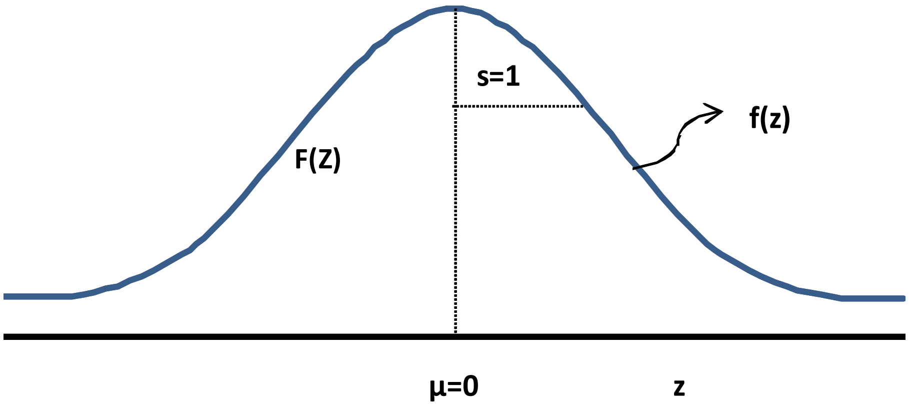
TOBB-ETÜ, İktisat Bölümü
İstatistik (İKT 253)
Normal Dağılım Çalışma Metni

- Ortalaması 0, standart sapması 1 olan normal dağılım standart normal dağılım olarak adlandırılır ve bu dağılımın değerleri z ile gösterilir.

$$Z \sim N(0,1)$$



- $f(z)$ z'nin olasılığını (gelme ihtimalini), $F(z)$ ise $-\infty$'dan z'ye kadar, yani z'den küçük tüm sayıların (kümülatif) gelme ihtimalini gösterir.
- Bu durumda aşağıdaki eşitlikler sağlanır

$$F(-\infty) = 0$$

$$F(\infty) = 1$$

- Ayrıca dağılım 0'ın etrafında simetrik olduğundan

$$F(0) = F(z < 0) = 0.5$$

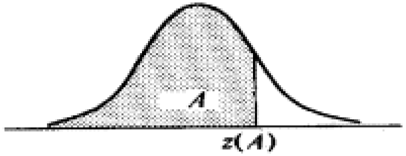
- $F(1)$ değeri ise ortalamadan 1 standart sapma daha büyük olan sayıdan küçük olan sayıların toplam oluşma ihtimalini gösterir.

$$F(1) = F(z < 1)$$

  Standart dağılım tablosundan baktığımızda ise bunun değeri $F(1) = 0.8413$'tür.

- Yani ortalamadan 1 standart sapma kadar büyük bir sayı bu dağılımdaki sayıların %84.13'ünden büyüktür.

Entry is area A under the standard normal curve from $-\infty$ to $z(A)$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890

- Herhangi bir normal dağılım için de aynı mantık geçerlidir. Öyle ki aşağıdaki normal dağılım için (ortalaması 4, standart sapması 3)

$$X \sim N(4, 9)$$

sayıların yarısı ortalamanın altında kalacak

$$F(4) = 0.5$$

sayıların %84.13'ü ise ortalamadan 1 standart sapma kadar büyük olan 7 sayısının altında kalacaktır

$$F(7) = 0.8413$$

- Dolayısıyla X dağılımında herhangi bir x sayısının altında kalan sayıların kümülatif ihtimalini–$F(x)$– bulmak için yapılması gereken bu x sayısının ortalamadan kaç standart sapma uzakta olduğuna bakmak (ortalamadan mesafesini standart sapma cinsinden ifade etmek)

