

Bilişimde İstatistiksel Analiz

Ders 3

Ali Mertcan KOSE Msc.

`amertcankose@ticaret.edu.tr`

İstanbul Ticaret Üniversitesi



İSTANBUL TİCARET
ÜNİVERSİTESİ

Tablo ve grafik bulguların, yazı metnini okumadan, okuyucuya kolay anlaşılır ve incelenebilir biçimde sunulmasını sağlayan bir araçtır.

Bir tablo yada grafiğin okur tarafından kolay anlaşılabilmesi için **karmaşık olmaması**, aksine çok kolay anlaşılır olması beklenir.

Karmaşık bir tabloyu ya da grafiği anlamak ve iletmek istediğiniz mesajı almak okuyucu için hem zor hem de sıkıcı olabilir. *Tablo ve grafik sunumunun amacına da ters düşer*

Bulguların tablo ve grafikte etkin biçimde okuyucuya sunulması bilgi, beceri ve deneyim isteyen bir işlemdir. Bulguları tablo ve grafikte sunarken sadece bir tablo ya da grafik yapmak yerine değişik düzenlerde birkaç deneme yapmak ve bunlar arasından bulguları en kolay ve anlaşılır biçimde sunabilecek olanını seçmek daha uygundur.

Elde edilen bulguların yazı metnine başvurmadan, açık kolay anlaşılır bir şekilde okuyucuya sunulmasını sağlamaktır.

Bulguların Tablo ile Gösterimi

Tablo, yazı metnine başvurmadan bulguların okuyucuya kolay anlaşılabileceği biçimde sunumudur.

Tablo ile verilerin sunumunda dikkat edilecek noktalar şunlardır.

- Her tablonun bir başlığı olmalıdır. Başlık tabloyu tanımlayıcı ve kısa olmalıdır. Başlık genellikle tablonun üstüne konur.
- Kolon ve satır başlıkları açık bir biçimde yazılmalıdır.
- Kolon ya da satırlarda gösterilen ölçekler ve birimler belirtilmelidir. (gr,cm,gün,ay,% vb)

Bulguların Tablo ile Gösterimi

- Gereken yerlerde kolon ve satır toplamaları gösterilmelidir.
- Yüzde, binde vb. oranlar yalnız başına değil, sayı ile birlikte gösterilmelidir.
- Gerektiğinde bilgilerin kolayca birbirinden ayrılmasını sağlamak için satır ve/veya kolonlar kalın çizgilerle ayrılmalıdır.

Marjinal Tablo(Sıklık Tablosu)

Gözlemlerin, incelenen herhangi bir değişkenin kategorilerine, değerlerine ya da oluşturulan sınıflara göre nasıl dağıldığını gösteren tablolara **marjinal (sıklık) tabloları** denir.

iki veya daha fazla değişkenin kategorilerinin kesiştiği yerde frekansların olduğu tablolara **çapraz tablo** denir

iki ya da daha çok değişkenin birlikte değişiminin incelenmesi çoğu zaman çapraz tablo yapımını gerektirir.

##	npreg	glu	bp	skin	bmi	ped	age	type
## 1	5	86	68	28	30.2	0.364	24	No
## 2	7	195	70	33	25.1	0.163	55	Yes
## 3	5	77	82	41	35.8	0.156	35	No
## 4	0	165	76	43	47.9	0.259	26	No
## 5	0	107	60	25	26.4	0.133	23	No
## 6	5	97	76	27	35.6	0.378	52	Yes

type	n	percent
No	132	0.66
Yes	68	0.34

Year	NumberofCigarette	percentange
1900	54	0.2380
1910	151	0.6655
1920	665	2.9310
1930	1485	6.5453
1940	1976	8.7094
1950	3522	15.5236
1960	4171	18.3841
1970	3985	17.5643
1980	3851	16.9737
1990	2828	12.4647

iki ya da daha çok değişkenin birlikte değişiminin incelenmesi çoğu zaman çapraz tablo yapımını gerektirir.

iki ve daha fazla değişkenin kategorilerinin kesiştiği yerde frekansların olduğu tablolara çapraz tablo denir.

