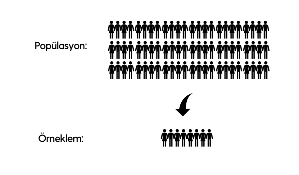
**VERİ OKUR YAZARLIĞI(16punto)**

**Temel Kavramlar**

**1-Popülasyon ve Örneklem**



Popülasyon: Verinin tamamını kapsar. Ana kütlemizdir.

Örneklem: Veriyi temsil eden alt kümeye örneklem denir.

Örneklem, popülasyonun niteliklerini ve özelliklerini temsil etmesi ve yansız olması gerekir. Ana kütleden örneklem çekmek için bazı yöntemler var: Tabakalı Örnekleme, Rasgele Örnekleme vs..

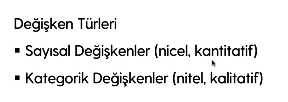
**2-Gözlem Birimi**

Araştırmada incelediğimiz birimlerdir. Örneğin seçim çalışması için yapılan analizde her bir mikrofon uzatılan bireyler gözlem birimidir. Ya da aşağıdaki tabloda satırlarımız gözlem birimi sütunlar ise değişkenlerimizdir.



**3-Değişken ve Değişken Türleri**

Değişken: Birimden birime farklı değerler alan niceliktir.



Sayısal Değişkenler: Tablodaki sayılar

Kategorik Değişkenler ve Kategorik Değişkenlerin Sınıfları:

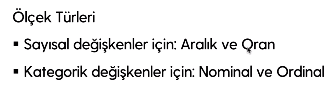
Yukarıda tablodaki vites türü, model ya da başka bir örnek olarak Cinsiyet “Kategorik Değişkendir”

Vites türündeki Manuel/Otomatik Cinsiyetteki Kadın/Erkek ise “Kategorik Değişken Sınıflarıdır.”



**4-Ölçek Türleri**

Bir değişkenin değerlerini insan okumak anlamaya bilmemiz adına bunu ölçmemiz gerekiyor. İşte buradaki ölçek türleri “Sayısal Değişkendir” için ve “Kategorik Değişkenler” için ikiye ayrılıyor.



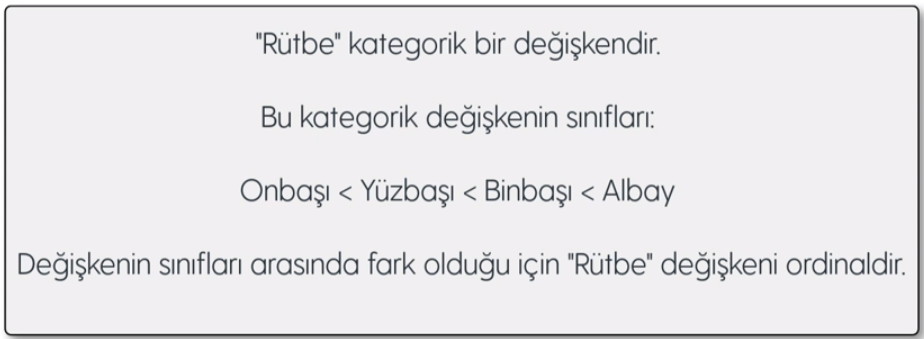
Elimizde sıcaklık değerlerinin olduğunu varsayalım. Sıcaklık değerlerin negatif veya pozitif olduğunu biliyoruz. Bu sebeple Negatif ve pozitif değerler alıyorsa “Aralık” ölçek türüdür.

Başlangıç noktasını sıfır kabul eden sayısal değişkenlerin ölçek türü ise Orandır.

String türünde olan Kategorik değişkenlerde eğer kategorilerin sınıfları arasında bir fark olmadığında buna Nominal ölçek türü denir. Örneğin Kadın/Erkek:

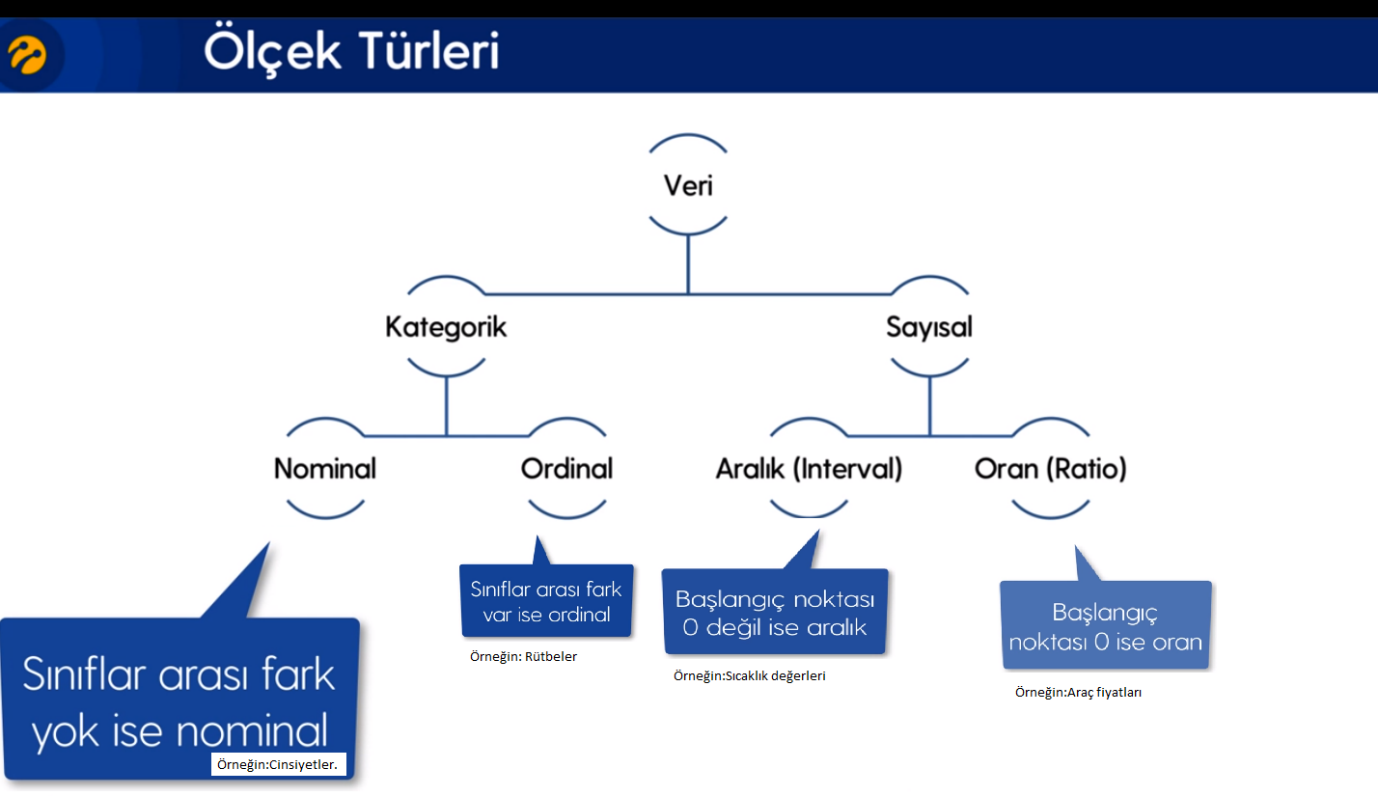


Sınıflar arasındaki fark olması durumunda (rütbe/eğitim durumu) bu ise “Ordinal” ölçek türüdür.





Özetleyecek Olursak:



Veri okuryazarlığı:

**Merkezi Eğilim Ölçüleri(14Pt)**

**1-Aritmetik Ortalama:**

Bir seride yer alan tüm elemanların değerlerinin toplamının eleman sayısına bölünmesi ile elde edilen istatistiktir.

