

Ankara Üniversitesi  
Siyasal Bilgiler Fakültesi  
İktisat Bölümü

2022-2023 Bahar

İKT 0208/0210/232/342 - İstatistik II Bütünleme Sınavı

Emrah Er

22-06-2023

---

Açıklama ve Uyarılar

- Bu açıklamalara harfiyen uyunuz! Cevaplarınızı bir R dosyasına yazınız! Sınavın bireysel olduğunu unutmayınız!
- Sınav, bu sayfa da dahil olmak üzere, toplam 6 sayfadır. Lütfen sayfaların eksik olup olmadığını kontrol ediniz. Sınav toplam 100 puan değerinde olup 5 sorudan oluşmaktadır.
- Cevaplarınızı 22-06-2023 tarihi saat 13:00'e kadar sisteme ([ekampus.ankara.edu.tr](http://ekampus.ankara.edu.tr)) proje ödevinizle birlikte sıkıştırılmış dosya olarak yükleyiniz. Gönderdiğiniz dosyada cevaplarınızın sıralı (önce 1. Soru sonra 2. Soru vs.) olmasına dikkat ediniz. Gönderdiğiniz dosyanın ismini Okul Numarası-İsimSoyisim.R şeklinde olmasına dikkat ediniz (Örneğin 18080000-EmrahEr.R).

---

Soru:	1	2	3	4	5	Toplam
Puan:	40	15	35	5	5	100
Skor:						

---

Başarılar!

**Soru 1.....(40 Puan)**

- (a) (40 Puan) Dönem boyunca Final Projesi için yaptığınız çalışmaların boşa gitmemesi için bu soruda Proje ödevinizden aldığımız notun %40'ı kullanılacaktır. Projenizin Github repo adresi nedir?

**Soru 2.....(15 Puan)**

Bir araştırmacı Titanic kazası hakkında bir çalışma için bir veri seti toplamıştır ve R'a aktarmıştır. Veri üzerinde bazı veri temizleme işlemleri yaptıktan sonra veri setini `titanic` adıyla bir `tibble` nesnesine kaydetmiştir (Verinin ham haline <https://bit.ly/3vTgDjZ> adresinden erişilebilir). Veride yer alan değişkenlere ait tanımlar şu şekildedir:

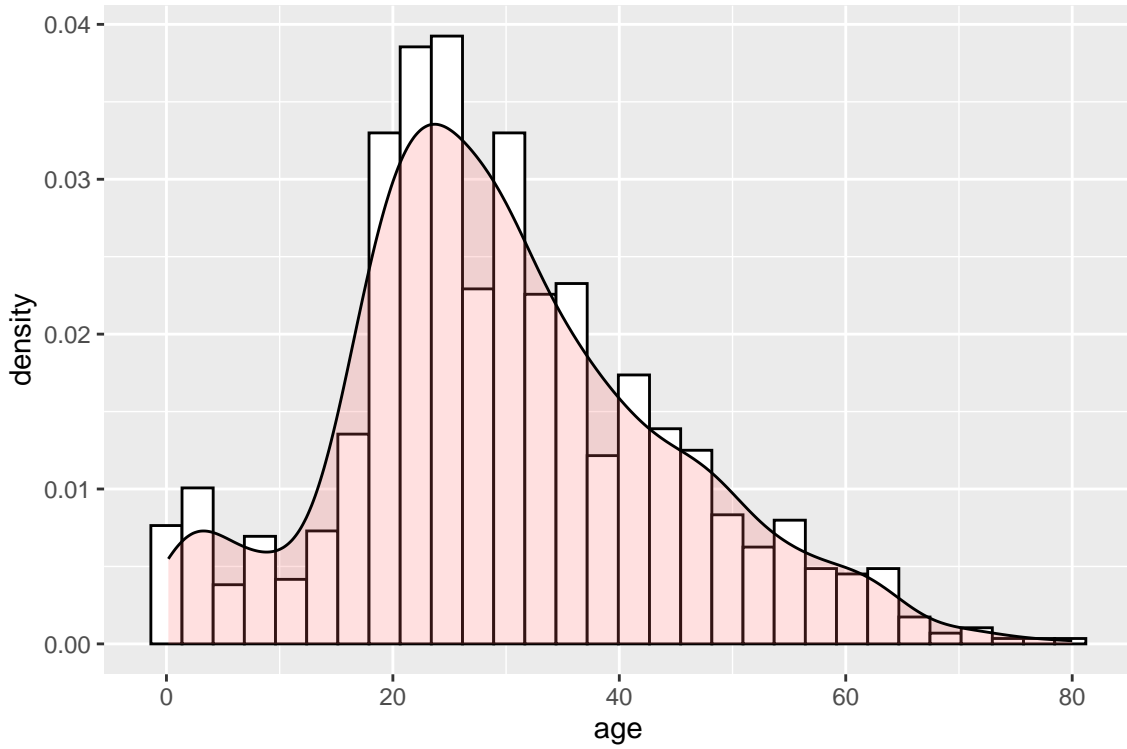
- **pclass:** Yolcunun hangi sınıfta yolculuk ettiği.
- **survived:** Yolcunun kazadan kurtulup kurtulmadığı. (1 = Kurtulmuş, 0 = Kurtulamamış)
- **sex:** Yolcunun cinsiyeti. (Female = Kadın, Male = Erkek)
- **age:** Yolcunun yaşı.
- **fare:** Yolcunun bilet için ödediği tutar.

