

Regresyon Analizi Raporu (R)

Konu: Yaşanabilir Alan (Gr.Liv.Area) ile Satış Fiyatı (SalePrice) Arasındaki İlişki

Veri Seti: `ames_train` (ames.csv)

1. Paketlerin Yüklenmesi ve Veri Yükleme

İlk olarak analizde kullanılacak paketler yüklendi:

- `library(tidyverse)`
- `library(broom)`
- `library(ggplot2)`
- `tidyverse`: Veri manipülasyonu ve görselleştirme için.
- `broom`: Model sonuçlarını düzenli bir formata getirmek için.
- `ggplot2`: Görselleştirme için.

Veri `ames.csv` dosyasından okundu:

- `ames_train <- read.csv("D:/ames.csv")`
- `head(ames_train)`

2. Basit Lineer Regresyon Modeli

Model tanımı şu şekilde yapıldı:

- `model1 <- lm(SalePrice ~ Gr.Liv.Area, data = ames_train)`
- `summary(model1)`

Model Özeti:

- **Bağımlı Değişken:** `SalePrice`
- **Bağımsız Değişken:** `Gr.Liv.Area`

Model özetiyle:

- `Gr.Liv.Area` katsayısı pozitif çıktı. Bu da yaşanabilir alan arttıkça satış fiyatının arttığını gösterir.
- **R² (Determination Coefficient):** Modelin açıklayıcılığı. Örneğin R² = 0.51 ise, model değişkenliğin %51'ini açıklıyor.

- **p-değeri:** Bağımsız değişkenin anlamlı olup olmadığını gösterir (genellikle $p < 0.05$ olmalı).

3. Tahmin ve Artık (Residual) Değerlerin Elde Edilmesi

Model sonuçlarına tahminler (`.fitted`) ve artıklar (`.resid`) eklendi:

- `model1_aug <- augment(model1, data = ames_train)`

Bu adım sayesinde her gözlem için:

- Tahmin edilen satış fiyatı,
- Gerçek fiyatla tahmin arasındaki fark (artık), elde edilmiş oldu.

4. Regresyon Doğrusu Grafiği

- `ggplot(model1_aug, aes(x = Gr.Liv.Area, y = SalePrice)) +`
- `geom_point(alpha = 0.5) +`
- `geom_smooth(method = "lm", se = FALSE, color = "blue", linewidth = 1) +`
- `labs(title = "Regresyon Doğrusu: Sale Price ~ Gr Liv Area")`

Yorum:

- Noktalar: Gerçek veri noktalarını temsil eder.
- Mavi çizgi: Tahmin edilen regresyon doğrusunu gösterir.
- **Pozitif eğim**, alan arttıkça fiyatın da arttığını görsel olarak da ortaya koyar.
- `alpha = 0.5` ile noktalar yarı saydam yapılarak üst üste binmelerin önüne geçildi.

5. Artıkların (Residuals) Dağılımı

- `ggplot(model1_aug, aes(x = Gr.Liv.Area, y = .resid)) +`
- `geom_point(alpha = 0.5) +`
- `geom_hline(yintercept = 0, color = "red", linetype = "dashed") +`
- `labs(title = "Artıkların Dağılımı", y = "Artıklar", x = "Gr Liv Area")`

Yorum:

- Artıklar sıfır etrafında dağılmalıdır. Bu durum modelin **hatalarının ortalama sıfır** olduğunu ve homojen olduğunu gösterir.

- Kırmızı kesikli çizgi sıfır artığı temsil eder.
- Eğer artıklar rastgele dağılmışsa, modelin varsayımlarının sağlandığı söylenebilir.
- Desene sahip bir dağılım modelin yetersizliğini veya başka değişkenlerin etkisini gösterebilir.

Genel Değerlendirme

- **Gr.Liv.Area**, **SalePrice** üzerinde güçlü ve pozitif bir etkiye sahiptir.
- Model, bu tek değişkenle fena sayılmayacak bir açıklama gücüne sahiptir.
- Artık analizi, modelin lineerlik varsayımını desteklemektedir.
- Gelecek adımda modele başka değişkenler de eklenerek **çoklu regresyon** modeli kurulabilir.