$$\frac{x-\mu}{\sigma}=z$$

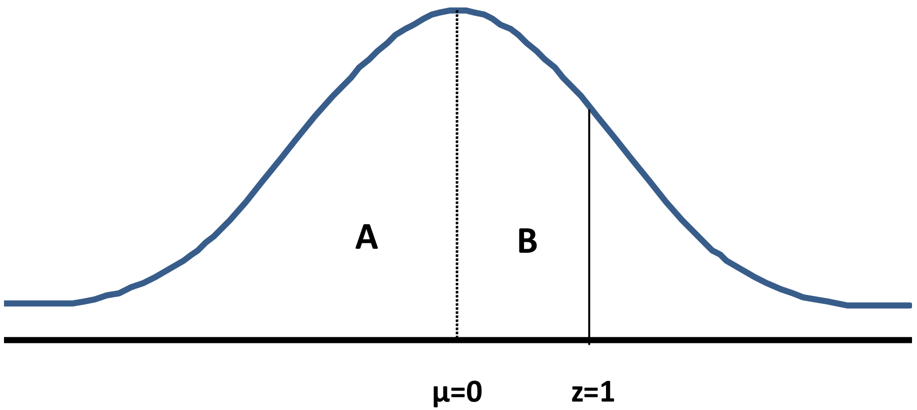
daha sonra ilgili $F(z)$ değerini standart normal dağılım tablosunda aramak olacaktır

- Dolayısıyla herhangi bir normal dağılım aşağıdaki gibi standardize edilebilir ve standart normal dağılım tablosundan yararlanılabilir

$$Z = \frac{X - \mu}{\sigma}$$

- Şimdi yukarıda bulduğumuz alanı aşağıda A+B şeklinde gösterelim ve B alanının ihtimalini bulalım

$$B = F(0 < z < 1)$$



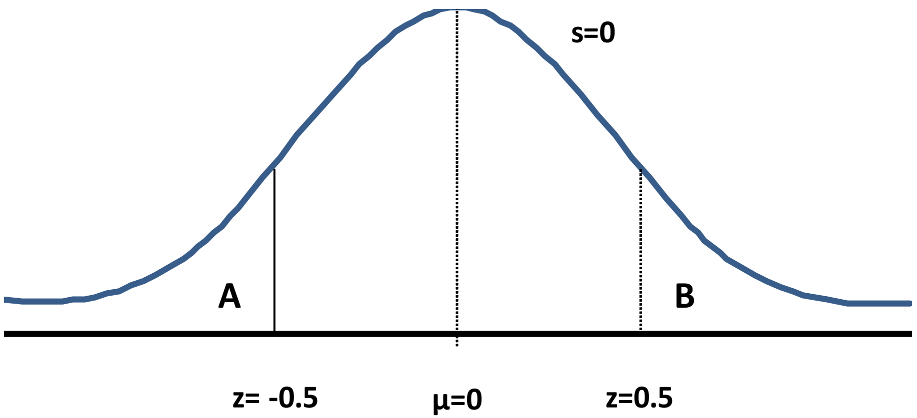
- B alanının ihtimali biraz önce bulduğumuz $F(1)$ alanından $F(0) = 0.5 = A$ alanını çıkararak bulunabilir.

$$\underbrace{F(0 < z < 1)}_{B} = \underbrace{F(1)}_{A+B} - \underbrace{F(0)}_{A}$$
$$= 0.8413 - 0.5 = 0.3413$$

- Ortalamadan en az 1 standart sapma daha büyük olan sayıların toplam oluşma ihtimali ise

$$F(1 < z) = F(1 < z < \infty) = F(\infty) - F(1)$$
$$= 1 - 0.8413 = 0.1587$$

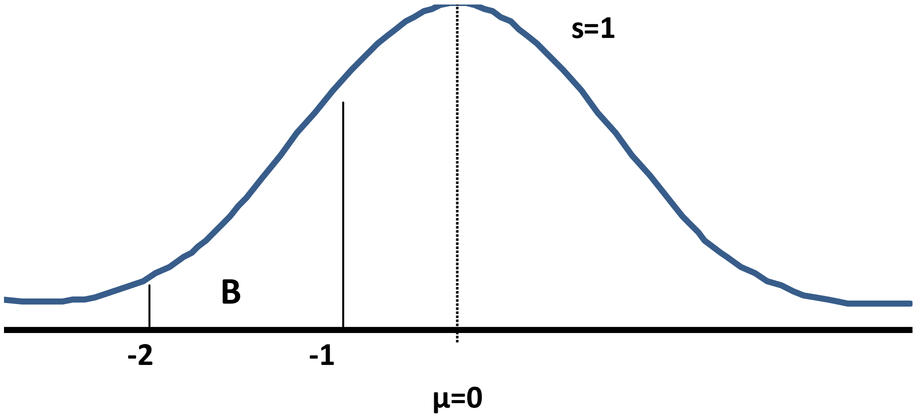
- Standart dağılım tablosu sadece pozitif z değerleri veriyor. Negatif değerler ise simetri yardımı ile bulunuyor.



