Process of converting data to **Z-scores**
- Centers data at **0** and scales it to **unit variance**

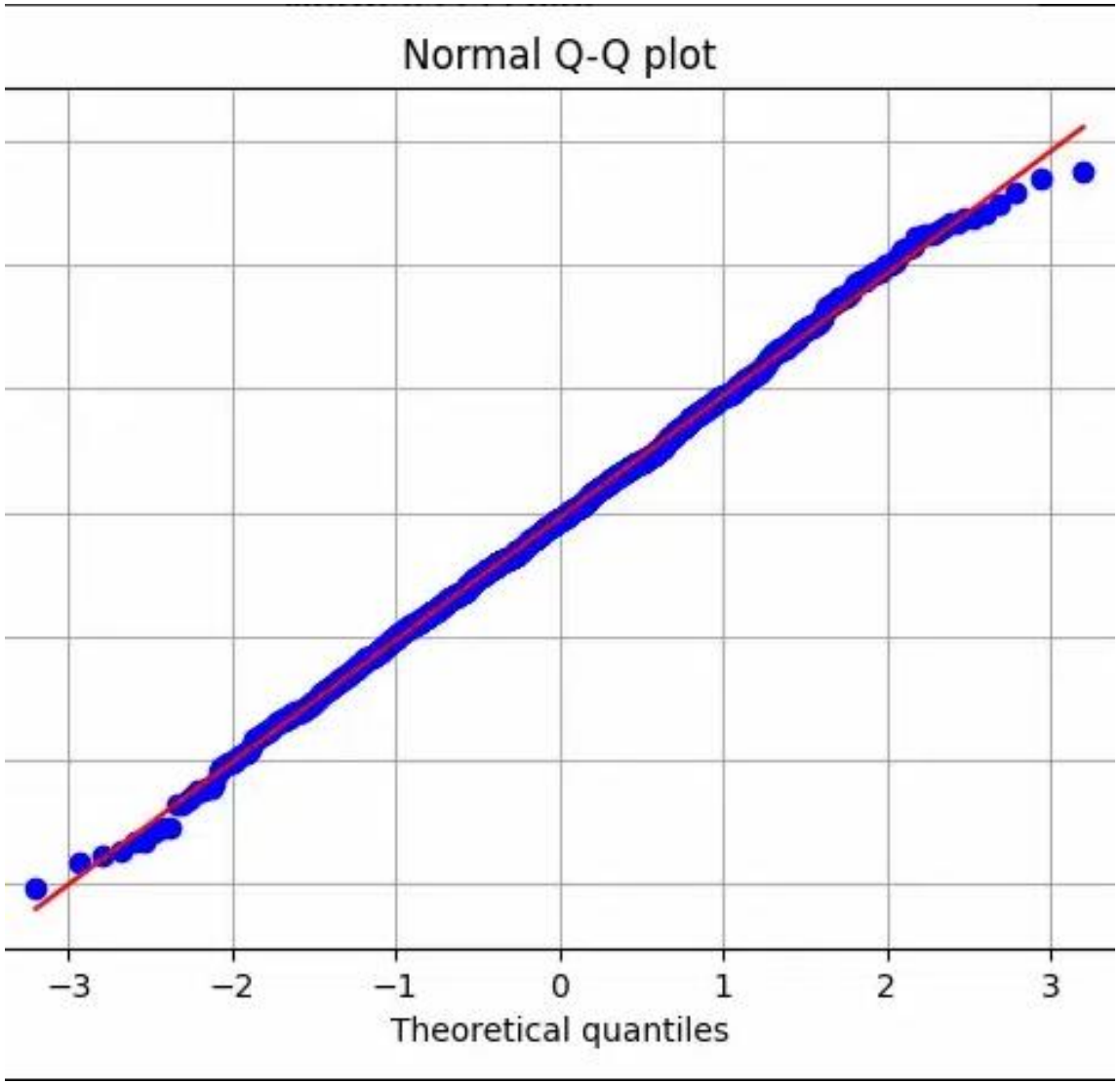
Central Limit Theorem

- Applies to **sampling distribution of the sample mean**
- For **large samples (n ≥ 30)**:
  - Distribution of sample mean → **approximately normal**
  - **Independent of population shape**
- **Mean = μ** (population mean)
- **Std. deviation = σ / √n** (standard error)
- Larger **n → less variability**
- Basis of **confidence intervals & hypothesis testing**



