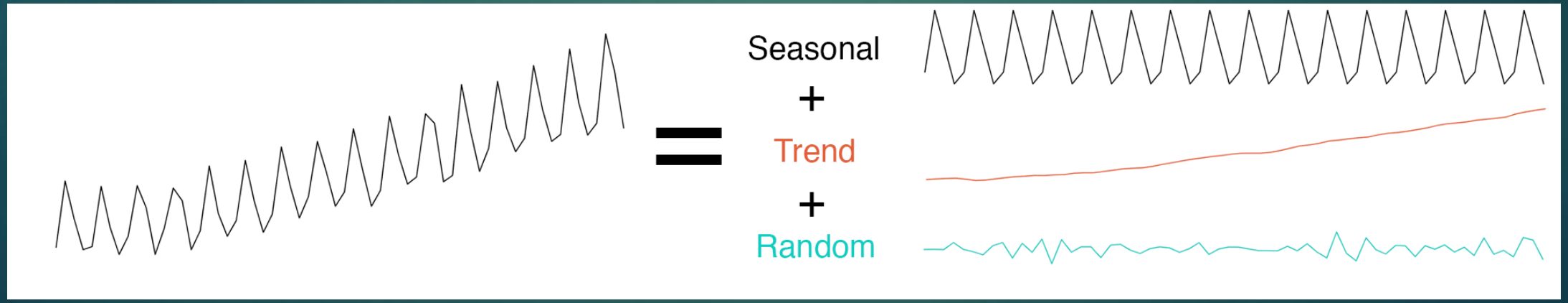
Time Series Anomaly Detection and Cause Analysis

