

Yaş Değişkenine Göre Ortalama Poliçe Sayısının Çubuk ve Gruplandırılmış Çubuk Grafiği İle Görselleştirilmesi

Rümeysa Kurt

26 Ekim 2022

ÖZET

Bu raporda “MASS” paketinde bulunan “Insurance” veri setini kullanarak iki farklı grafik hazırlanmıştır.Öncelikle veri setine erişmek için `install.packages(“MASS”)`, `ggplot` görselleştirme araçlarını kullanmak için `install.packages(“ggplot2”)` ve `pipe` operatörü kullanabilmek/veri manipülasyonu işlemlerini yapabilmek için `install.packages(“dplyr”)` paketleri indirilmiştir.İlk grafik yaşa göre ortalama poliçe sayısının bulunduğu bir çubuk grafiği olup ikinci grafik ise yaş değişkenine göre ortalama poliçe sayısının, araç motor hacmine göre gruplandırılmış çubuk grafiği ile görselleştirilmesidir.Son olarak oluşturulan grafikler yorumlanmıştır.

```
install.packages("MASS")
library(MASS)
data(Insurance)
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyverse")
library(ggplot2)
library(dplyr)
library(tidyverse)
```

Yaş(Age) Değişkenine Göre Ortalama Poliçe Sayısının(Claims) Görselleştirilmesi

Öncelikle kullandığımız kategorik değişkenin (Age) factor tipinde bir R değişkeni olduğundan emin olmak için `class()` komutu ile kontrol ettik. `levels()` fonksiyonu sayesinde de Age değişkeninde kaç farklı düzey/seviye olduğunu görmüş olduk.

```
class(Insurance$Age)
```

```
[1] "ordered" "factor"
```

```
levels(Insurance$Age)
```

```
[1] "<25" "25-29" "30-35" ">35"
```

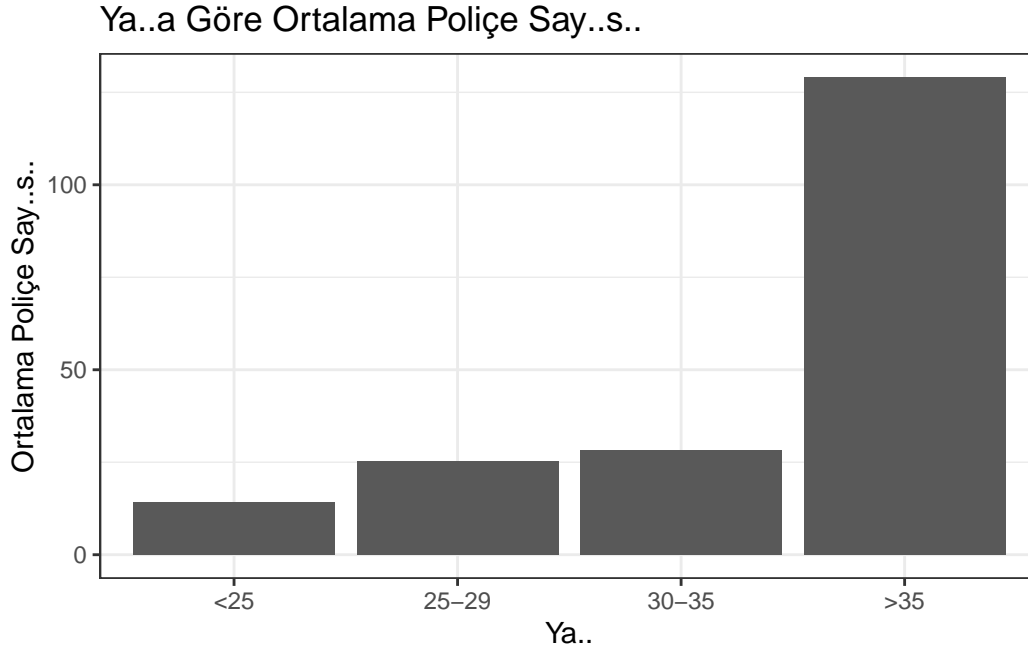
Bizden yaş değişkenine göre ortalama poliçe sayısını istediği için yaşa göre gruplayıp ortalamalarını aldık. Aşağıdaki taloda her bir yaş grubunda ortalama kaç poliçe olduğu görülüyor. Şimdi grafiği çizebiliriz.

```
mean_insurance <- Insurance %>%  
  group_by(Age) %>%  
  summarise(mean_claims = mean(Claims))  
mean_insurance
```

```
# A tibble: 4 x 2  
  Age    mean_claims  
  <ord>    <dbl>  
1 <25      14.3  
2 25-29    25.2  
3 30-35    28.3  
4 >35     129.
```

Grafiğin Çizilmesi

```
ggplot(mean_insurance, aes(x = Age, y = mean_claims)) +  
  geom_bar(stat = "identity") +  
  labs(x = "Yaş",  
       y = "Ortalama Poliçe Sayısı",  
       title = "Yaşa Göre Ortalama Poliçe Sayısı") +  
  theme_bw()
```



Grafiğin Yorumlanması

4 farklı yaş aralığındaki ortalama poliçe sayısını çubuk grafiği ile görselleştirilmiştir. Grafikte x eksenine yaş (kategorik/sıralı), y eksenine ortalama poliçe sayısı (nümerik/sürekli) konumlandırılmıştır. Görüldüğü üzere ortalama poliçe sayısının en yüksek olduğu yaş grubu >35 en az olduğu yaş grubu ise <25'tir. >35 yaş grubu hariç geri kalan üç grup arasındaki fark oldukça azdır.