Quantile–Quantile (Q–Q) Plot

- Graphical tool to **compare a dataset with a theoretical distribution**
- Commonly used to **check normality**
- Plots:
  - **X-axis:** theoretical quantiles
  - **Y-axis:** sample quantiles
- **Interpretation:**
  - Points ≈ straight line → data fits distribution
  - S-shape → skewed data
  - Curved ends → heavy/light tails
  - Distant points → outliers
- More reliable than histogram for **distribution comparison**



Chebyshev's Inequality

- Applies to **any distribution** (normal or non-normal)
- Gives a **minimum guarantee** of data within **k standard deviations**
- Formula:
  $P(|X - \mu| < k\sigma) \geq 1 - 1/k^2$, $k > 1$
- Guaranteed coverage:
  - k = 2 → **≥ 75%**
  - k = 3 → **≥ 88.9%**
  - k = 4 → **≥ 93.75%**
- Requires only **mean (μ)** and **standard deviation (σ)**
- Conservative but **distribution-free**

ABOUT CHEBYSHEV'S THEOREM

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- X: Represents a random variable (your data).
- μ: Represents the population mean of X.
- σ: Represents the population standard deviation of X.
- k: Represents the number of standard deviations away from the mean. k must be greater than 1.
- $P(|X - \mu| \geq k\sigma)$: Represents the probability that a value of X will differ from the mean (μ by k standard deviations or more (in either direction). This is the probability of being outside the range.

Bernoulli Distribution

- Models a **single random experiment** with **two outcomes**
- Outcomes:
  - **Success (1)** with probability **p**
  - **Failure (0)** with probability **1 – p**
- **Probability Mass Function (PMF):**
  $$P(X = x) = p^x(1 - p)^{1-x}, x \in \{0,1\}$$
- **Mean:** E(X) = **p**
- **Variance:** Var(X) = **p(1 – p)**

Binomial Distribution

- Models the number of **successes** in **n independent Bernoulli trials**
- Each trial has:
- Two outcomes (Success / Failure)
- Constant probability of success **p**
- **Random variable (X):** number of successes
- **Probability Mass Function (PMF):**
  $$P(X = x) = C(n, x)\, p^{x}\, (1 - p)^{n-x}$$
- **Mean:** E(X) = **np**
- **Variance:** Var(X) = **np(1 – p)**
- Discrete distribution; **Bernoulli is a special case (n = 1)**