Aritmetik ortalama Seri dağılımının simetrik olduğu bilindiğinde kullanılabilir

**2-Medyan (Ortanca):**

Bir seriyi küçükten büyüğe sıraladığımızda tam ortak noktadan seriyi iki eşit parçaya ayıran değere medyan adı verilir.

n tek ise ---> (n+1)/2

n çift ise---> (n/2 + ((n/2)+1))/2

Aritmetik ortalama Seri dağılımının simetrik olduğu bilindiğinde kullanılabilir. Simetrik değil de medyanı kullanmak daha uygundur.

Örnek

13,10,15,12,17,12,19,18,11,12,190

Veri setimiz 190 olması sebebiyle simetrik değildir

Ortalaması = 28,5

Medyanı = 13 ---> Simetrik olmadığı için medyanı kullanmak önemlidir.

**3-Mod(Tepe Değeri):**

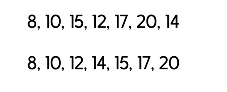
Bir seride en çok tekrar eden değere Mod denir.

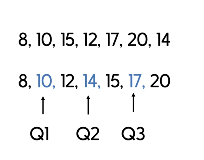
Örnek

13,10,13,12,17,13,14 ---> Modu 13 tür.

**4-Kartiller:**

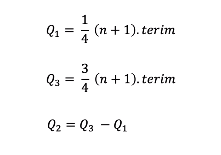
Küçükten büyüğe sıralanan bir seriyi 4 parçaya ayıran değerlere kartiller denir. Verimizin Yüzde 25 lik kısmına sağdan ve soldan gitmek istediğimizde kartilleri kullanabiliriz.





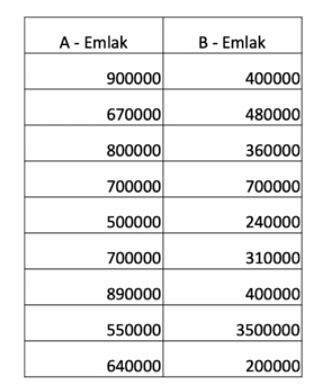
Q1=1.Çeyrek Değeri Q2=2.Çeyrek Değeri Q3=3.Çeyrek Değeri

Aşağıdaki gibi bulabiliriz:

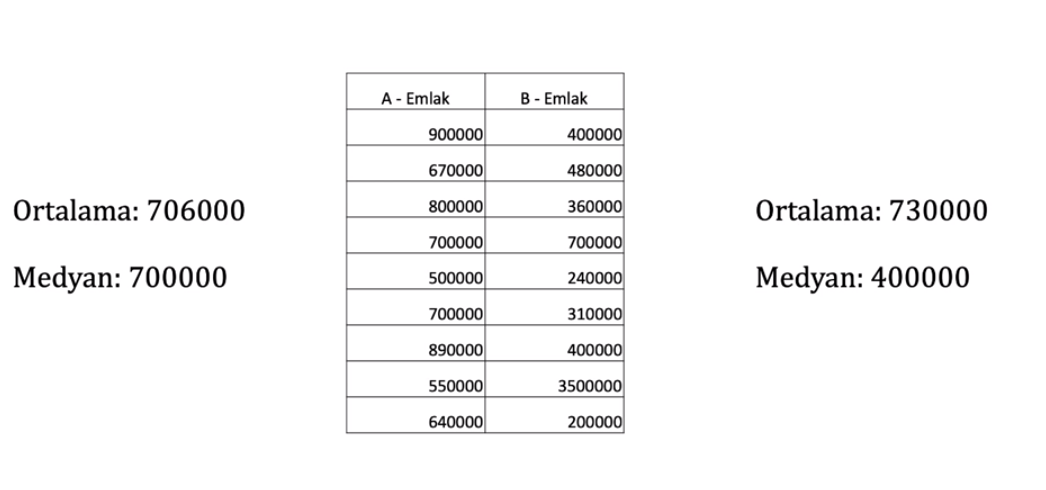


**4-Merkezi Eğilim:**

İki emlakçının da üst segmentlere hitap ettiğini düşünüyorlar ancak veri biliminde istatistiksel olarak gerçekte hangisi daha üst segmentlere hitap ettiğine bakalım.



Portfoy temsil istatistiğini yani ortalama, medyan... vs gibi istatistikleri ortaya koymamız gerek.



Yukarda da görüldüğü gibi A Emlakçısının Değerleri birbirine yakın yani homojen ancak B Emlakçısının verisinde sadece tek bir fiyat (aykırı bir veri) ortalamayı yükseltti bu sebeple ortalamayı ele alamayız ve aslında A emlakçısının daha üst segmentlere hitap etmektedir diyebiliriz.

**Dağım Ölçüler**



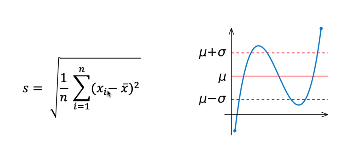
**1-Değişim Aralığı:**

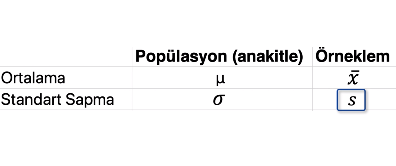
Değişim Aralığı= Max Değer- Min Değer

Örneğin bir şirketteki maaşları ele alalım. Maaşlardaki “değişim aralığı” düşük olan yerde daha adil(Departmanlar arasındaki farkın az olduğu) maaş dağılımı olduğu yüksek olan yerde ise bunun tam tersi söz konusudur

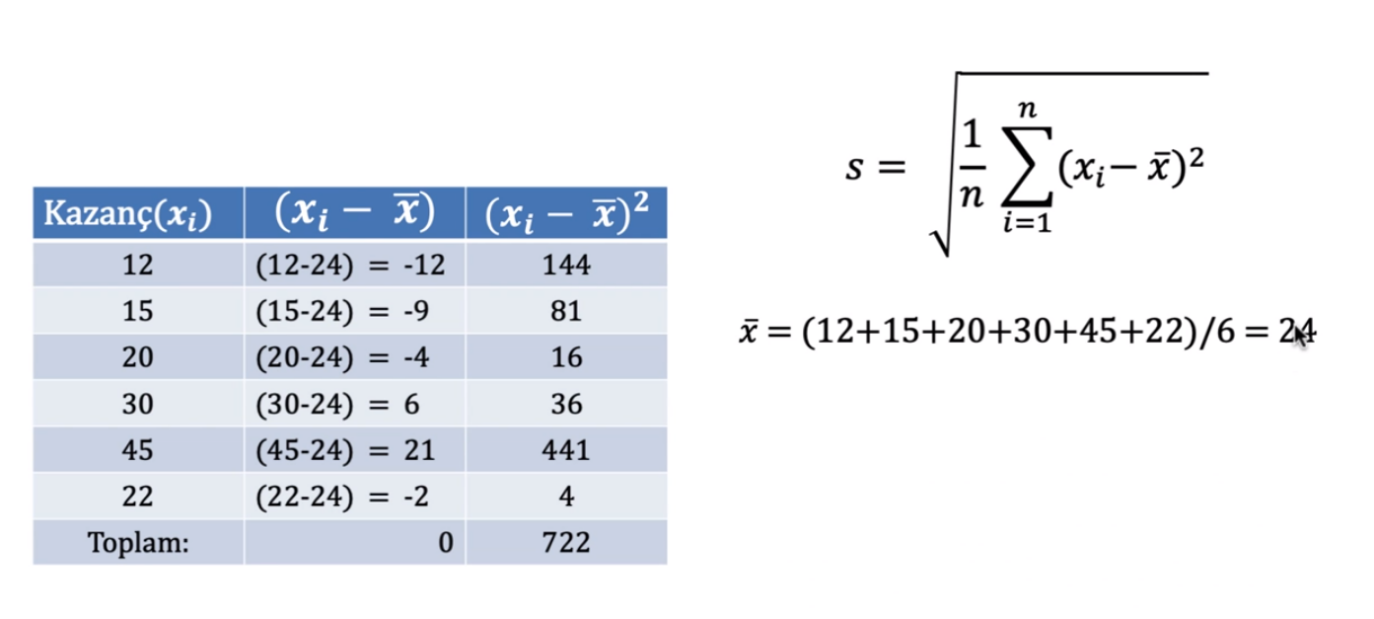
**2-Standart Sapma:**

Ortalamadan olan sapmaların genel bir ölçüsüdür yani ortalamaya olan uzaklıktır. Diğer bir ifadeyle ortalamadan olan sapmaların ortalamasıdır.