Yaş(Age) Değişkenine Göre Ortalama Poliçe Sayısının(Claims) Araç Motor Hacmine(Group) Göre Gösterdiği Değişimin Görselleştirilmesi

Öncelikle bu kısımda kullanacağımız kategorik değişkenin (Group) factor tipinde bir R değişkeni olduğundan emin olmak için `class()` komutu ile kontrol ettik. `levels()` fonksiyonu sayesinde de Group değişkeninde kaç farklı düzey/seviye olduğunu görmüş olduk. 4 farklı araç motor hacmi düzeyi bulunmaktadır.

```
class(Insurance$Group)
```

```
[1] "ordered" "factor"
```

```
levels(Insurance$Group)
```

```
[1] "<11"      "1-1.51" "1.5-21" ">21"
```

Bu kısımda yaş değişkenine göre ortalama poliçe sayısının araç motor hacmine göre değişimini görmek istediğimiz için Insurance veri setini yaşa ve araç motor hacmine göre gruplayıp ortalamalarını aldık. Görselleştirme kısmına geçebiliriz.

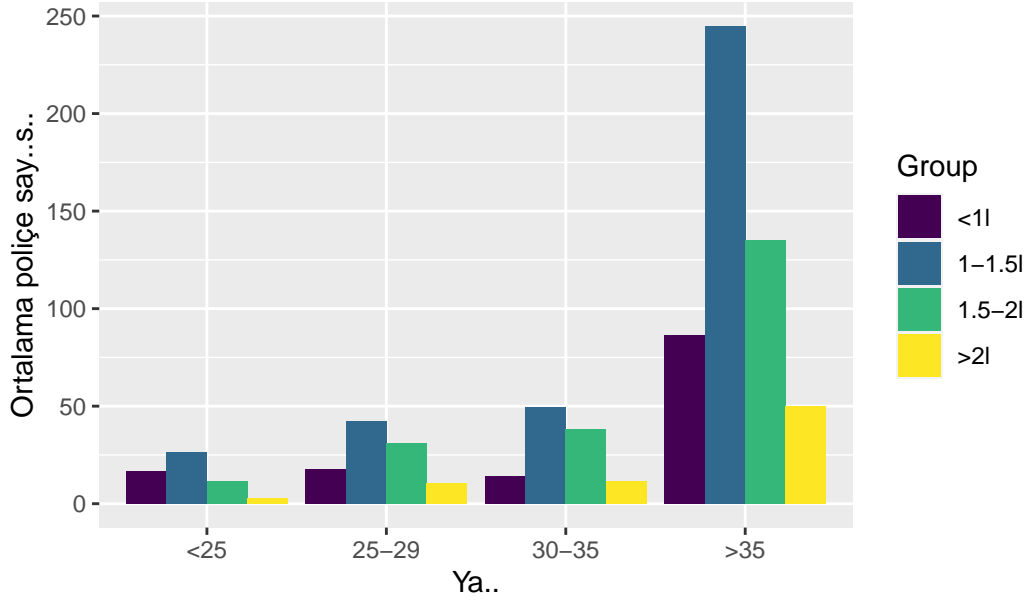
```
mean_insurance2 <- Insurance %>%
  group_by(Age, Group) %>%
  summarise(mean_claims = mean(Claims))
mean_insurance2
```

```
# A tibble: 16 x 3
# Groups:   Age [4]
   Age   Group mean_claims
  <ord> <ord>      <dbl>
1 <25   <11        16.8
2 <25   1-1.51     26.2
3 <25   1.5-21     11.5
4 <25   >21         2.75
5 25-29 <11        17.5
6 25-29 1-1.51     42.2
7 25-29 1.5-21     31
8 25-29 >21       10.2
9 30-35 <11        14
10 30-35 1-1.51    49.2
11 30-35 1.5-21    38.2
12 30-35 >21       11.8
13 >35   <11       86.5
14 >35   1-1.51    245.
15 >35   1.5-21    135
16 >35   >21       50
```

Grafiğin Çizilmesi

```
ggplot(mean_insurance2, aes(x = Age, y = mean_claims, fill = Group)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Yaş",
       y = "Ortalama poliçe sayısı",
       fill = "Group",
       title = "Yaş Göre Ortalama Poliçe Sayısının Araç Motor Hacmine Göre Değişimi")
```

Yaş'a Göre Ortalama Poliçe Sayısının Araç Motor Hacmine Göre Gruplandırılması



Grafik Yorumlanması

Yukarıdaki grafik dört farklı yaş aralığındaki ortalama poliçe sayısının araç motor hacmine göre, gruplandırılmış çubuk grafiği ile görselleştirilmiştir. Grafikte x eksenine yaş (kategorik/sıralı), y eksenine ortalama poliçe sayısı (nümerik/süreklî) konumlandırılmış olup her bir araç motor hacmi düzeyleri farklı renklendirilmiştir. Grafikte baktığımızda yine 35 yaşından büyüklerin araç motor hacmi gruplarına göre ortalama poliçe sayılarının en fazla olduğunu söyleyebiliriz. Her araç motor hacmi düzeyinde ortalama poliçe sayısının en az olduğu yaş grubu ise <25'tir. Buna göre 35 yaşından büyük olup aracının motor hacmi 1-1.5 litre arasında bulunan kişilerin ortalama poliçe sayısı en fazladır diyebiliriz. Yaşı 25'ten küçük olup araç motor hacmi 2'den fazla olan kişilerin ise ortalama poliçe sayısı en azdır yorumunu yapabiliriz.