Statistical Outlier Detection Notes

Outlier Detection Using Mean and Standard Deviation

To detect outliers in a dataset Δ, we use the mean and standard deviation:

- $\mu(\Delta)$: Mean of the data
- $\sigma(\Delta)$: Standard deviation of the data

Normal Range

The normal range is defined as:

$$\mu(\Delta) \pm 2\sigma(\Delta)$$

This means most data points (about 95% if normally distributed) are expected to lie within this range.

Outlier Condition

A value is considered an outlier if:

$$\Delta < \mu(\Delta) - 2\sigma(\Delta) \quad \text{or} \quad \Delta > \mu(\Delta) + 2\sigma(\Delta)$$

Python Equivalent

This logic can be implemented in Python as follows:

```
mu, std = delta.mean(), delta.std()
lower, upper = mu - 2*std, mu + 2*std
outliers = (delta < lower) | (delta > upper)
```

Notes

- This method assumes data is roughly normally distributed.

- Using 2σ captures approximately 95% of data points under a normal distribution.

- You can adjust the multiplier (e.g., 3σ) for stricter or looser thresholds.