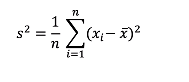


Bir tane örnek ele alalım:

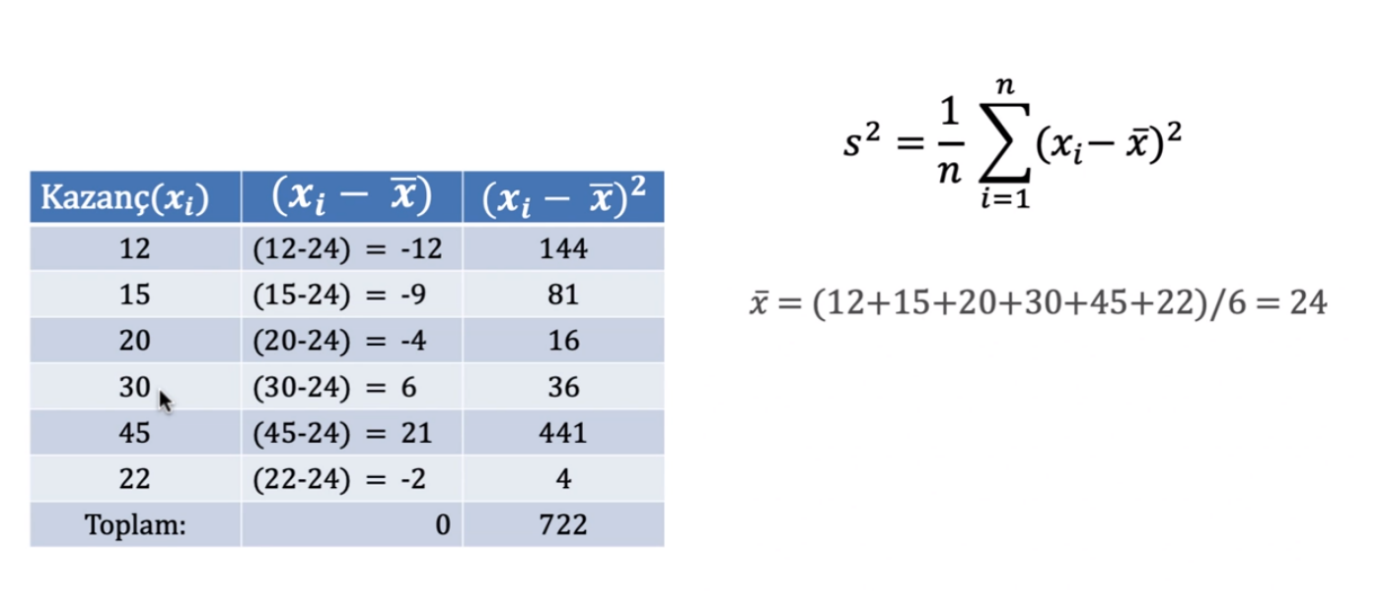


**3-Varyans:**

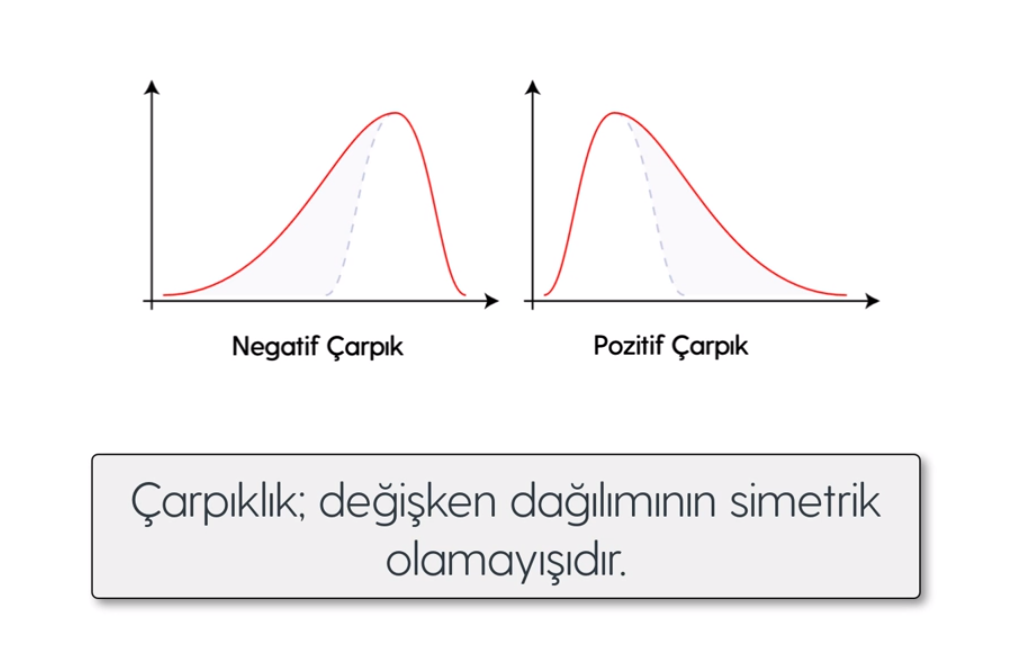
Birden fazla değişkenin dağılımını kıyaslamak için kullanılabilir.

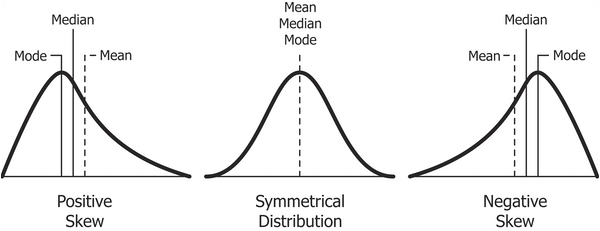


Örnek:

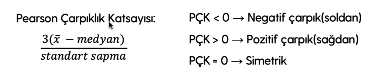


**4-Çarpıklık:**





Not: Simetrik bir dağılımda Ortalama ve Medyan eşit veya birbirine yakın olur. Diğer bir değişle Ortalama ve Medyan birbirine uzak ise bu da simetrik bir dağılım olmadığının işaretini verir.



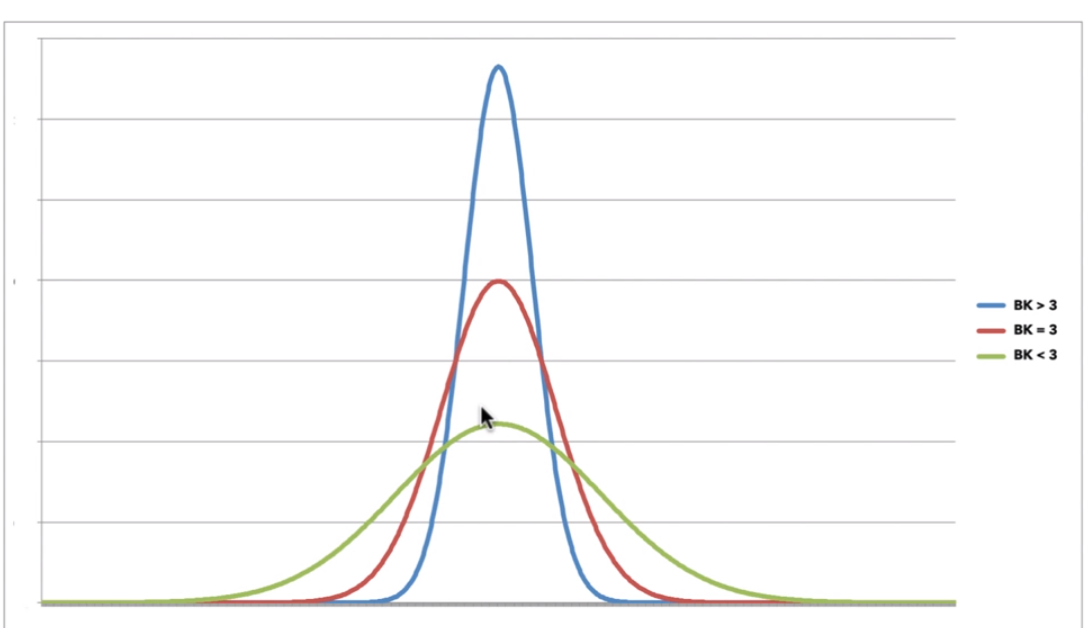
Bir örnek ile hesaplayalım ve yorumlayalım:

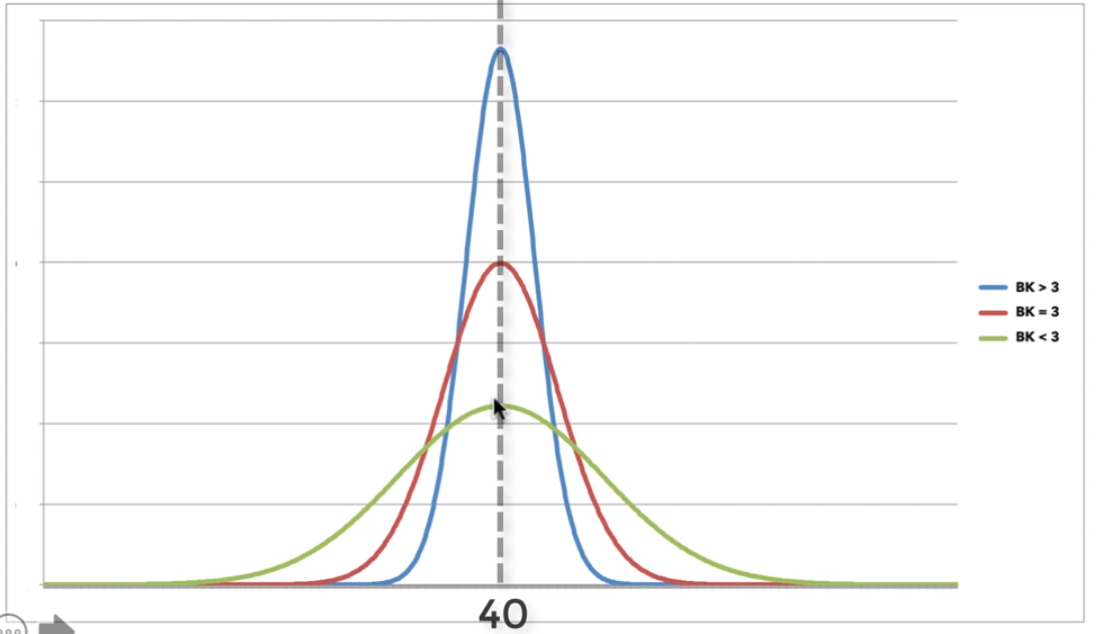


Yukarıda da görüldüğü gibi Pearson Çarpıklık Katsayısı 0,82 olarak çıktı yani pozitif çarpıklık söz konusudur.

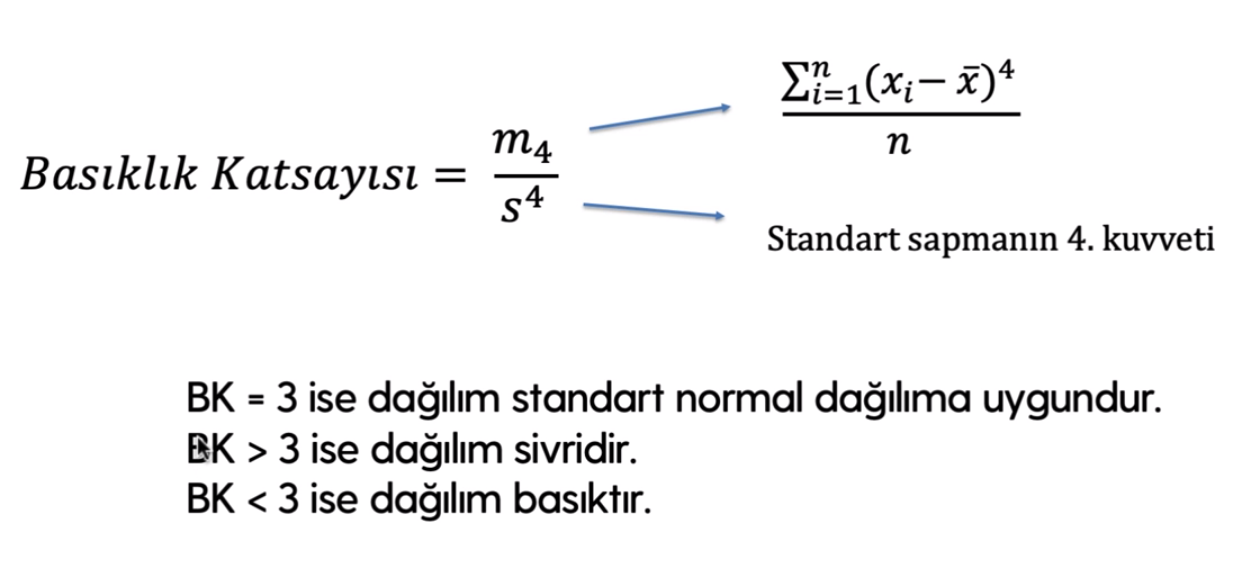
**5-Basıklık:**

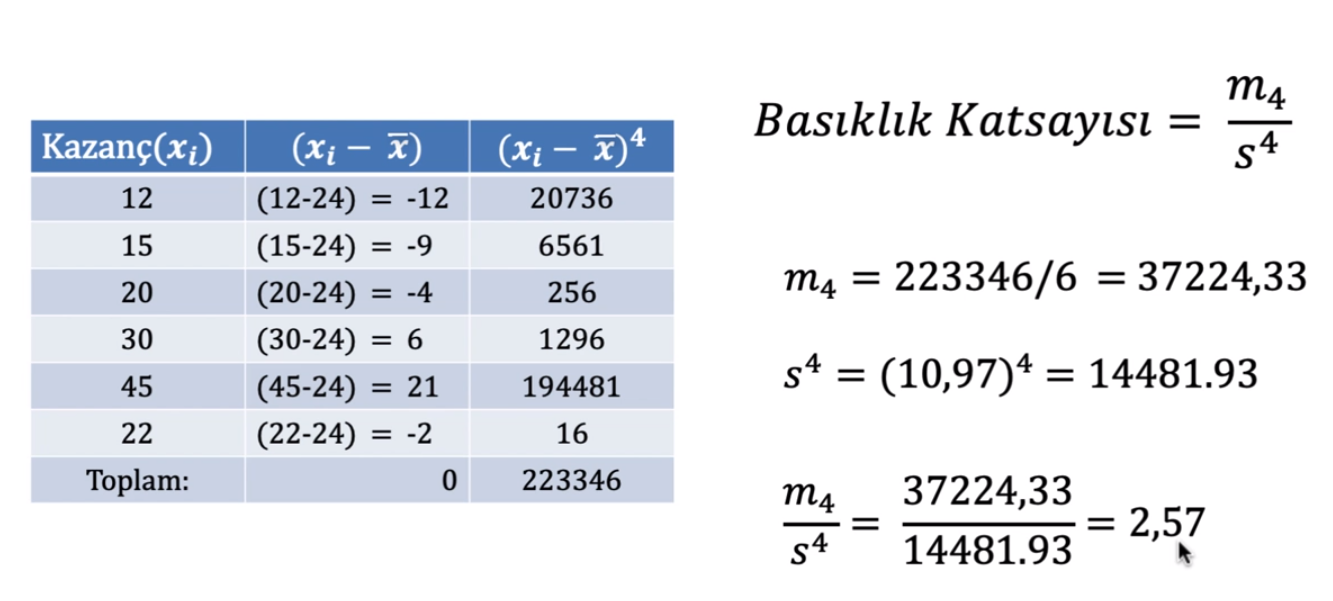
Değişken dağılımın sivriliğini/basıklığını gösteren istatistiktir.