- Örneğin $F(z < -0.5)$ alanını A ile gösterelim. Dağılım simetrik olduğundan bu alan $z = 0.5$'in üstünde kalan B alanına eşittir.
- B alanı ise tüm alan olan 1'den, $F(0.5)$ alanının çıkarılmasıyla bulunabilir.

$$A = B \quad \Rightarrow \quad F(-0.5) = 1 - F(0.5) = 1 - 0.6915 \approx 0.30$$

- Şimdi de aşağıdaki B alanını bulalım.



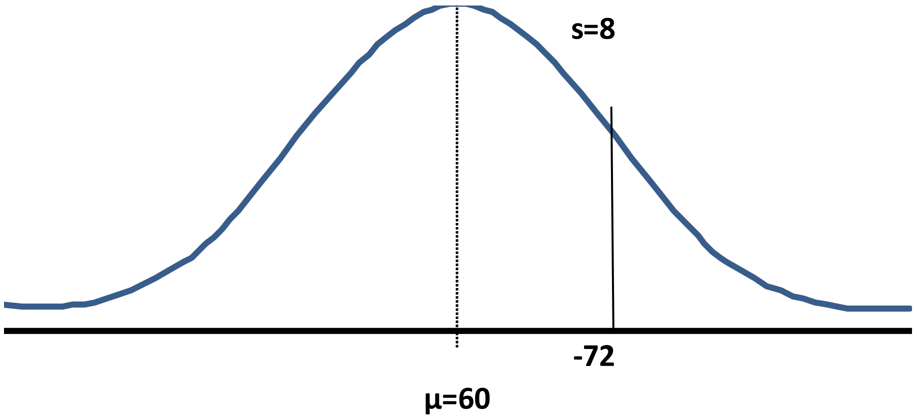
$$B = F(-1) - F(-2)$$

- Yani -∞'dan -1'e kadar olan sayıların gelme ihtimalinden, -∞'dan -2'ye kadar olan sayıların gelme ihtimallerini çıkarırsak -2 ile -1 arasındaki sayıların gelme ihtimallerini buluruz.

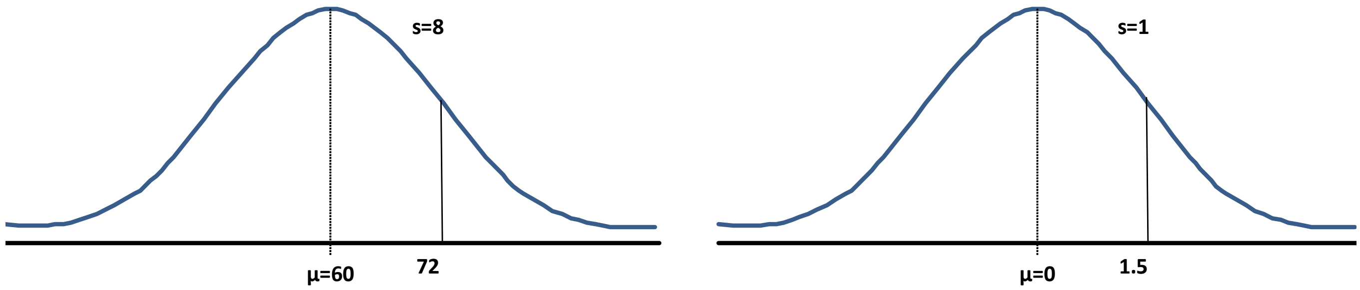
$$\begin{aligned} B &= F(-1) - F(-2) = [1 - F(1)] - [1 - F(2)] \\ &= F(2) - F(-1) \\ &= (1 - 0.8413) - (1 - 0.9772) \\ &\approx 0.13 \end{aligned}$$

- Peki standart olmayan bir dağılım için bu tip ihtimaller nasıl bulunur?