Eğer incelenecek iki değişken varsa, bu iki değişkenin birlikte değişimini göstermek amacıyla oluşturulan tabloya **ikili çapraz tablo** denir. Üç değişkenin birlikte değişimini incelemek amacıyla oluşturulan tabloya **üçlü çapraz tablo** denir.

Bağımlı değişken: Bir ya da daha çok etkenin(faktörün) etkisi ile ortaya çıkabilen ve bu etkenler ile ilişkisi aranan değişkene bağımlı değişken denir.

Sadece ölçülür veya kayedilir değişkenlere denir

Bağımsız Değişken: Bağımlı değişkeni etkilediği düşünülen değişkendir.

Araştırmalarda değerlerini belirli düzeylere ayarlayabildiğimiz değişkenlerdir.

	Heart attack	No heart attack	Total
Placebo	189	10845	11034
Aspirin	104	10933	11037
Total	293	21778	22071

Hospital		A	B	Total
Deaths	Confirmed Accurate	157	268	425
	Inaccurate No Change	18	44	62
	Incorrect Recording	54	34	88
	Total	229	346	575

Grafik: Tablo olarak özetlenen bilgiler grafiklerle de sunulabilir. Grafikler elde edilen sonuçların şekillerle ifade edilerek açık ve kolay anlaşılır biçiminde sunulmasını sağlar.

Bu nedenle bir bilimsel yazıda tablo olmadan tek başına grafik konması hatalı bir davranış olur.

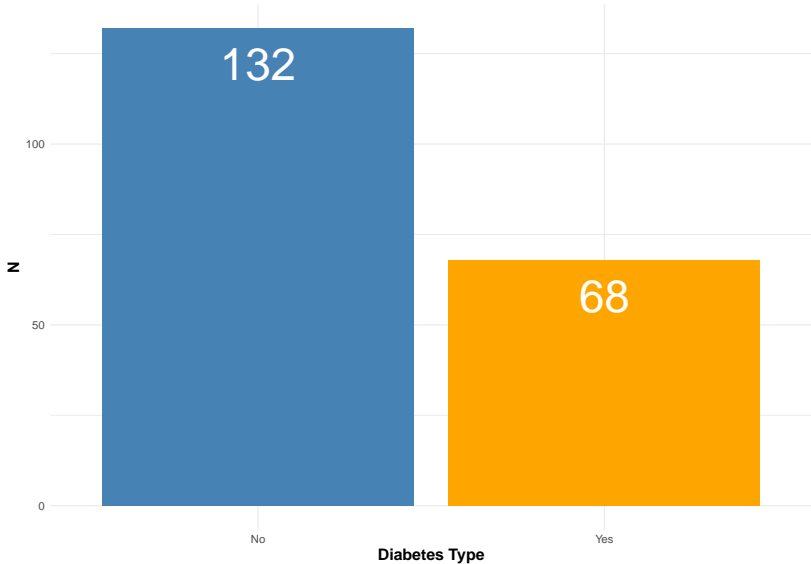
Bulguların grafikte sunumunda dikkat edilecek noktalar

- 1 Her grafiğin bir başlığı olmalıdır. Başlık grafiğin altına yada üstüne konabilir.
- 2 Eksenlerin neyi ifade ettiği belirtilmelidir. Genellikle x eksenine değişkenler y eksenine frekanslar ya da oranlar(yüzde,binde vb) konur.
- 3 Grafikte kullanılan ölçekler ve işaretler hakkında açıklayıcı bilgi konulmalıdır.
- 4 Grafik karışık olmamalıdır. Çok karışık işaretler ve çizgiler kullanılmamalıdır. Kendisini açıklayıcı olmalıdır.