Not:40 değeri Temsiliğidir !





Basıklık 3 ten küçük yani dağılım basıktır ! YANİ: Veri setimizin daha çok ortalamadan(ort=24) uzak değerlerden oluştuğunu söyleyebiliriz !

**Bunun haricinde söylenen varsayımlar:**

+1.0 -1.0 için verebileceğiniz kaynak: Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2013). Multivariate Data Analysis: Pearson Education Limited.

+1.5 -1.5 için kaynak: Tabachnick and Fidell, 2013 B.G. Tabachnick, L.S. Fidell Using Multivariate Statistics (sixth ed.)Pearson, Boston (2013)

+2.0 -2.0 için kaynak: George, D., & Mallery, M. (2010). SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update (10a ed.) Boston: Pearson

ilgili atıflarda cümlelerin geçtiği yerler ve ilgili ifadeler aşağıdaki şekildedir:

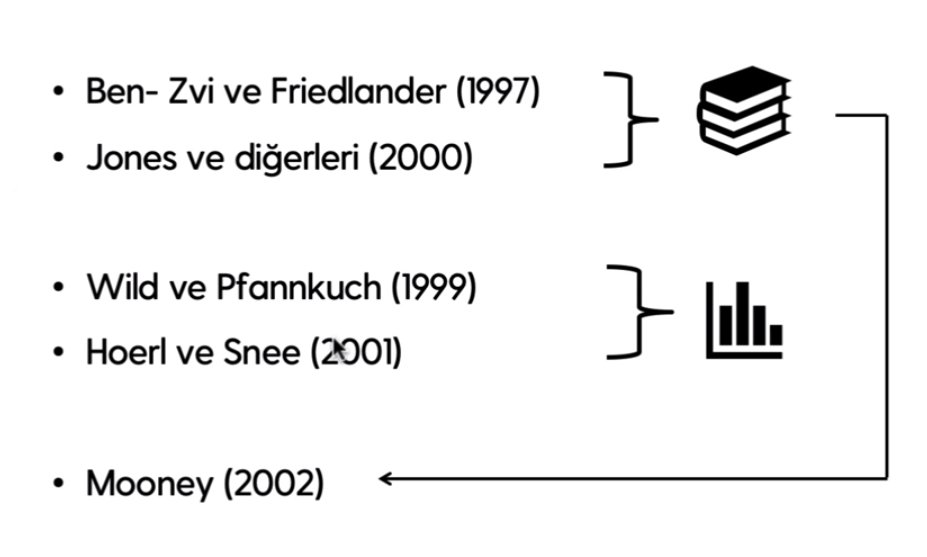
"A kurtosis value between ±1.0 is considered excellent for most psychometric purposes, but a value between ±2.0 is in many cases also acceptable, depending on the particular application." (George & Mallery (2012)) "Skewness Measure of the symmetry of a distribution; in most instances the comparison is made to a normal distribution. A positively skewed distribution has relatively few large values and tails off to the right, and a negatively skewed distribution has relatively few small values and tails off to the left. Skewness values falling outside the range of -1 to +1 indicate a substantially skewed distribution." (Hair, Black, Babin & Anderson (2013)).

**İstatistiksel Düşünce**

**İstatistiksel Düşünce Modelleri**

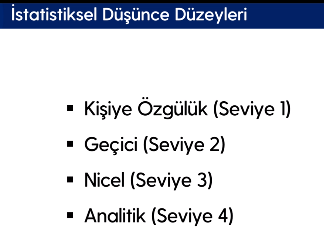
Veri okuryazarlığından veri analitiğine giden yolu modelleyen yol göstericilerdir.

Bazı Teorik çalışma modelleri aşağıda listelenmiştir.



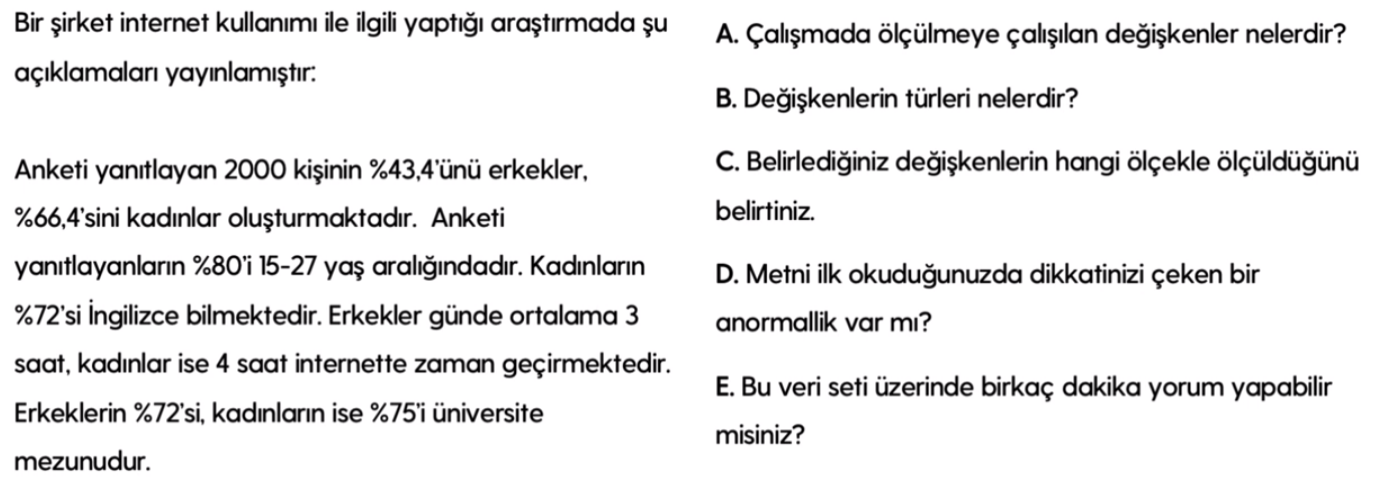
Biz burada Mooney modelini ele alarak konumuza devam edeceğiz. Mooney modeli 4 aşamadan oluşur. Bunlar:

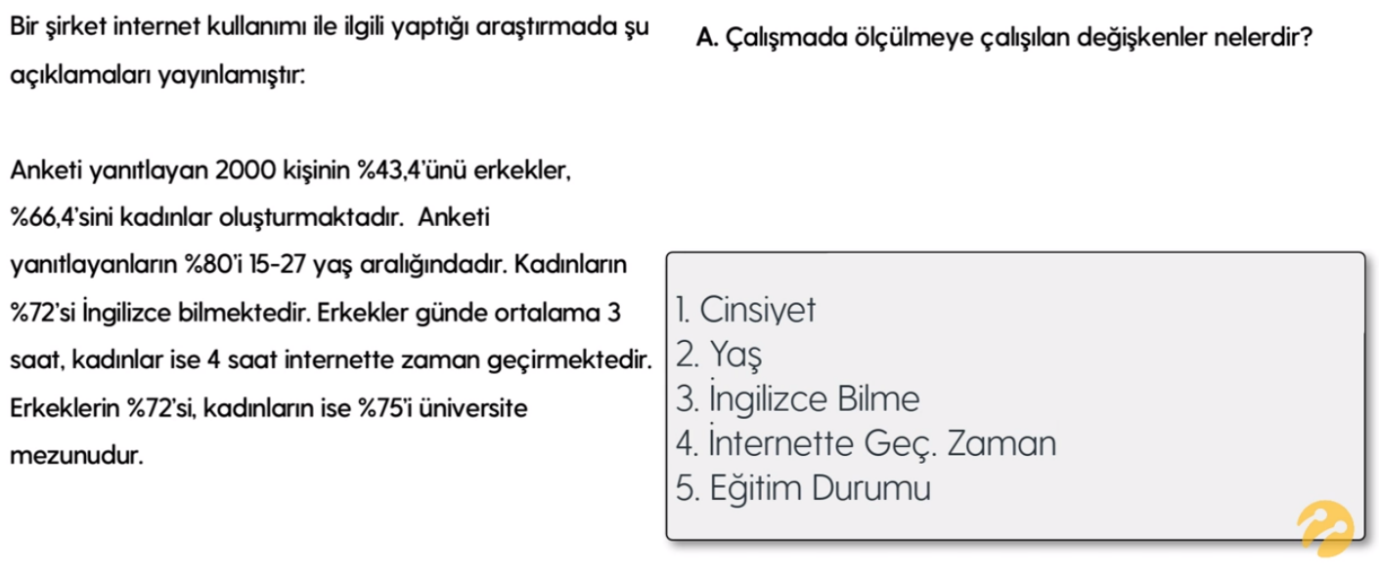


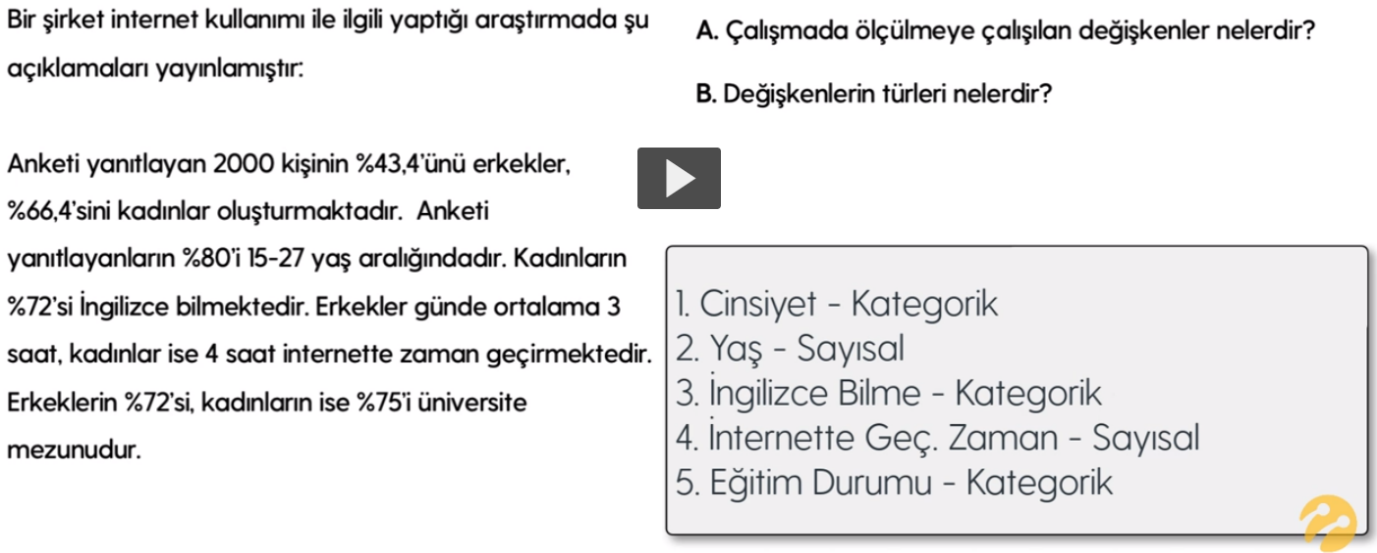


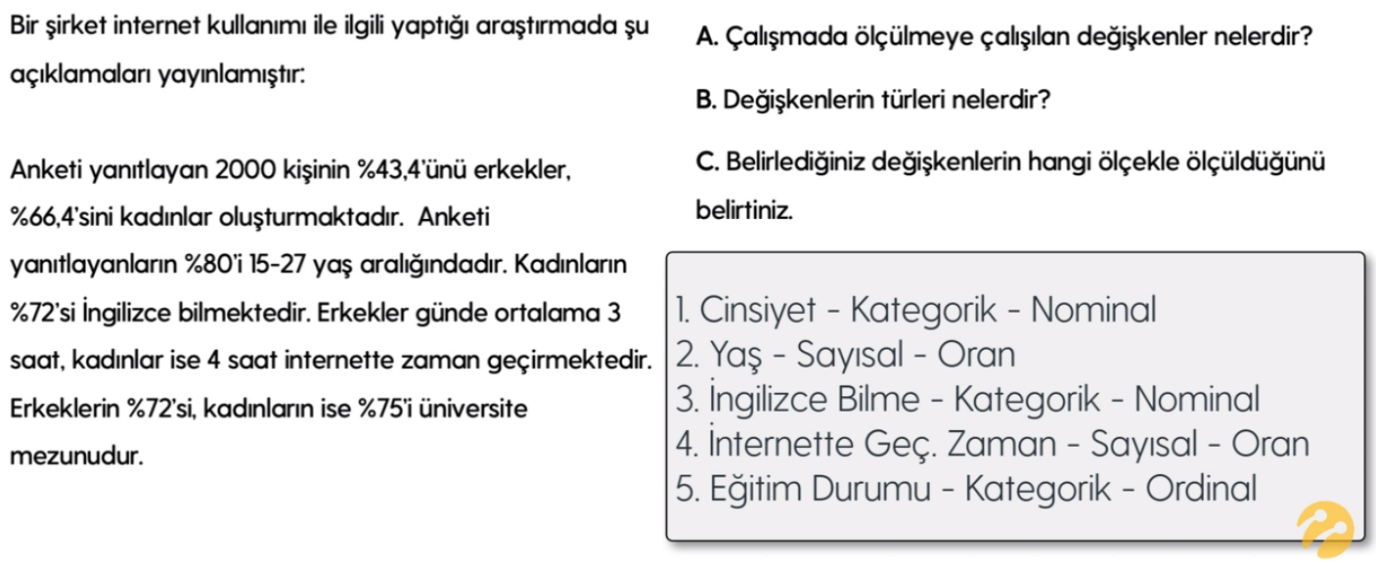
**1-Verinin Tanımlanması**

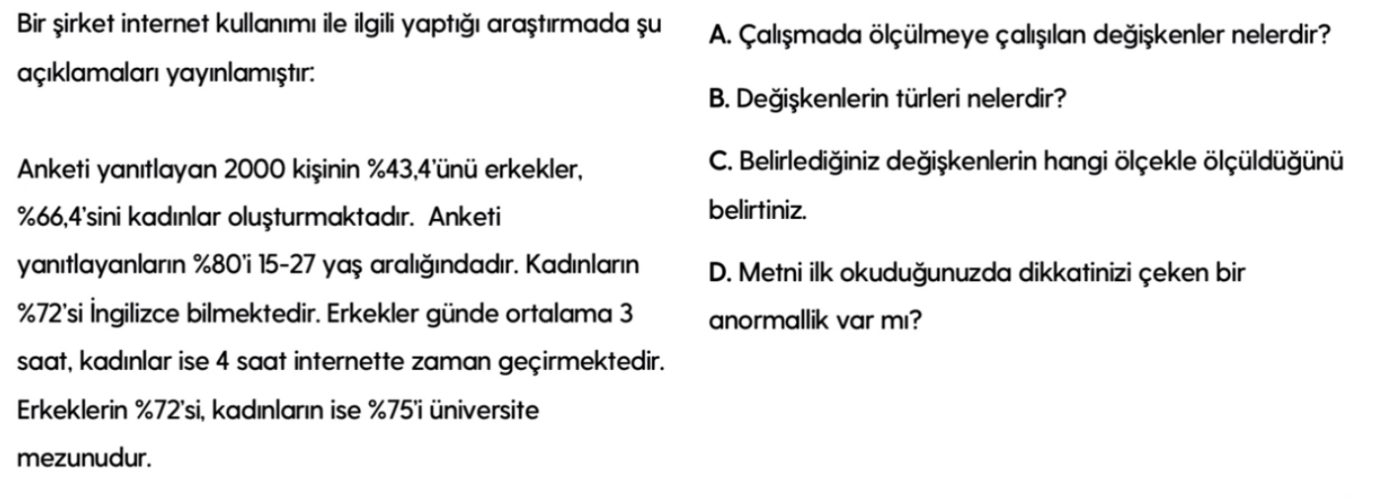
**Örnek:**





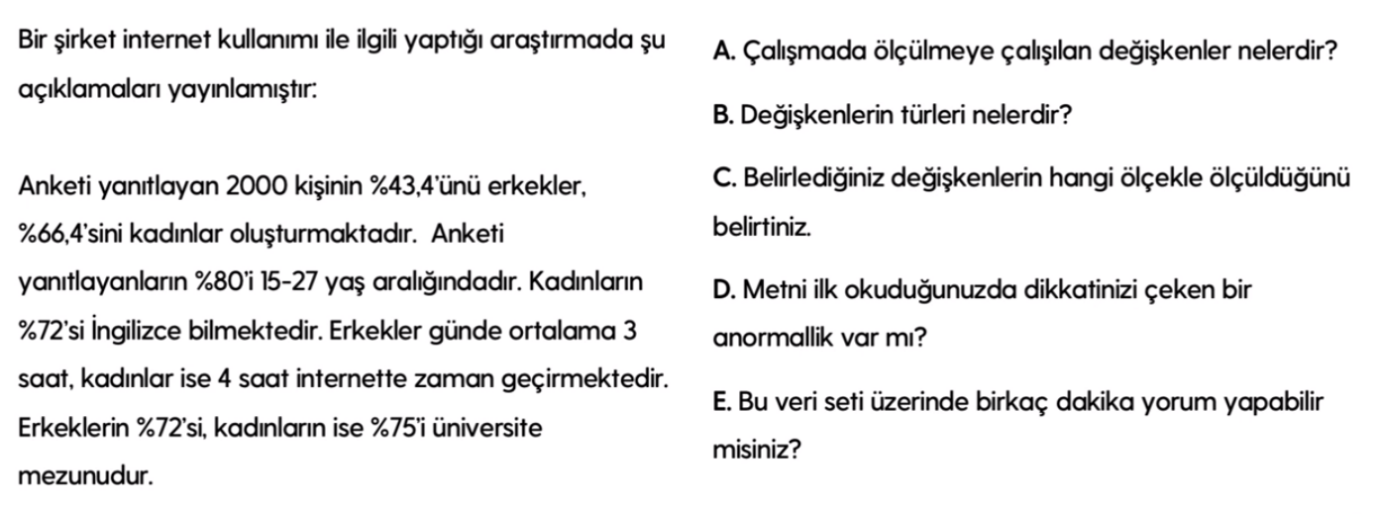






Aslında bakarsanız istatistiksel düşünce modellerinin analitik düşünceyi geliştirme yönelik olarak geliştirilen modellerin hedeflediği kişinin dikkatini ölçmeye yönelik bir soru tarzıdır. Dikkatinimizi çekmesi gereken:

* 43,4 erkek ve 66,4 kadın toplandığında yüzdelikler toplamı 100 olmuyor.
* Dil konusunda bir eksik anlatım tarzı var. İngilizceye kıyasla bir dil var mı ya da başka diller arasında mı ingilizce yüzdeliği verilmiş gibi bu soruların cevabı yok.