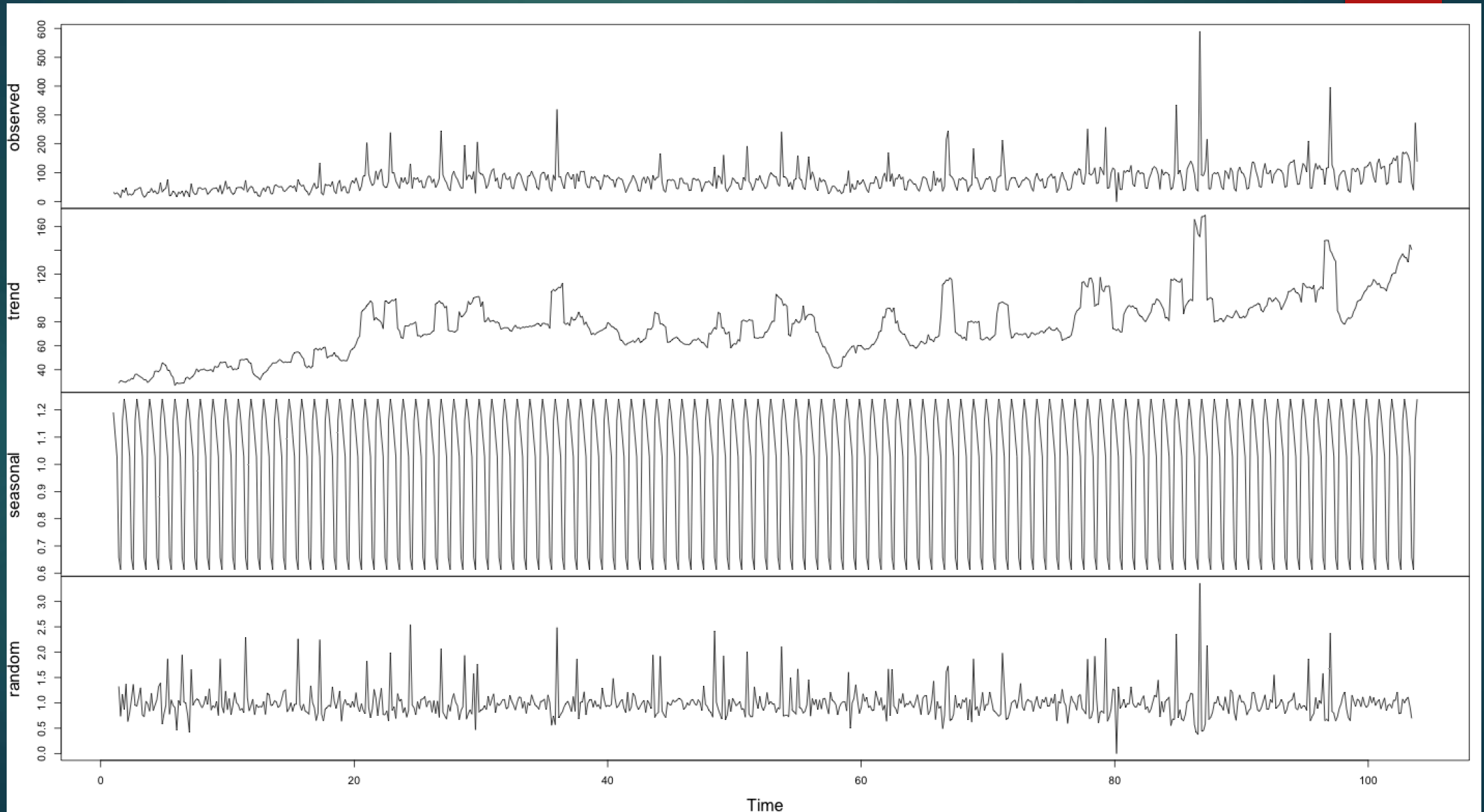
- VINOD ADWANI
- HARSHIL SHAH

Components of Time Series Data

- The Components of time series data are given below -
- Trend
- Seasonal
- Random