- Niteliksel Veriler için
 - Çubuk grafiği
 - Bindirmeli Çubuk Grafiği
 - Pasta Grafiği
- Niceliksel Veriler için
 - Histogram
 - Ortalama ve Standart Sapma Grafiği
 - Kutu ve Çizgi Grafiği
 - Saçılım Grafiği
 - Dal ve Yaprak Grafiği

- Nitelik veriler için kullanılan çubuk grafiklerde verilerdeki kategoriler ayrı ayrı çubuklarla gösterilir. Çubukların enleri birbirine eşittir.
- Yatay eksene(x) incelenen değişkene ait kategoriler, dikey eksene (y) ise bu kategorilere ait sayı ya da yüzde değerleri konur.

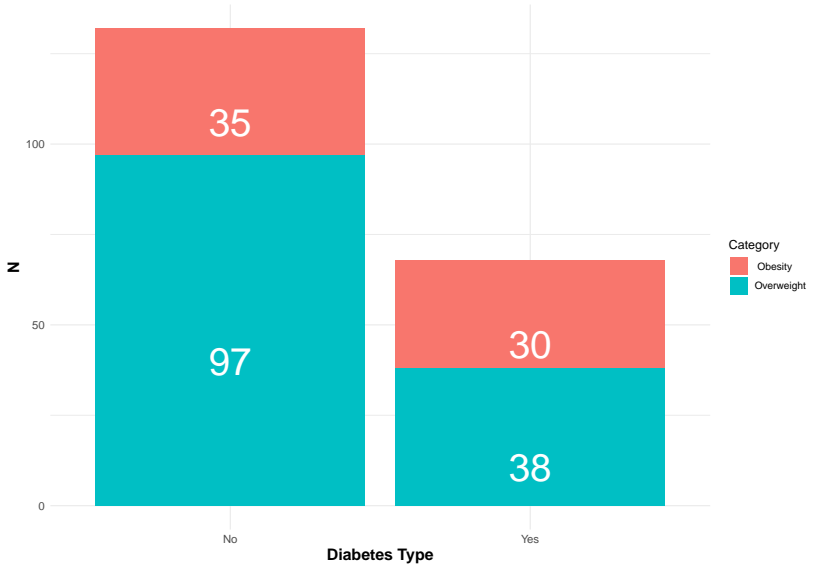
Çubuk Grafiği



Bindirmeli Çubuk Grafiği

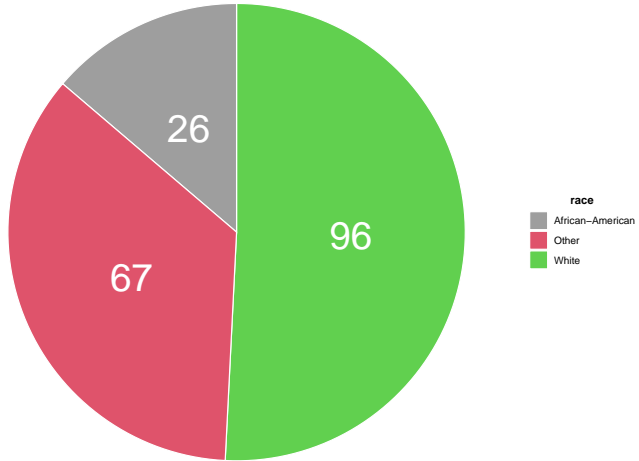
##	npreg	glu	bp	skin	bmi	ped	age	type	Category
## 1	5	86	68	28	30.2	0.364	24	No	Overweight
## 2	7	195	70	33	25.1	0.163	55	Yes	Overweight
## 3	5	77	82	41	35.8	0.156	35	No	Obesity
## 4	0	165	76	43	47.9	0.259	26	No	Obesity
## 5	0	107	60	25	26.4	0.133	23	No	Overweight
## 6	5	97	76	27	35.6	0.378	52	Yes	Obesity