Comparison

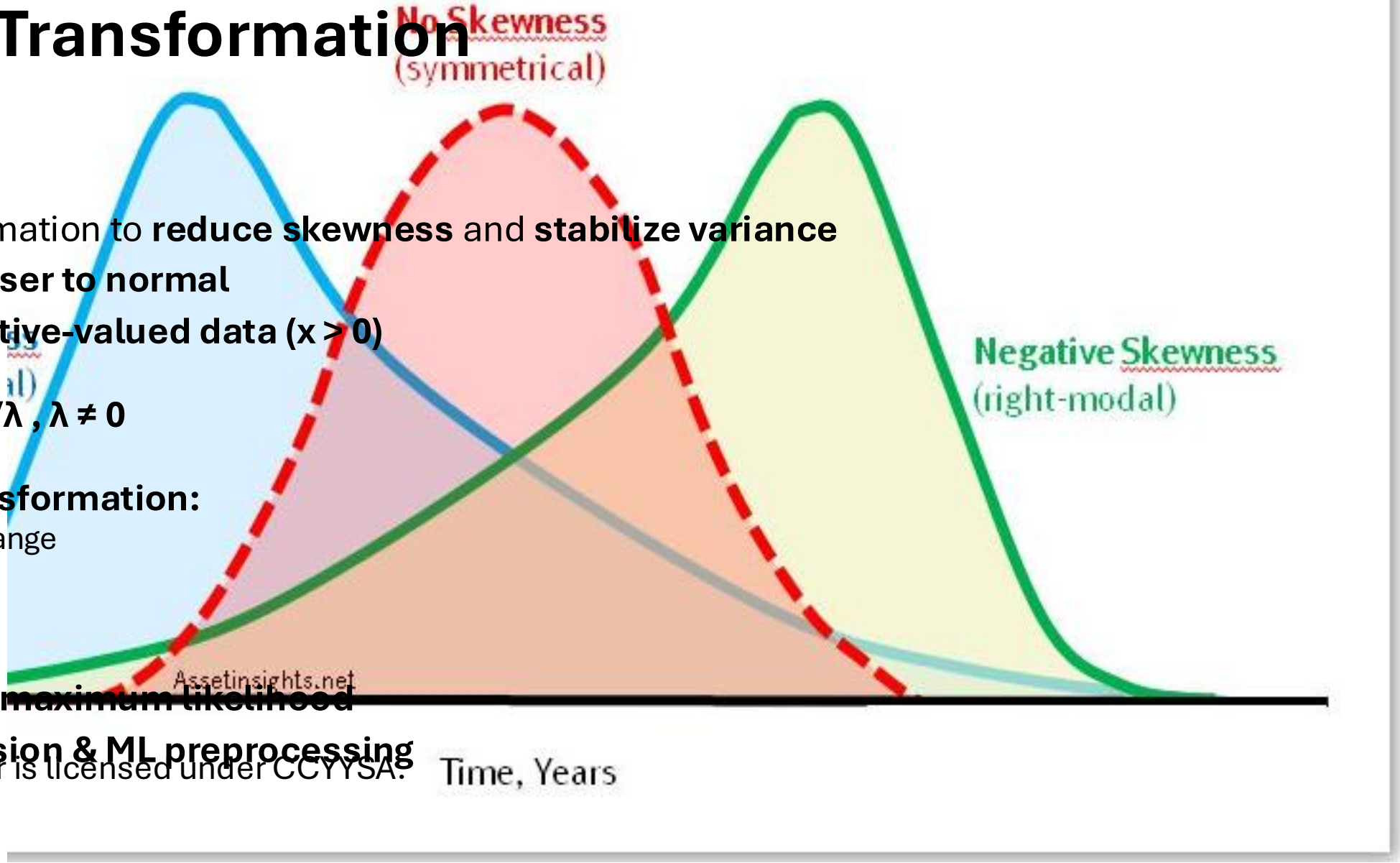
Feature	**Bernoulli Distribution**	**Binomial Distribution**
Number of trials	**Single trial (n = 1)**	**Multiple trials (n ≥ 1)**
Possible outcomes	Two (Success / Failure)	Counts number of successes
Random variable (X)	0 or 1	0, 1, 2, ..., n
Probability parameter	p	p
PMF	$P(X = x) = p^x(1-p)^{1-x}$	$P(X = x) = C(n,x)p^x(1-p)^{n-x}$
Mean	p	np
Variance	$p(1-p)$	$np(1-p)$
Relationship	Basic model	Sum of n Bernoulli trials
Example	Single coin toss	Number of heads in n tosses

Power Law / Pareto Distribution

- Models phenomena where **few large values dominate many small ones**
- Follows the **80–20 (Pareto) principle**
- **Right-skewed, heavy-tailed** distribution
- **PDF:**
  $f(x) = \alpha x_m^{\alpha} / x^{\alpha+1}$, $x \geq x_m$
- **Shape parameter (α):**
  - Small α → heavier tail
  - Large α → faster decay
- Mean exists if **α > 1**; variance if **α > 2**
- Common in **wealth, city sizes, networks**

Box–Cox Transformation

- Power transformation to **reduce skewness** and **stabilize variance**
- Makes data **closer to normal**
- Applied to **positive-valued data (x > 0)**
- **Formula:**
  $y(\lambda) = (x^{\lambda} - 1)/\lambda$, $\lambda \neq 0$
  $y(0) = \ln(x)$
- **λ controls transformation:**
  - λ = 1 → no change
  - λ = 0 → log
  - λ = 0.5 → √x
  - λ = −1 → 1/x
- λ chosen using **maximum likelihood**
- Used in **regression & ML preprocessing**



• Covariance

- Covariance & Correlation Measures
- Measures **direction of relationship** between two variables
- Formula:
  $Cov(X, Y) = \Sigma[(X - \mu_x)(Y - \mu_Y)] / n$
- Interpretation:
  - Positive → move in same direction
  - Negative → move in opposite direction
  - Zero → no linear relationship
- **Magnitude depends on units** (hard to interpret)

Pearson Correlation Coefficient (r)

- Measures **strength & direction of linear relationship**
- **Standardized covariance**
- Formula:
  $\mathbf{r = Cov(X, Y) / (\sigma_x \sigma_Y)}$
- Range: **−1 to +1**
  - +1 → perfect positive linear
  - −1 → perfect negative linear
  - 0 → no linear correlation
- Assumes **linearity & normality**

Spearman Rank Correlation (ρ)

- Measures **monotonic relationship using ranks**
- Non-parametric (no normality assumption)
- Formula:
  $\rho = 1 - [6\Sigma d^2 / n(n^2 - 1)]$
- Uses **rank differences (d)**
- Robust to **outliers & non-linear trends**