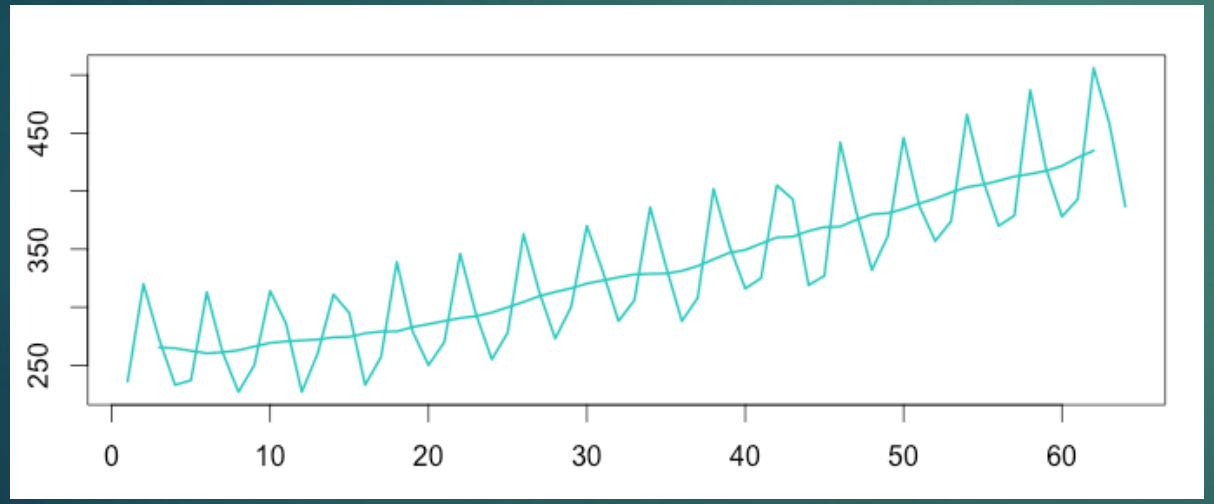


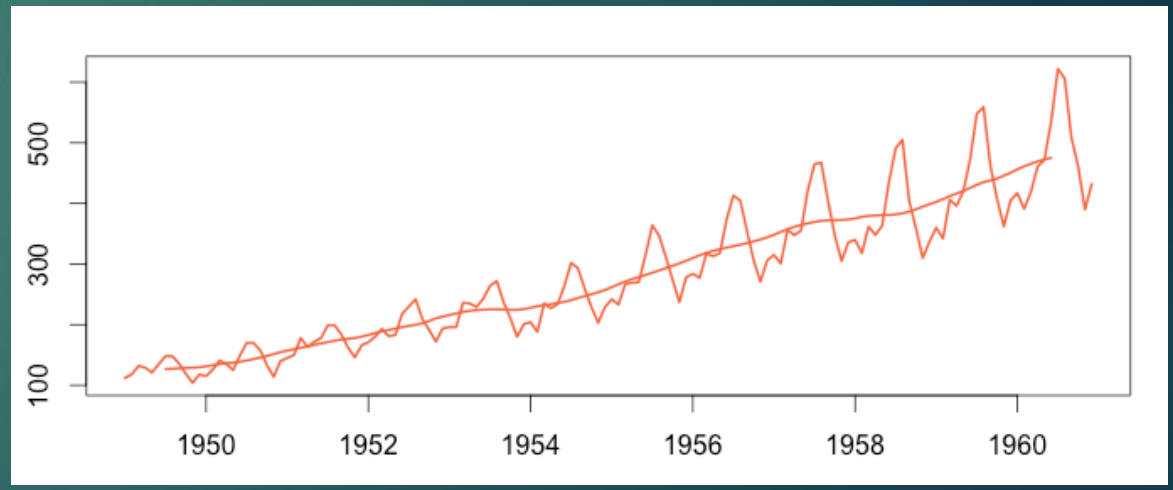
Example:


observed
0
100
200
300
400
500
600
trend
40
60
80
120
160
seasonal
0.6
0.7
0.8
0.9
1.0
1.1
1.2
random
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
20
40
60
80
100
Time

Models of Time Series:

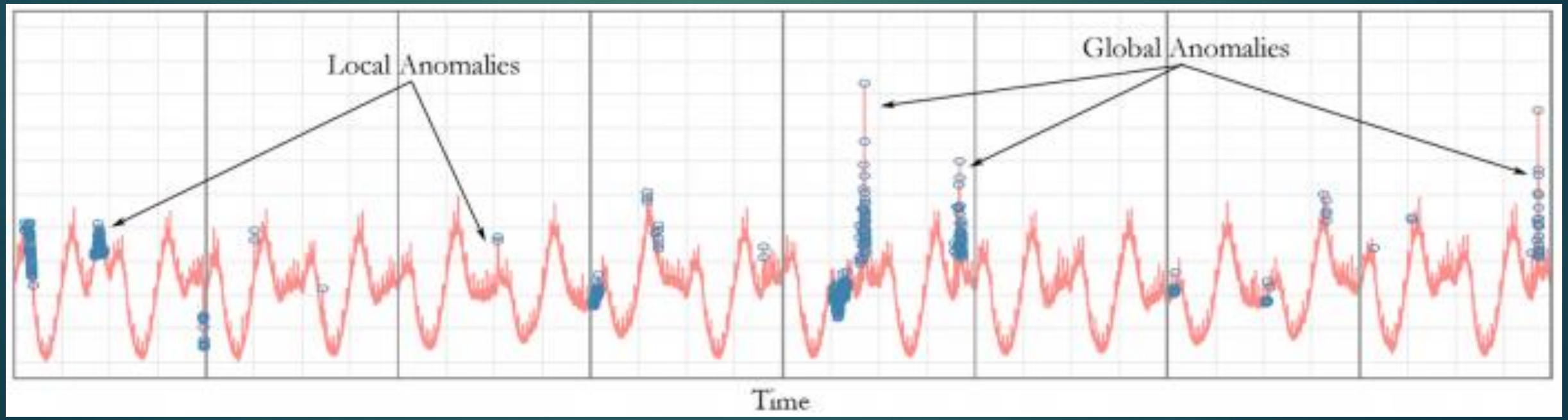
- **Additive Model**
- **Multiplicative Model**