`summary(titanic)` komutu kullanılarak veri için elde edilen özet istatistikler şu şekildedir:

```
##  pclass  survived      sex          age          fare
##  1:323    0:809    female:466  Min.    : 0.17  Min.    : 0.0
##  2:277    1:500    male  :843   1st Qu.:21.00  1st Qu.: 7.9
##  3:709                                Median :28.00  Median : 14.5
##                                Mean   :29.88  Mean   : 33.3
##                                3rd Qu.:39.00  3rd Qu.: 31.3
##                                Max.    :80.00  Max.    :512.3
##                                NA's    :263    NA's    :1
```

Aşağıda yer alan soruları özet istatistiklerde yer alan bilgileri de göz önünde bulundurarak cevaplayınız.

- (a) (5 Puan) “Titanic’te yolculuk yapmak için ortalamada kim daha fazla ödeme yapmıştır, **kadınlar** mı, **erkekler** mi?” sorusuna cevap vermek istediğimizi varsayalım. `dplyr` paketinde yer alan fonksiyonları kullanarak R’da bunu hangi kodla yapabileceğimizi yazınız.
- (b) (5 Puan) Kayıp (NA) olmayan gözlemler için cinsiyete göre yaşlara ait kutu grafiğini çizmek istediğimizi varsayalım. `tidyverse` paketinde yer alan `ggplot2` fonksiyonlarını kullanarak R’da bunu hangi kodla yapabileceğimizi yazınız.