Örneğin ortalaması 60, standart sapması 6 olan bu dağılımdaki sayıların % kaçı 72'den küçüktür?



Bu soruyu cevaplamak için aşağıdaki 2 şekli karşılaştıral



2 şekilde de eğrilerin altında kalan tüm alan 1'dir (yani %100, yani 1 ihtimalle olası sonuçlardan biri gelmek zorundadır).

İlk şekilde 72'nin altında kalan lan ile ikinci şekilde 1.5'in altında kalan alan,birbirlerinin aynıdır.

Zira 72 sayısı, kendi ortalaması olan 60 sayısından1.5 standart sapma ($1.5 \times 8 = 12$) uzaktadır. Yine 2. şekilde 1.5 sayısı dağılımın kendi ortalaması olan 0 sayısından 1.5 standart sapma uzaktadır. (İki çizgi de alanı 1 olan şekilleri benzer noktadan kesmiş oluyorlar.) Dolayısıyla $P(x < 72) = P(z < 1.5)$'tur. İşte x 'i bu şekilde z'ye çevirip

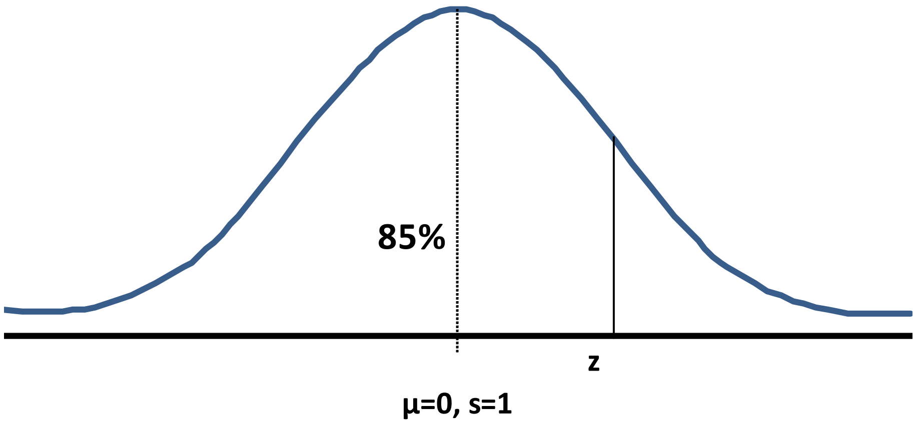
$$z = (72 - 60)/8 = 1.5$$

Artık standart dağılım tablosunu kullanabiliriz.

$$P(x < 72) = F(1.5) = 0.9332$$

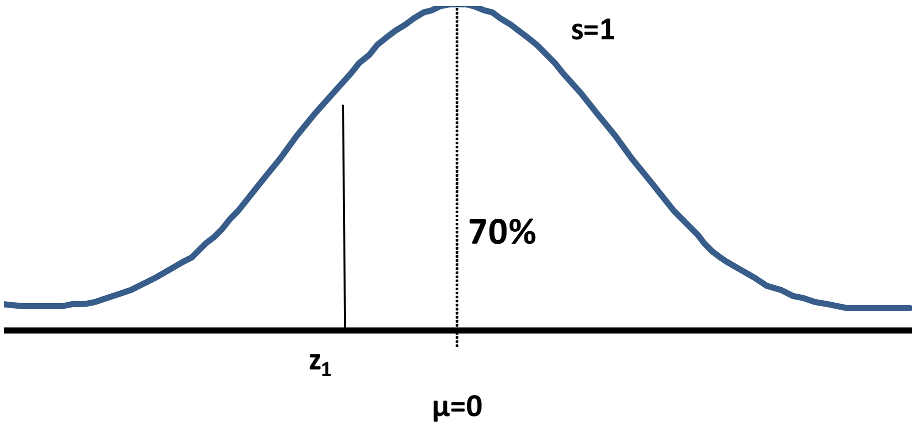
72 sayısı kendi dağılımındaki sayıların %93'ünden büyüktür.