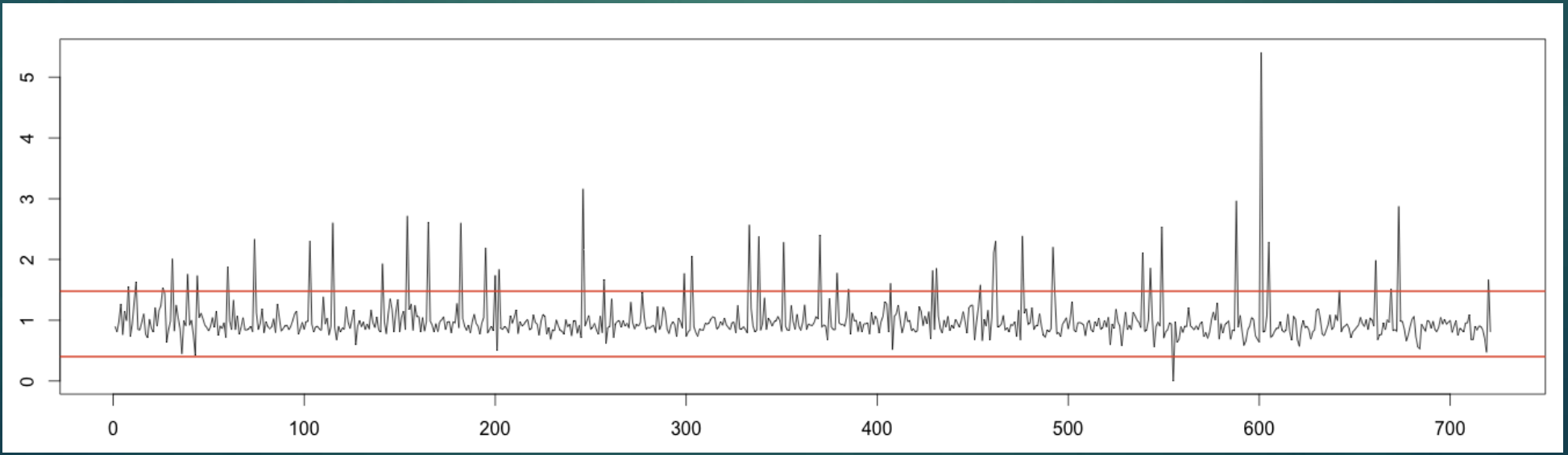
Types of Anomalies

- Global/local Anomalies
- Positive/Negative



How to detect Anomalies?

- Use time series decomposition to split your time series into three parts: **seasonal, trend and residue**
- ***Introduce some threshold for residue and you'll get an anomaly detection algorithm.***
- Trivial way to do this will be to use threshold as { μ + 4 * **σ** , μ - 4 * **σ** }



Better way!

- Generalized ESD test for detecting anomalies.
- Central idea – Use different thresholds for different points $\{ \mu + \lambda i * \boldsymbol{\sigma} , \mu - \lambda i * \boldsymbol{\sigma} \}$
- Using Seasonality information, it can be used to detect both global and local anomalies.
- To make it Robust, use robust statistical metrics(Median and median Absolute deviation)

Generalized ESD (Extreme Student Deviate) test

- Given the upper bound, *r*, the ESD test essentially performs *r* iterations. On ith iteration, it will find ith anomaly point.
- On ith iteration, it finds the point P_i with maximum value of z-score
- $\mathbf{Z}_i$ = max(Z) and compares it with thresold value λi.
- The Outliers are determined by finding the largest *i* such that $Z_i > \lambda_i$.

$$\lambda_i = \frac{(n-i)\, t_{p,\, n-i-1}}{\sqrt{(n-i-1+t^2_{p,\, n-i-1})(n-i+1)}} \qquad i = 1, 2, \ldots, r$$

$$p = 1 - \frac{\alpha}{2(n-i+1)}$$