

## 2. Đánh giá từng cột

Dựa trên các tiêu chí trên và kết quả phân tích trước, tôi đánh giá từng cột:

### Cột FLC

- **Thống kê:**
  - Số điểm: 1342
  - Trung bình: 7.74, Trung vị: 7.05
  - Skewness: 2.40 (lệch phải mạnh)
  - Kurtosis: 8.02 (đuôi rất nặng)
  - Shapiro-Wilk: p-value = 0.0000 (không chuẩn)
- **Ngoại lai:**
  - Z-score: 36 ngoại lai (2.68%) – Tỷ lệ thấp, hợp lý.
  - IQR: 0 ngoại lai – Có thể do khoảng IQR quá rộng với phân phối đuôi nặng.
- **Dữ liệu thiếu:** Không thiếu trong dữ liệu gốc, 676 ngày thiếu sau reindex (đã điền).
- **Phân tích:**
  - **Ưu điểm:**
    - Tỷ lệ ngoại lai thấp (2.68% với Z-score), cho thấy dữ liệu tương đối ổn định.
    - **FLC** (bất động sản) có thể đại diện cho một phân khúc thị trường quan trọng.
  - **Nhược điểm:**
    - Phân phối lệch phải mạnh và đuôi nặng, cho thấy có một số giá trị cực đoan (giá tăng đột biến).
    - Skewness (2.40) và Kurtosis (8.02) cao nhất trong 4 cột, có thể gây khó khăn cho mô hình giả định phân phối chuẩn.
  - **Kết luận:** Không nên loại bỏ **FLC**, vì tỷ lệ ngoại lai thấp và dữ liệu có thể xử lý được (đã thay thế ngoại lai bằng nội suy). Tuy nhiên, cần lưu ý phân phối đuôi nặng khi phân tích.

### Cột HSG

- **Thống kê:**
  - Số điểm: 1342
  - Trung bình: 12.29, Trung vị: 11.15
  - Skewness: 0.89 (lệch phải vừa)
  - Kurtosis: 0.05 (đuôi gần chuẩn)
  - Shapiro-Wilk: p-value = 0.0000 (không chuẩn)
- **Ngoại lai:**
  - Z-score: 0 ngoại lai – Do phân phối lệch và độ lệch chuẩn lớn (7.12).
  - IQR: 371 ngoại lai (27.58%) – Tỷ lệ rất cao, đáng lo ngại.
- **Dữ liệu thiếu:** Không thiếu trong dữ liệu gốc, 676 ngày thiếu sau reindex (đã điền).
- **Phân tích:**
  - **Ưu điểm:**
    - **HSG** (thép) là cổ phiếu quan trọng, phản ánh biến động ngành thép (tăng mạnh 2016–2017).
    - Skewness (0.89) và Kurtosis (0.05) vừa phải, dễ xử lý hơn **FLC**.
  - **Nhược điểm:**

- Tỷ lệ ngoại lai 27.58% (IQR) quá cao, cho thấy nhiều biến động giá bị coi là ngoại lai. Điều này có thể do hệ số IQR 1.5 quá nghiêm ngặt, hoặc **HSG** thực sự có nhiều giai đoạn bất thường.
- Độ lệch chuẩn cao (7.12), cho thấy biến động lớn, có thể làm giảm độ chính xác của mô hình.
- **Kết luận: Cân nhắc loại bỏ **HSG****, nhưng chỉ nên làm sau khi thử điều chỉnh hệ số IQR (ví dụ: 2.0) để giảm tỷ lệ ngoại lai. Tỷ lệ 27.58% cho thấy dữ liệu có thể không ổn định, nhưng **HSG** vẫn có giá trị nếu biến động này phản ánh đúng thị trường thép.

## Cột KDC

- **Thống kê:**
  - Số điểm: 1342
  - Trung bình: 23.72, Trung vị: 23.12
  - Skewness: 0.37 (lệch phải nhẹ)
  - Kurtosis: -0.17 (đuôi nhẹ)
  - Shapiro-Wilk: p-value = 0.0000 (không chuẩn)
- **Ngoại lai:**
  - Z-score: 0 ngoại lai – Do phân phối gần chuẩn và độ lệch chuẩn lớn (7.57).
  - IQR: 138 ngoại lai (10.26%) – Tỷ lệ chấp nhận được.
- **Dữ liệu thiếu:** Không thiếu trong dữ liệu gốc, 676 ngày thiếu sau reindex (đã điền).
- **Phân tích:**
  - **Ưu điểm:**
    - Tỷ lệ ngoại lai thấp (10.26% với IQR), cho thấy dữ liệu tương đối ổn định.
    - Phân phối gần đối xứng (skewness = 0.37), đuôi nhẹ (kurtosis = -0.17), phù hợp cho phân tích.
    - **KDC** (thực phẩm) đại diện cho ngành ổn định hơn **HSG**.
  - **Nhược điểm:**
    - Độ lệch chuẩn cao (7.57), cho thấy biến động giá đáng kể.
  - **Kết luận:** Không nên loại bỏ **KDC**, vì tỷ lệ ngoại lai hợp lý và phân phối gần đối xứng, phù hợp cho phân tích hoặc mô hình hóa.

## Cột PPC

- **Thống kê:**
  - Số điểm: 1342
  - Trung bình: 13.48, Trung vị: 14.32
  - Skewness: -0.81 (lệch trái vừa)
  - Kurtosis: -0.42 (đuôi nhẹ)
  - Shapiro-Wilk: p-value = 0.0000 (không chuẩn)
- **Ngoại lai:**
  - Z-score: 0 ngoại lai – Do phân phối lệch và độ lệch chuẩn (4.08).
  - IQR: 241 ngoại lai (17.92%) – Tỷ lệ cao, cần xem xét.
- **Dữ liệu thiếu:** Không thiếu trong dữ liệu gốc, 676 ngày thiếu sau reindex (đã điền).
- **Phân tích:**
  - **Ưu điểm:**
    - **PPC** (điện) đại diện cho ngành ổn định, với độ lệch chuẩn thấp (4.08) so với **HSG**, **KDC**.
    - Skewness (-0.81) và Kurtosis (-0.42) vừa phải, dễ xử lý.

- **Nhược điểm:**

- Tỷ lệ ngoại lai 17.92% (IQR) khá cao, có thể do các giai đoạn giá giảm (lệch trái).
- Hệ số IQR 1.5 có thể quá nghiêm ngặt, dẫn đến phát hiện quá nhiều ngoại lai.

- **Kết luận:** Không nên loại bỏ **PPC** ngay, nhưng cần thử điều chỉnh hệ số IQR (ví dụ: 2.0) để giảm tỷ lệ ngoại lai. Tỷ lệ 17.92% cho thấy biến động, nhưng **PPC** vẫn có giá trị nếu phản ánh đúng thị trường điện.