- Peki sorumuz şu şekilde olsaydı;



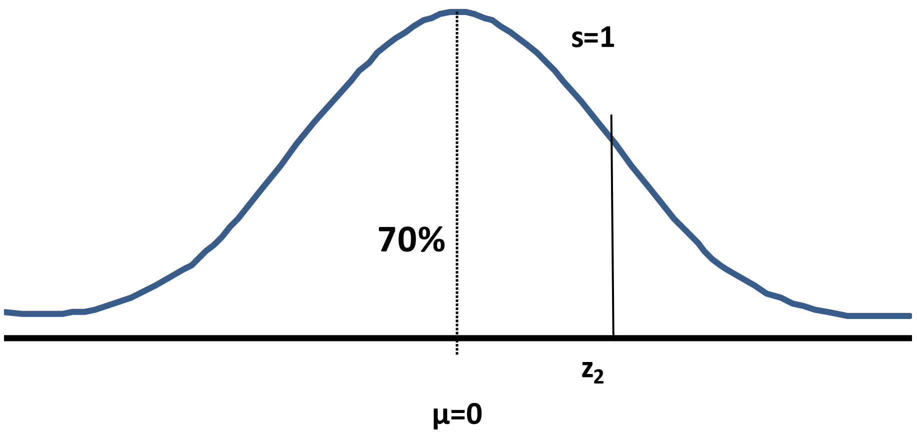
Hangi z sayısı toplam dağılımdaki sayıların %85'inden büyüktür? Bu sefer tersten gidip, standart dağılım tablosunun içindeki sayılardan 0.85'i bulup; buna karşılık gelen z'yi bulacağız. **Cevap:**1.04

- Yeni soru:



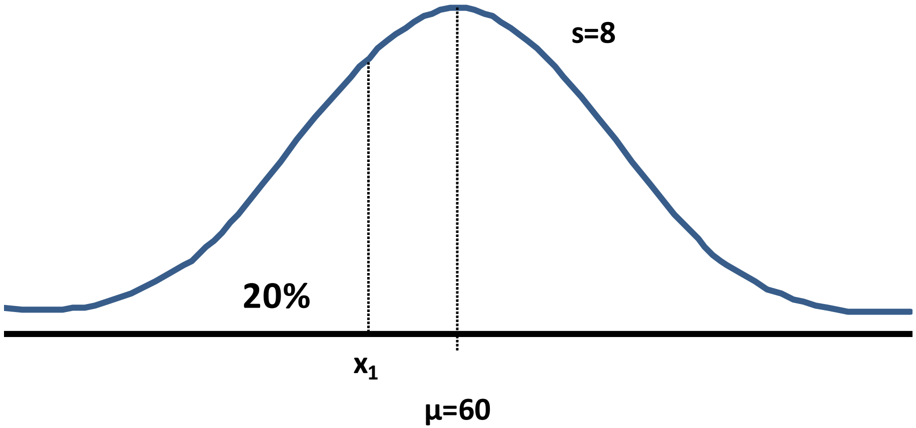
Hangi z_1 sayısı toplam sayıların %70'inden küçüktür?

z_1 bu sefer (-) bir sayı olacağı için onu tabloda bulamayacağınızdan, soruyu şu şekilde çevirip aşağıdaki z_2 ‘yi bulabiliriz.(z_1, bu z_2’nin (-)’lisi olacaktır.)



$$F(z_2) = 0.7 \Rightarrow z_2 = 0.53$$
$$\Rightarrow z_1 = -0.53$$

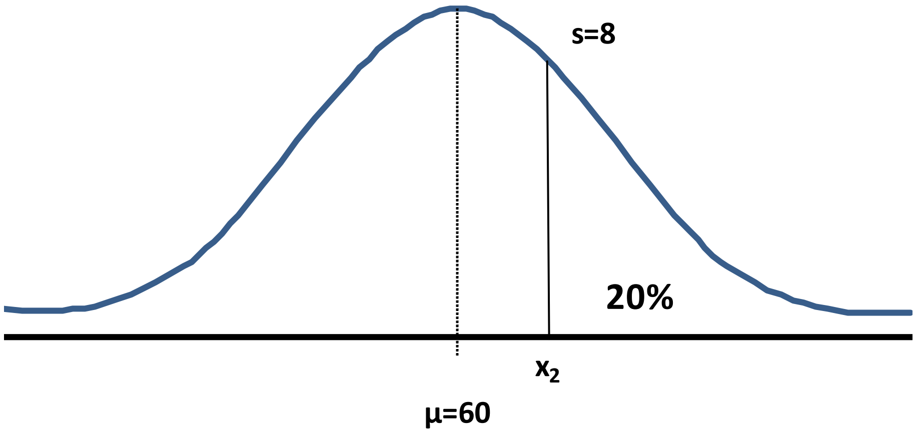
- Yeni soru:



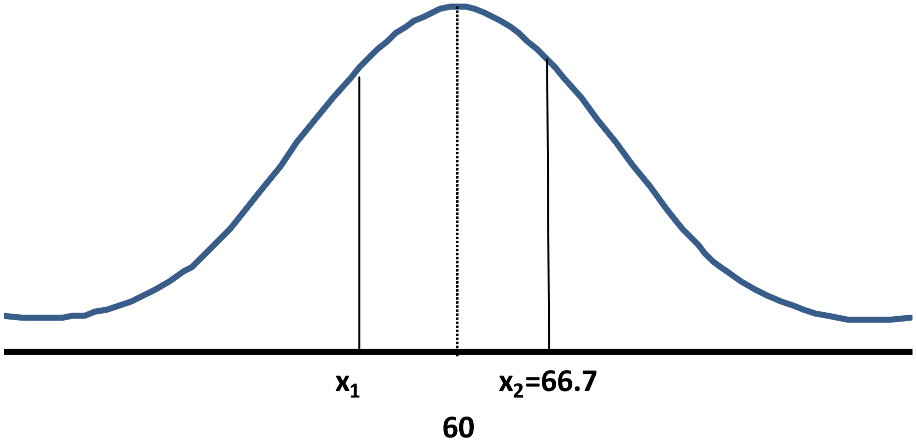
Bu normal dağılımda -$X\tilde{}N(60,6)$– hangi x_1 sayısı sayıların %20'sinden büyüktür?

Cevap: Bu değer, alttaki şekilde sayıların %20'sinden küçük x_2 değerinin sol taraftaki simetriğidir. Dolayısıyla önce x_2'yi bulalım.

Onu da sayıların %80'inden büyük bir değer gibi düşünebiliriz.

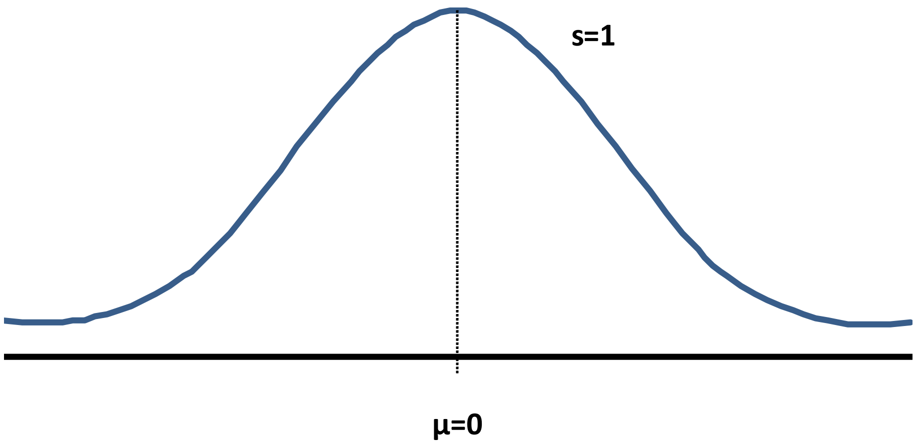


$$F(z) = 0.8 \quad \Rightarrow \quad z = 0.84$$
$$(x_2 - 60)/8 = z \quad \Rightarrow \quad x_2 = 66.7$$