Bindirmeli Çubuk Grafiği

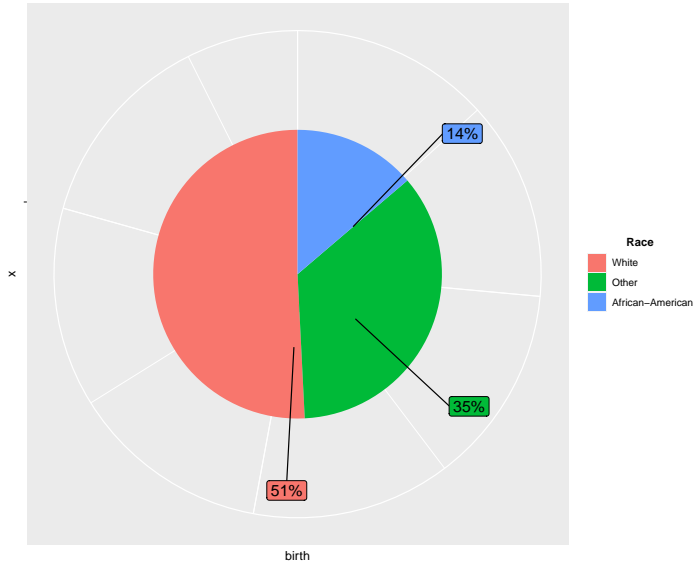


##	birthwt.race	birthwt.smoke
## 1	African-American	No
## 2	Other	No
## 3	White	Yes
## 4	White	Yes
## 5	White	Yes
## 6	Other	No

Pasta Grafiği



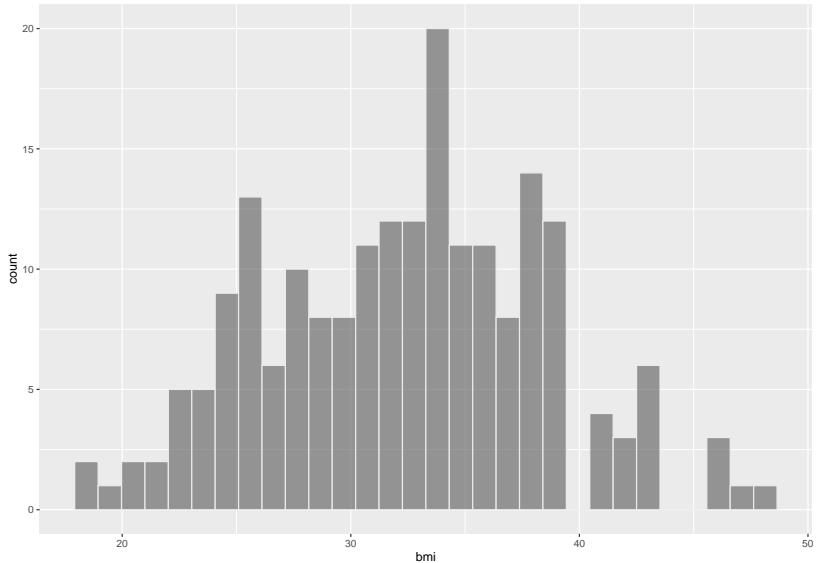
Pasta Grafiği



Histogram Grafiği

- Sürekli değişkenler için kullanılır. Çubuk grafiklerden farklı olarak her sınıftaki frekansları gösteren çubuklar birbirine bitişik olarak çizilir.
- Sayı ya da yüzde kullanmak grafiğin şeklini değiştirmez.
- Yatay ekseninde sınıf değeri dikey ekseninde sayı ya da yüzde bulunur. (Yatay eksene alt sınır ve üst sınır değerleri de yazılabilir).
- Dağılımın yapısı, tabloya göre daha belirgin hale gelir.

