

Bir örnekte aykırı değerlerin (outlier) ortaya çıkmasının çeşitli sebepleri vardır. Bu sebeplerin anlaşılması, doğru veri analizi yapabilmek için önemlidir. İşte aykırı değerlerin başlıca nedenleri:

1. Veri Girişi Hataları

- **İnsan hatası:** Veri girişindeki yazım hataları veya birimlerin yanlış yorumlanması gibi hatalar, aykırı değerlere neden olabilir.
- **Yanlış ölçüm:** Aletlerin kalibrasyon sorunları veya yanlış veri toplama prosedürleri, beklenmedik değerler üretebilir.

2. Örneklem Hataları

- **Temsili olmayan örneklem:** Eğer örneklem, kitleyi doğru şekilde temsil etmiyorsa, nadir veya alışılmadık değerler içerebilir ve bu da aykırı değerlere yol açar.
- **Rasgele değişkenlik:** Rasgele seçilen bir örneklemde, tamamen şansa bağlı olarak bazı uç değerler ortaya çıkabilir.

3. Doğal Değişkenlik

- **Doğal varyasyon:** Bazı veri kümelerinde, özellikle biyolojik, ekonomik veya sosyal fenomenlerle ilgili verilerde, uç değerler sistemin doğal değişkenliğinin bir parçası olarak ortaya çıkabilir.
- **Gerçek ama olağan dışı vakalar:** Örneğin, gelir verilerinde aşırı zengin bireyler aykırı değerler olabilir, ancak bu durum geçerli olabilir.

4. Temel Sürecin Değişmesi

- **Süreçteki değişiklikler:** Veri üreten sistem veya süreç değiştiğinde (örneğin, yeni bir teknoloji, piyasa değişiklikleri veya ani dış faktörler), bu yeni uç değerlere neden olabilir.
- **Dönem değişiklikleri:** Zaman serisi verilerinde, temel eğilimlerin veya desenlerin değişmesi, geçiş dönemlerinde aykırı değerlere yol açabilir.

5. Veri Dönüşümü veya Toplama

- **Yanlış dönüşüm:** Bazen, veriye uygulanan yanlış dönüşümler (örneğin, logaritma, ölçekleme gibi) aykırı değerlere neden olabilir.
- **Toplama hataları:** Farklı kaynaklardan veya zaman dilimlerinden gelen veriler birleştirildiğinde, verilerin tutarsız bir şekilde birleşmesi sonucu aykırı değerler ortaya çıkabilir.

6. Ölçüm Sınırlamaları

- **Ölçüm araçlarının sınırlamaları:** Bazen, aletlerin sınırlı bir ölçüm aralığı veya hassasiyeti vardır ve bu da olağandışı okumaların aykırı değer olarak görünmesine neden olabilir. Örneğin, bir sensör maksimum limitinin üzerine çıktığında aykırı bir değer verebilir.

- **Sansürleme veya kesilme:** Bazı aykırı değerler, verilerin kesilmesinden (örneğin, ölçüm cihazı sınırlı bir aralığa sahip olduğunda) veya sansürlenmesinden (örneğin, belirli bir eşğin üzerindeki değerlerin kaydedilmemesi) kaynaklanabilir.

7. Gerçek İstisnai Olaylar

- **Nadir veya uç olaylar:** Aykırı değerler, nadir, alışılmadık veya uç olayları temsil edebilir. Örneğin, finansal verilerde, bir ekonomik kriz veya küresel bir salgın, normal veri desenlerinden büyük sapmalara neden olabilir.
- **Anormal gözlemler:** Bazı aykırı değerler, önemli bulgular veya iç görüleri temsil edebilir (örneğin, tıbbi araştırmalarda bir hastalığın yeni bir varyantını belirten bir aykırı değer).

8. İstatistiksel Model Hataları

- **Uygun olmayan model kullanımı:** Eğer kullanılan istatistiksel model veri kümesi için uygun değilse, model verilerdeki değişkenlikleri doğru şekilde dikkate almayabilir ve bu da aykırı değerler ortaya çıkarabilir.
- **Aşırı genelleme (ekstrapolasyon):** Veriler, modelin öngördüğü alanın dışında bir yere düştüğünde, model aşırı uç değerler tahmin edebilir.

9. Yanlılık (Bias) veya Sistemik Faktörler

- **Veri toplama yanlılığı:** Bazen, veri toplama sırasında sistemik yanlılıklar olabilir (örneğin, sadece belirli bir alt grup üzerinde anket yapılması), bu da aykırı değerlere yol açabilir.
- **Veri yönlendirilmesi(manipülasyonu):** Bazı durumlarda, veriler kasıtlı olarak yönlendirilebilir veya çarpıtılabilir, bu da gerçekliği yansıtmayan uç değerlere neden olabilir.

10. Dışsal veya Çevresel Faktörler

- **Çevresel değişiklikler:** Çevredeki değişiklikler (örneğin, hava durumu, politika değişiklikleri veya toplumsal olaylar) veriyi etkileyebilir ve aykırı değerlere neden olabilir.
- **Dışsal kesintiler:** Doğal afetler, politik değişiklikler veya büyük toplumsal değişimler gibi olaylar, verilerde olağan dışı değerlerin ortaya çıkmasına neden olabilir.

Sonuç

Aykırı değerler, birçok farklı faktöre bağlı olarak ortaya çıkabilir ve bunların nedenlerini anlamak, nasıl ele alınacaklarına karar verirken önemlidir. Bazen aykırı değerler, geçerli ve önemli uç durumları temsil ederken, diğer durumlarda veri toplama sürecindeki hatalar veya sorunlar olabilir. Aykırı değerlerin bağlamını ve nedenlerini analiz etmek, onları dışlamak mı yoksa daha derinlemesine incelemek mi gerektiğine karar verirken rehberlik sağlar.