Bu veri seti üzerinde birkaç dakika yorum yapabilir misiniz ?

Bilimsel anlamda eğer verilerde rastgeleliğe sahip ve yansız ise 2000/2500 Türkiye için tahmin konusunda ideal bir örneklem sayısıdır.

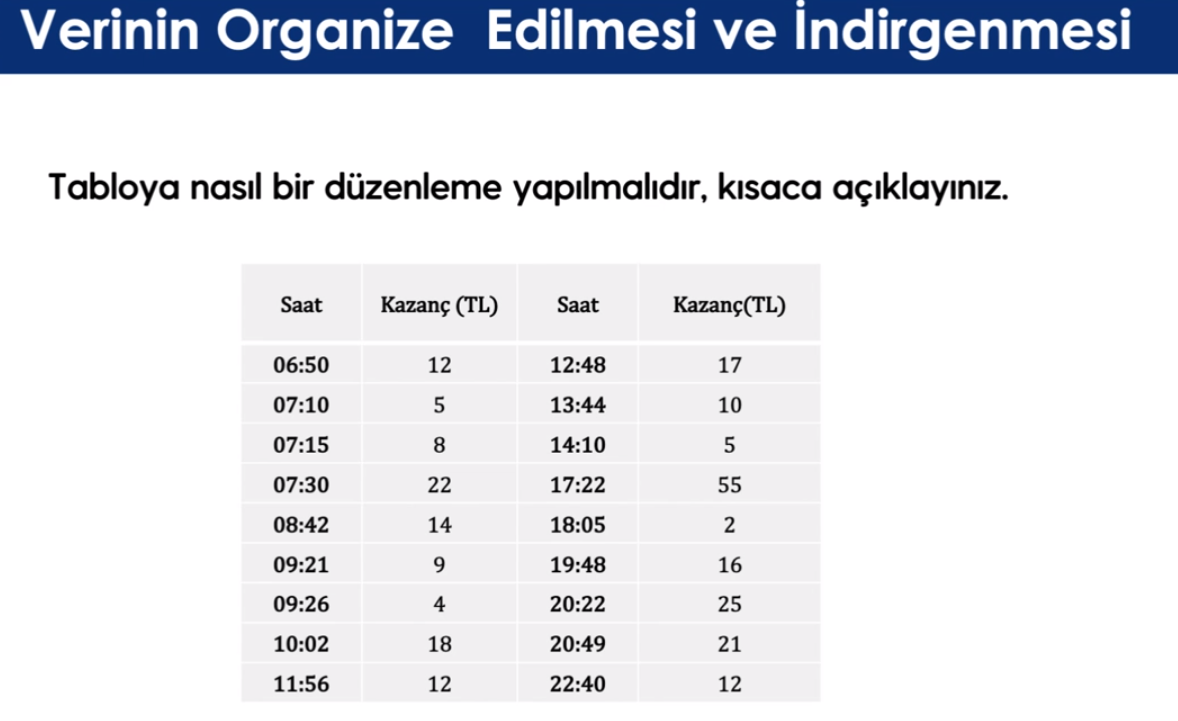
Kadın erkek oranları Türkiye için tam uygun değil gibi çünkü Türkiye'de kadın erkek oranı %50 civarında. Netice olarak düzgün bir oran.

Bu şirketin odak kitlesi 15-25 yaş arasında yani genç nüfus ile ilgili genelleme yapılır.

Kadınların %72 i İngilizce bilmektedir. Bize verilen her bilgi doğru olmaya bilir. Burada diğer kalan kısım ise ne bilmektedir sorunun cevabı yok.