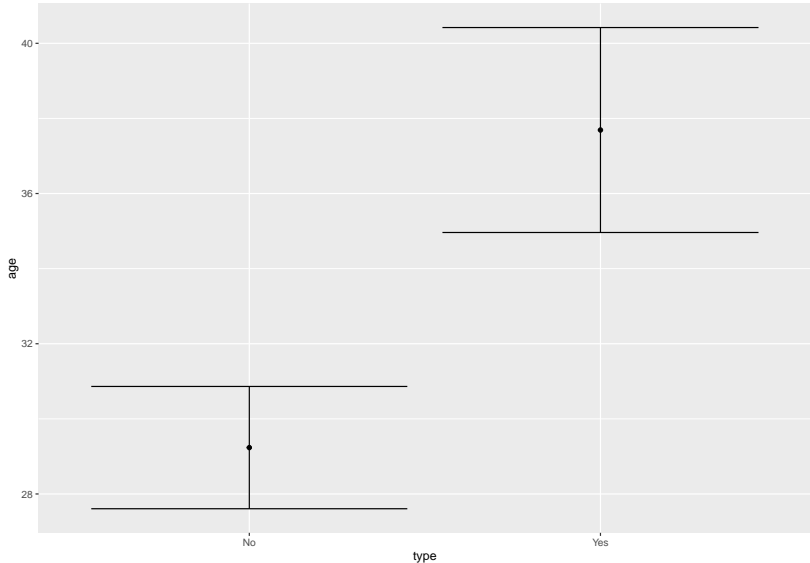
Histogram Grafiği



Ortalama ve Standart Sapma Grafiği

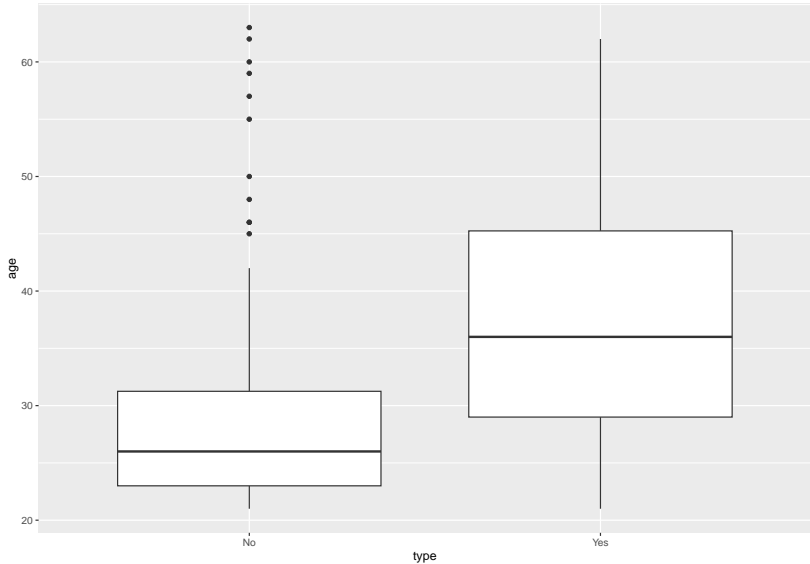
- Sürekli değişkenler için kullanılan bir grafik türüdür.
- Simetrik ve tek tepeli dağılımları tanımlamak için kullanılır.
- Genellikle aritmetik ortalamaya 1 ya da 2 standart sapmalık uzaklıkların çizgiyle birleştirilmesiyle elde edilir.