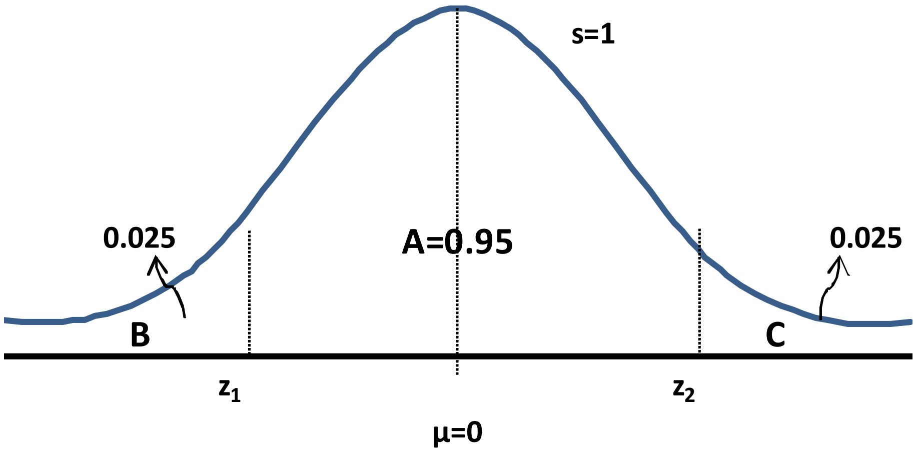


$$\Rightarrow \quad x_1 = 60 - [66.7 - 60] = 53.4$$

Güven Aralığı Oluşturmak:



Bu standart normal dağılımda hangi aralıktaki sayılar $\mu = 0$'ın etrafındaki ona en yakın sayıların %95'ini içerir?



Yani tüm sayıların %95'inin istemek demek, kenardaki %2.5'luk bölümleri istememek demektir.

- Peki $z_1$ve z_2 nasıl bulunur?

z_2'den başlayalım. z_2 tüm sayıların %97.5'undan büyük bir sayıdır.

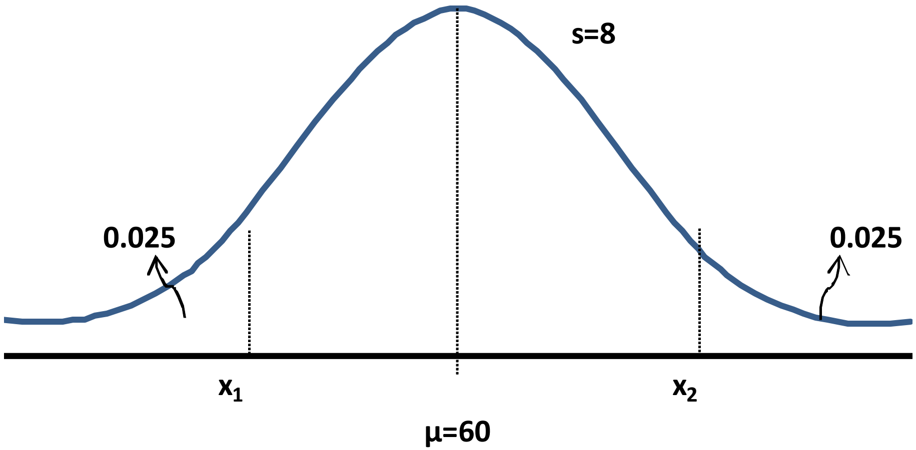
$$\Rightarrow F(z_2) = 0.975 \Rightarrow z_2 = 1.96$$

Simetrik olarak ; $z_1 = -1.96$

Yani $[-1.96, 1.96]$ aralığı tüm sayıların %95'ini içerir.

- Bu dağılımda ortalamanın etrafındaki sayıların

%95'ini içeren güven aralığı nedir?



Bu dağılımda ortalamanın etrafındaki sayıların %95'ini içeren güven aralığı nedir?

Cevap: Zaten $z_1$ve z_2, -1.96 ve 1.96 olarak bulundu.

$$(x_1 - \mu)/s = (x_1 - 60)/8 = -1.96 \Rightarrow x_1 \approx 44$$
$$(x_2 - \mu)/s = (x_2 - 60)/8 = 1.96 \Rightarrow x_2 \approx 76$$

Genel formül:

$$(x - \mu)/s = \pm\, z_{1-\alpha/2} \text{ (burada } \alpha = 0.05)$$

dolayısıyla güven aralığı

$$[\ \mu - z_{1-\alpha/2}.s\ ,\ \mu + z_{1-\alpha/2}.s]$$