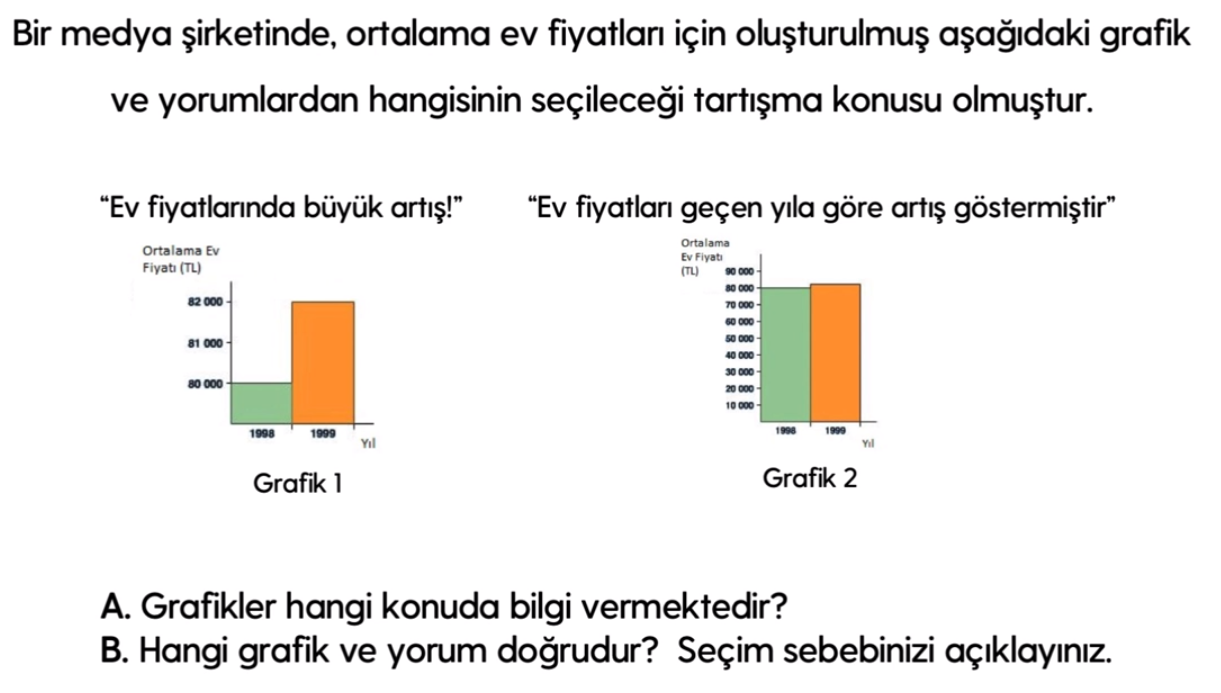
Bu değerlerin üzerine diğer yıllardaki analizi de karşılaştırarak doğrulanabilir.

**2-Verinin Organize Edilmesi ve İndirgenmesi**





Böylece iş saatlerini baz alarak bazı yorumlar yapabiliriz. Eğer ki bakkal iş yerlerine yakın olsaydı saat 12:00-15:00 arasında daha fazla kazaç elde edebilirdi. Ancak Bakkal iş yerlerine yakın gözükmüyor. Çünkü o saatlerdeki kazancı az.



1. İçin:

Eksenlerde ortalama ev fiyatı ve yılı mevcut. Grafikler Ev fiyatlarındaki artışı vermektedir.

1. İçin:

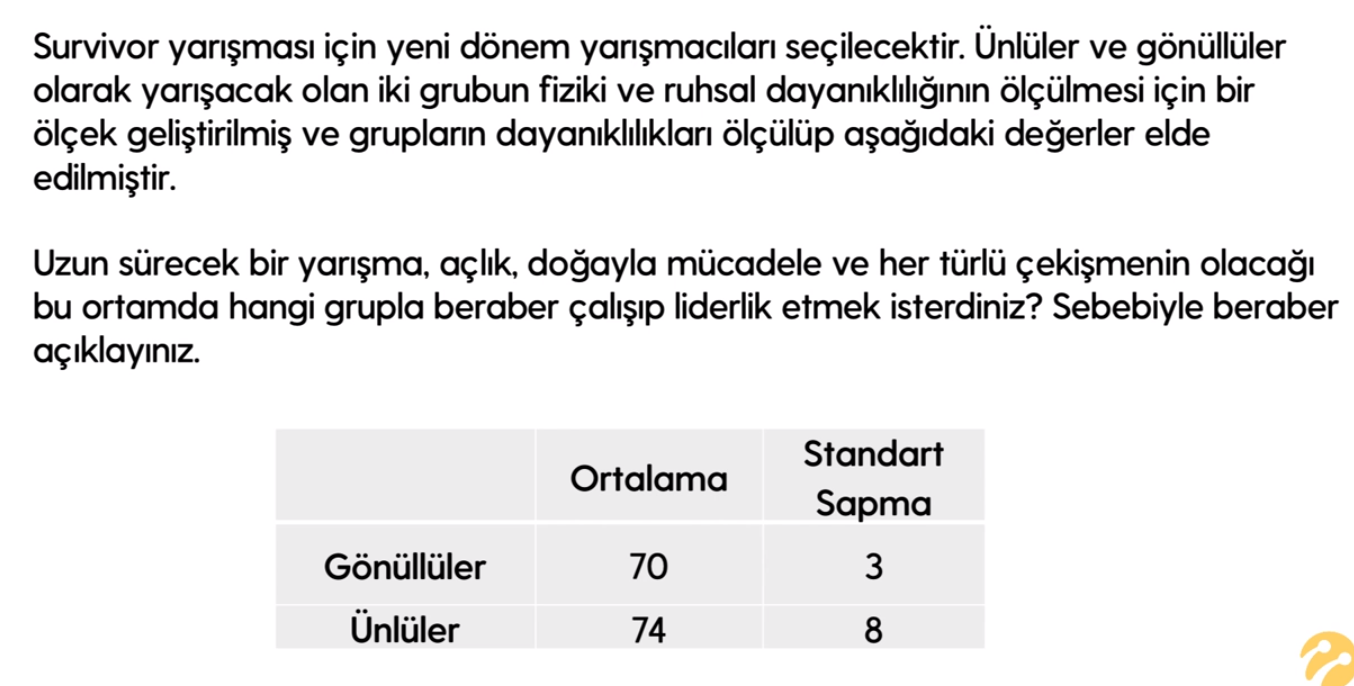
Grafik 1. Medya şirketi için daha uygundur ve daha ilgi çekicidir ve tıklaması konusunda yüksek sayılara ulaşılabilir.

Ayrıca:

Grafik 1 de eksen 0 dan başlamamıştır. Eksenlerin başlangıç noktaları 0 dan başlanmalıdır. Bu suiistimal edilmiş. Buna göre sağdaki(Grafik 2) daha doğru bir yaklaşım.

Soldaki grafikte ev fiyatları 1000 er artıyorken sağdaki grafikte 10000 artıyor. Bu sebeple 81000’e çıkan fiyat Grafik 1 daha çok daha yüksekmiş gibi gösterilmektedir. Ev fiyatlarında 1000 TL lik bir artış genel olarak pek te doğru bir yaklaşım olmaz.

**3-Verinin Analiz Edilmesi ve Yorumlanması**



Bu durumda normalde ortalamanın yüksek olduğu Ünlüler seçilirdi. Burada basit bir dağılımdan daha fazla bilgi barındırmaktadır.

Genel ortalaması diğer gruba göre yüksek olanla mı çalışmak isterdiniz yoksa grup içi davranış biçimleri birbirine yakın grupla mı çalışmak isterdiniz.

**NOT: Standart sapma küçükse veriler ortalamaya yakın olarak dağılım gösterir. Büyükse ortalamadan uzakta dağılım gösterirler.**

Gönüllülerde sapma 3 olduğu için bireyler ortalama 67 73 arasında değerlere sahiptir. Bunun anlamı gönüllüler fiziksel ve ruhsal dayanılılık ölçümleri olarak birbirine çok yakın yani homojon dağıldığı için ve bu yarışmada takım çalışması önemli olduğu için Gönüllülerde olmak daha uygudur.

Diğer yandan Ünlülerin ölçümleri 66 ile 82 arasında değişmektedir. Bu da takımda hem çok iyi hem de çok kötü yarışmacıların olduğu aynı zamanda fiziksel ve ruhsal dayanıklılık ölçümleri olarak birbirine yakın olmayan farklılıkların olduğu yorumlanabilir. Bu da her ne kadar iyi yarışmacılar olsa da, birkaç yarışmayı iyi götürseler de bu yarışmada takım çalışması önemlidir ve bu durumda ünlüler seçilmemelidir.