Ortalama ve Standart Sapma Grafiđi

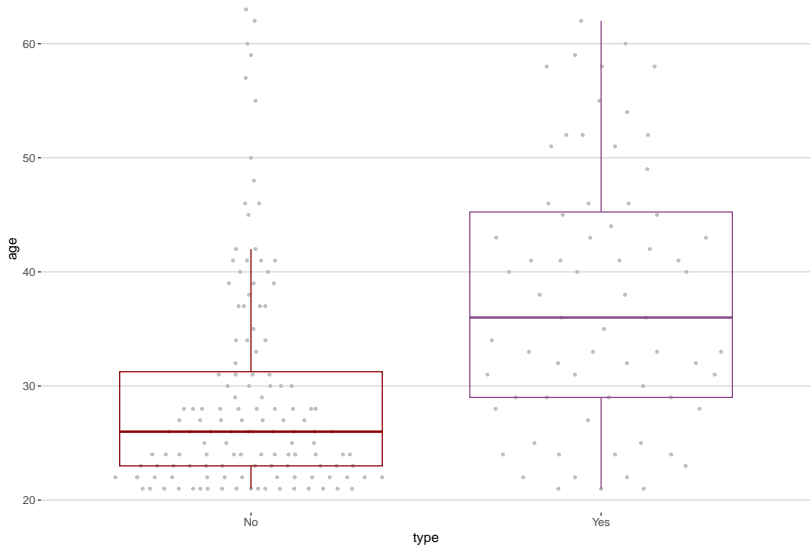


- Kutu-çizgi grafikleri, dağılımdaki 1,2,3,4. çeyrek yüzdeler ile en küçük ve en büyük değerler kullanılarak çizilir.
- Daha çok dağılım çarpık olduğunda kullanılır.
- Kutu-çizgi grafikleri, aşırı gözlemlerin varlığı hakkında bilgi verir.
 - **Aşırı Gözlem:** Kutunun alt ve üst sınırlarından bir buçuk ile üç kutu boyu arası uzakta yer alan gözlemler aşırı gözlemler olarak adlandırılır.
 - **Aykırı Gözlem:** Kutunun alt ve üst sınırlarından üç kutu boyu ve daha uzak gözlemler aykırı gözlem olarak adlandırılır.

Kutu Grafiği

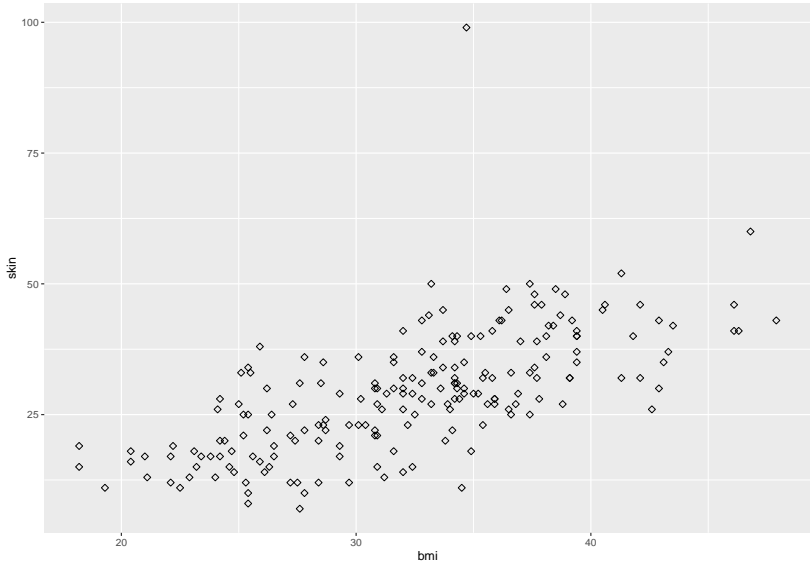


Kutu Grafiği



- Sayısal değişkenlerin birbirleriyle ilişkili olup olmadığını göstermek amacıyla kullanılan bir grafik yöntemidir.
- Bağımsız değişken x , bağımlı değişken y ekseninde yer alır. Her gözlem çiftinin kesiştiği yere bir nokta konulması ile çizilir.

Saçılım Grafiği



- Veri kümesini özetlemek için basit ve kullanışlı bir grafik türüdür. Hem grafiğin şeklini hem de dağılımdaki gözlem değerlerinin tümü bu grafik üzerinde görülebilir. *bp(diyastolik kan basıncı) değerine ilişkin dal yaprak grafiği* örneği

Dal-Yaprak Grafiği

```
##
## The decimal point is 1 digit(s) to the right of the |
##
## 3 | 8
## 4 | 0
## 4 | 888
## 5 | 002224444
## 5 | 56668888888
## 6 | 000000000000022222222224444444444
## 6 | 566666666666668888888888888888
## 7 | 000000000000000000222222222244444444444444444
## 7 | 566666666666668888888888888888
## 8 | 00000000000222222222224444444
## 8 | 556668888
## 9 | 00000244
## 9 | 5
## 10 | 2
## 10 | 6
## 11 | 0
```