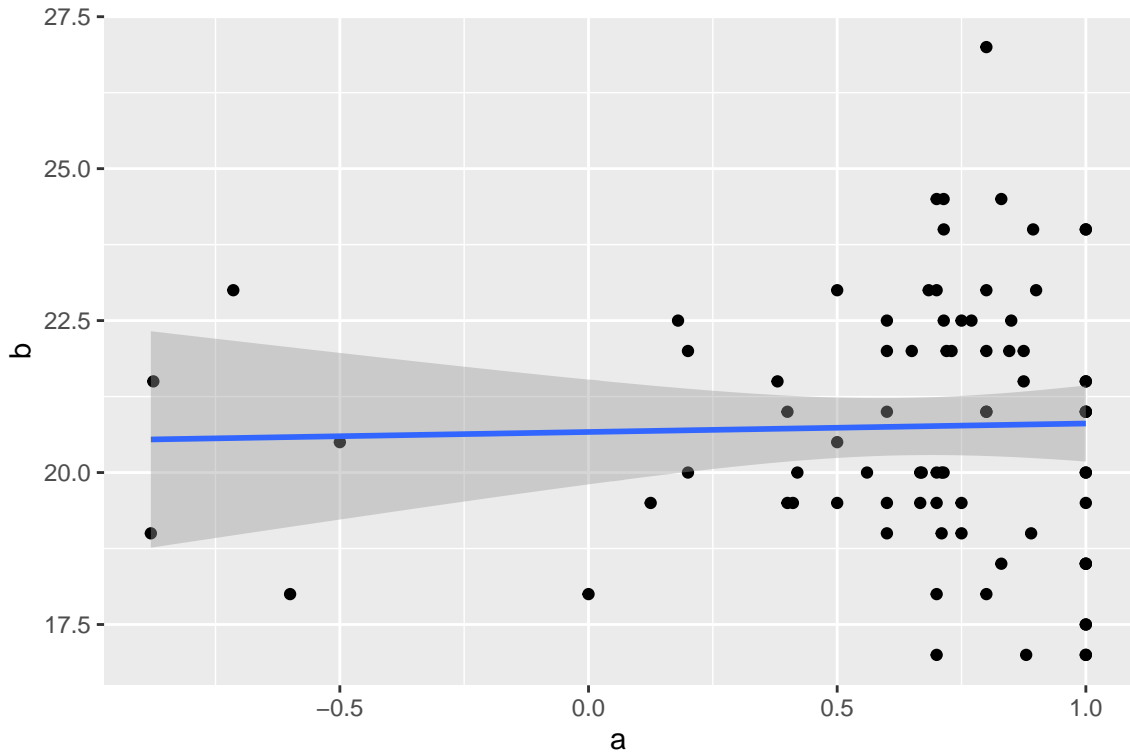


Şekil 1: Histogram

(c) (5 Puan) Şekil 1’de yer alan grafiği R’da türetmek için gerekli olan kodu yazınız.

**Soru 3.....(35 Puan)**

- (a) (5 Puan) `x` vektörü R’da `x <- 10:20` komutu kullanılarak tanımlanmışsa, `x[seq(1, 5, by = 3)]` komutunun sonucu nedir?
- (b) (5 Puan) `dat1` ve `dat2` gibi iki `tibble` olduğunu varsayalım. `dat1`’den ve `dat2`’den eşleşen değerlerin olduğu tüm satırları ve `dat1`’deki ve `dat2`’deki tüm sütunları döndürüp `dat3` isminde bir `tibble`’a kaydetmek istersek hangi komutu çalıştırmamız gerekir?
- (c) (5 Puan) Şekil 2’de `dat` `tibble` nesnesi kullanılarak çizilmiş bir grafik yer almaktadır. Verilen grafiği çizmek için hangi `ggplot2` komutları kullanılmalıdır?
- (d) (5 Puan) Aşağıda verilen komutlar çalıştırıldığında `myresult` ne olur?
- ```
mylist <- list(1:3, c(3:5, NA))
myresult <- map(mylist, ~ mean(.x, na.rm = TRUE)) %>% unlist()
```
- (e) (5 Puan)  $X_1, \dots, X_{25} \stackrel{iid}{\sim} N(\mu, \sigma)$  olsun.  $Z = 5(\bar{X}_n - \mu)/S$  şeklinde tanımlansın.  $P(Z \leq 1)$  olasılığını hesaplamak için gerekli olan R kodu nedir?
- (f) (5 Puan) Tavla oynarken kullanmak amacıyla zar atışlarını (tavla 2 adet zar ile oynanmaktadır)



Şekil 2: Saçılım Grafiği

simüle eden bir fonksiyona ihtiyacımız olduğunu varsayalım. Derste gördüğümüz R komutlarını ve fonksiyonlarını kullanarak zar atışlarını taklit eden ve her bir zarın sonucunu gösteren bir fonksiyon yazınız.

- (g) (5 Puan) Soru 2'deki `titanic` verisini kullanarak Titanik felaketinden kurtulan yolcuların (`survived`), kurtulamayan yolcular ile aynı yaş (`age`) ortalamasına sahip olup olmadığını test etmek istediğimizi varsayalım. Bunun için R'da hangi komut(lar) kullanılmalıdır? (Her iki grubun eşit varyansa sahip olduğunu varsayınız.)

**Soru 4.....(5 Puan)**

Aşağıda verilen **dat** veri seti kullanılarak

```
## # A tibble: 2 × 4
##   country   '2018' '2019' '2020'
##   <fct>     <dbl>  <dbl>  <dbl>
## 1 İngiltere  8000    8100    8500
## 2 Almanya   10000   11000   10200
```

aşağıdaki **dat2** veri seti oluşturulmuştur.

```
## # A tibble: 6 × 3
##   country   year   gdp
##   <fct>     <chr> <dbl>
## 1 İngiltere 2018    8000
## 2 İngiltere 2019    8100
## 3 İngiltere 2020    8500
## 4 Almanya   2018   10000
## 5 Almanya   2019   11000
## 6 Almanya   2020   10200
```

