

Eksik Veri ve Doldurma Yöntemleri Nedir?

Çalışma Kağıdı

Eksik veri, bir veri setindeki bazı gözlemlerin veya değişkenlerin mevcut olmaması durumudur. Eksik verileri doldurma yöntemleri ise bu boşlukları istatistiksel olarak tahmin edip tamamlama teknikleridir.

Sorular

- Aşağıdakilerden hangisi eksik veri doldurma yöntemlerinden biri DEĞİLDİR?
A) Ortalama ile Doldurma
B) Regresyon ile Doldurma
C) Veri Silme (Listwise Deletion)
D) Çoklu Doldurma
- Eksik verileri ortalama ile doldurmanın en büyük dezavantajı nedir?
A) Hesaplanması zordur.
B) Değişken varyansını azaltır.
C) Analizleri çok karmaşık hale getirir.
D) Tüm veri setini siler.
- Hangi doldurma yöntemi, eksik değerler için birden fazla olası değer üreterek belirsizliği daha iyi ele alır?
A) Medyan ile Doldurma
B) Son Gözlem Taşıma (Last Observation Carried Forward)
C) Çoklu Doldurma
D) Mod ile Doldurma
- Basit bir ortalama doldurma örneği verin.
- Regresyon ile doldurma nasıl yapılır?
- Çoklu doldurma (Multiple Imputation) nedir?
- Tanımla: Eksik Veri Nedir?
- Tanımla: Neden Eksik Veri Önemlidir?
- Tanımla: En Basit Doldurma Yöntemi Hangisidir?

Cevap Anahtarı

1. C) Veri Silme (Listwise Deletion) — Veri silme (listwise deletion), eksik veri içeren gözlemleri tamamen analizden çıkarmaktır; doldurma yöntemi değildir.
2. B) Değişken varyansını azaltır. — Ortalama ile doldurma, veri setindeki değişkenliğin gerçekte olduğundan daha az görünmesine neden olur.
3. C) Çoklu Doldurma — Çoklu doldurma, eksik değerler için birden fazla tahmin üretir ve bu tahminler üzerinden analizler yaparak belirsizliği hesaba katar.
4. Bir veri setinde 'Yaş' değişkeninde birkaç eksik değer varsa, mevcut 'Yaş' değerlerinin ortalaması hesaplanır. Ardından, eksik olan her bir yaş değeri, bu ortalama ile doldurulur.
5. Eksik veriye sahip değişken ile diğer değişkenler arasındaki ilişki regresyon modeli ile belirlenir. Model kullanılarak eksik değerler tahmin edilir ve yerine konulur.
6. Eksik değerlerin her biri için birden fazla olası değer üretilir ve bu şekilde birden fazla tamamlanmış veri seti oluşturulur. Analizler bu veri setlerinin her birinde yapılır ve sonuçlar birleştirilir.
7. Bir veri setinde bazı gözlemlerin veya değişkenlerin mevcut olmaması durumudur.
8. Analizlerin yanlı olmasına, istatistiksel gücün azalmasına ve sonuçların güvenilirliğinin düşmesine neden olabilir.
9. Ortalama, medyan veya mod ile doldurma.

Bounlu

Tüm kartlar, adım adım çözümler ve AI hoca desteği Notek uygulamasında.
Sınav tarihlerini Promy otomatik hatırlatıcıya çevirir.