

Çözümler: Veri Import ve Export Mantığı

R Programlama ve İstatistik Akademisi

Dr. M. Fatih Tüzen

Genel Not

Bu dosya, alıştırmaların çözümlerini içerir.
Kodları adım adım çalıştırıp çıktıları inceleyin.

Veri kaynağı:

<https://raw.githubusercontent.com/mwaskom/seaborn-data/master/tips.csv>

Çözüm 1 — Çalışma Dizini ve Dosya Kontrolü

```
# Mevcut çalışma dizini
getwd()

# tips.csv dosyasını çalışma dizinine indir
url_tips <- "https://raw.githubusercontent.com/mwaskom/seaborn-data/master/tips.csv"
download.file(url_tips, destfile = "tips.csv", mode = "wb")

# Dosya indirildi mi?
file.exists("tips.csv")
```

Çözüm 2 — CSV Dosyasını Okuma (Virgül ile Ayrılmış)

```
tips <- read.csv("tips.csv")

# Boyut

dim(tips)

# İlk satırlar

head(tips)

# Değişken tipleri

str(tips)
```

Sayısal değişkenler genellikle `total_bill`, `tip`, `size`; kategorik değişkenler `sex`, `smoker`, `day`, `time` olarak gelir.

Çözüm 3 — CSV Dosyasını Okuma (`read.table`)

```
tips_rt <- read.table(
  file = "tips.csv",
  header = TRUE,
  sep = ",",
  dec = ".",
  stringsAsFactors = FALSE
)

# Kontroller

dim(tips_rt)
head(tips_rt)
str(tips_rt)
```

Genellikle sonuçlar aynı olur. Fark olursa çoğu zaman:

- `stringsAsFactors` varsayımları,
- dosyada beklenmedik ayırıcı/boşluklar,
- encoding gibi nedenlerden kaynaklanır.

Çözüm 4 — CSV Dosyasını R’den Üretme

```
write.csv(  
  tips,  
  file = "tips_comma_exercise.csv",  
  row.names = FALSE  
)  
  
file.exists("tips_comma_exercise.csv")
```

`row.names = FALSE` satır numaralarının dosyaya yazılmasını engeller.

Çözüm 5 — Noktalı Virgül ile CSV Üretme

```
write.csv2(  
  tips,  
  file = "tips_semicolon_exercise.csv",  
  row.names = FALSE  
)  
  
file.exists("tips_semicolon_exercise.csv")
```

```
tips_sc <- read.csv2("tips_semicolon_exercise.csv")  
  
head(tips_sc)  
str(tips_sc)  
  
# Sayısal değişken kontrolü  
  
is.numeric(tips_sc$total_bill)  
is.numeric(tips_sc$tip)  
is.numeric(tips_sc$size)
```

`read.csv()` ile okunursa çoğunlukla:

- ayırıcı yanlış olacağından sütunlar bozulabilir,
- sayısal alanlar karakter olarak gelebilir veya tüm satır tek sütuna “yığılabilir”.

Çözüm 6 — Excel Dosyasına Veri Yazma

```
# Paket (yükli değilse önce install.packages("openxlsx"))

library(openxlsx)

write.xlsx(
  x = tips,
  file = "tips_excel_exercise.xlsx",
  sheetName = "tips"
)

file.exists("tips_excel_exercise.xlsx")

# Sheet isimlerini listele

getSheetNames("tips_excel_exercise.xlsx")
```

Çözüm 7 — Excel Dosyasından Veri Okuma

```
tips_excel <- read.xlsx(
  xlsxFile = "tips_excel_exercise.xlsx",
  sheet = 1
)

head(tips_excel)
str(tips_excel)

# Basit karşılaştırma: boyutlar aynı mı?

dim(tips) == dim(tips_excel)
```

Veri yapısı genel olarak korunur. Farklılık olursa çoğu zaman:

- Excel'de saklama biçimi,
- sütun tiplerinin otomatik algılanması,
- karakter/sayı dönüşümleri gibi nedenlerden kaynaklanır.

Çözüm 8 — Çoklu Sheet ile Excel Raporu

```
# Günlere göre ortalama bahşis

summary_tips <- aggregate(tip ~ day, data = tips, FUN = mean)
names(summary_tips) <- c("day", "mean_tip")

summary_tips

# Workbook oluşturun

wb <- createWorkbook()

# Ham veri

addWorksheet(wb, "Ham Veri")
writeData(wb, sheet = "Ham Veri", x = tips)

# Özet

addWorksheet(wb, "Ozet")
writeData(wb, sheet = "Ozet", x = summary_tips)

# Kaydet

saveWorkbook(wb, file = "tips_rapor_exercise.xlsx", overwrite = TRUE)

file.exists("tips_rapor_exercise.xlsx")

# Sheet isimlerini kontrol et

getSheetNames("tips_rapor_exercise.xlsx")
```

Çözüm 9 — Kontrol Alışkanlığı

1. Bir veri setinin doğru okunup okunmadığını kontrol etmek için:

```
# Temel kontroller

head(tips)
```

```
str(tips)
summary(tips)
```

- `dim()` ile satır/sütun sayısı kontrol edilebilir
 - Beklenen değişken tipleri (numeric/character) doğrulanmalıdır
2. CSV dosyalarında sessiz hatalara neden olabilecek durumlar:
- Yanlış `sep` (ayırıcı)
 - Yanlış `dec` (ondalık ayracı)
 - Encoding problemleri
 - Metinlerin otomatik factor'a dönüşmesi (eski R davranışları)
 - Excel'de açıp kaydedince dosyanın yapısının fark edilmeden değişmesi
3. Excel dosyalarıyla çalışırken sık yapılan hatalar:
- Yanlış sheet'in okunması
 - Sheet isimlerini kontrol etmeden okuma
 - Dosyanın nereye kaydedildiğinin bilinmemesi
 - `overwrite = TRUE` yüzünden dosyanın istemeden üzerine yazılması