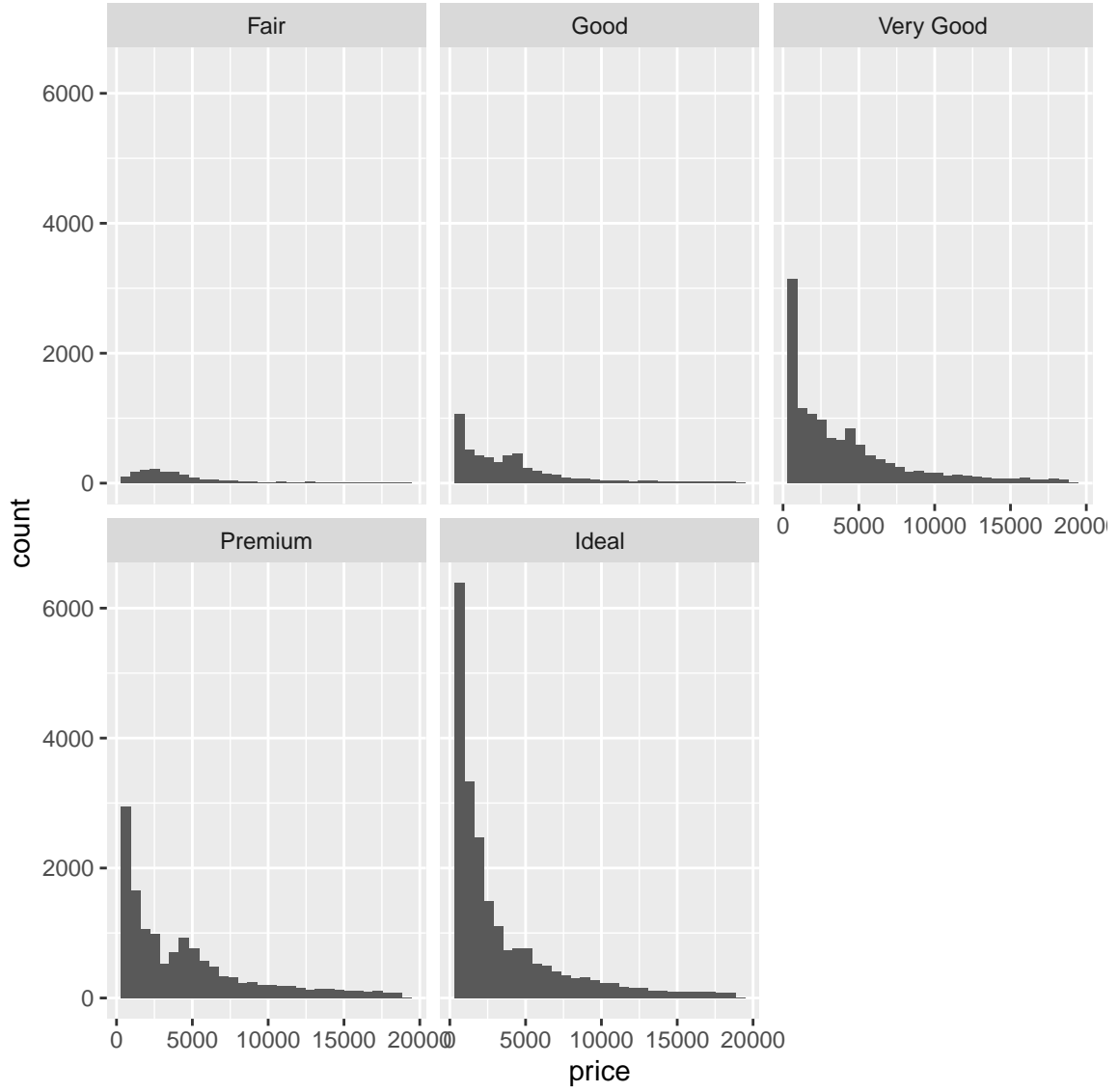
(a) (5 Puan) **dat2** veri setini oluşturan **tidyr** komutlarını yazınız.

**Soru 5.....(5 Puan)**

Aşağıda ilk 5 satırı verilen veri seti (**dat**) kullanılarak Şekil 3'de yer alan grafik çizilmiştir.

```
## # A tibble: 5 × 4
##   price cut    depth color
##   <int> <ord>    <dbl> <ord>
## 1   326 Ideal   61.5   E
## 2   326 Premium 59.8   E
## 3   327 Good    56.9   E
## 4   334 Premium 62.4   I
## 5   335 Good    63.3   J
```

(a) (5 Puan) Bu grafiği çizmek için hangi **ggplot2** komutları kullanılmalıdır?



Şekil 3: