

BETİMLEYİCİ İSTATİSTİĞE GİRİŞ

Niceliksel ölçmede, dört temel ölçü biçimi vardır. Ölçünün türü temsil ettiği kavram veya yapıyı yansıtır. Bazı ölçüler oldukça yapay veya ham ayrımlar yapar; bazıları daha yüksek ve daha ayrıntılandırılmış düzeyde. Ölçüler başlangıçta ikiye ayrılır (Tablo 4.1). Sürekli ve parçalı (İngilizce continuous ve discrete). Parçalı ölçüler isimsel veya kademelidirler. Sürekli olanlarsa matematiksel bir devamlılıkta nerede oldukları saptanabilen değerlerdir ve ölçü derece derece artan biçimdedir. Oransal ve mesafeli sayılar, sürekli türe örnektir. Demek ki, zamanın ölçüsü, saniyeler veya dakikalar olabilir; mesafe ölçüsü uzunluğun farklı birimleridir. Öte yandan sıçramalı olanlar görece olarak sabit özelliklere göredirler.

Tablo: Değişken (Ölçü) Türleri

Parametrik Olmayan		Parametrik	
Sıçramalı		Devamlı	
İsimsel	Kademeli (rütbeli)	Mesafeli	Oransal
Erkek=0 veya ♀ Kadın=1 veya ♂	K e s i n l i k l e Katılmıyorum=1 Katılmıyorum=2; Bilmiyorum=3 Katılıyorum=4 K e s i n l i k l e Katılıyorum=5	25 Santigrat; 50 santigrat (ısı iki katı değildir). 50 IQ; 100 IQ'nun iki katı değildir.	24 Yaş 50 metre 500 TL 120 Kg.

Ancak sürekli ölçülerden mesafeli ölçünün anlaşılması diğerlerine oranla daha zordur. Burada kastedilen mesafeler arasında ölçeğin eşit olmasıdır. Yani bir ısıölçer öğlen 24 derece Santigratı gösteriyorsa ve akşam aynı ısıölçer 12 Santigratı gösteriyorsa, sıcaklık yarı yarıya azalmış diyemiyoruz. Çünkü, ısının gerçek anlamda 0 noktası yoktur. Yani ısısızlık durumu söz konusu değildir. Bu nedenle 0 noktası “keyfidir”. Zeka testlerinde de aynı durum söz konusudur. Bir kişi zeka testinden 50 IQ almış, diğeri 100 almışsa, bu ikincinin iki kat daha zeki olduğu anlamına gelmez. Çünkü sıfır noktası burada da keyfi konulmuştur.

Ölçüler arasında en basiti isimsel olandır ve az sayıda matematiksel işleme tabi tutulabilir. En gelişkin matematiksel işleme sokulabilenler oransal olanlardır. Farklı ölçüler üzerinde durmamızın nedeni, istatistiki işlemlerde her tür ölçüye uygulanabilecek işlemler vardır. Farklı işlemler uygulandığında hata yapılmış olur.

Araştırmalarda kullanılan iki tip istatistik vardır. Bunlardan biri betimleyici istatistik, diğeri çıkarımsal istatistiktir. Betimleyici istatistik araştırma sonucu ortaya çıkan ham verilerin genel görünümünü veya özetini topluca görmemizi sağlar. Gündelik araştırmalar da göz önüne alınırsa, en çok kullanılan istatistik türünün bu olduğunu söyleyebiliriz. Çıkarımsal istatistikse, eldeki malzemeden, belli koşullar yerine geldikçe dış dünyaya genelleme yapmamıza izin veren istatistik biçimidir. Bazı istatistikçiler bu ayrımı parametrik istatistik, parametrik olmayan istatistik olarak da sınıflandırır. Parametrik istatistikler mesafeli veya oransal değişkenler için, parametrik olmayan istatistik isimsel ve kademeli veriler için kullanılır. Parametrik istatistiğin sonuçları nüfusun bütününe genellenebilir, oysa parametrik olmayan istatistiğin sonuçları genellenemez.

1. TEK DEĞİŞKENLİ ÇÖZÜMLEME

Tek deęişkenli çözümlenmede, tek bir deęişkenin birimleri incelenir. Bu deęişken yaş, sınavda alınan notlar, belli bir kümenin ağırlığı, geliri, kişilerin boylarının uzunluğu olabilir. Tek deęişkenli çözümlenmelerde şunlara bakılır:

- Dağılım
- Merkezi Eğilim
- Saçınıklık

Şekil 4.1’de 94 öğrencinin sınavdan aldıkları notların dağılımı, not kategorilerine göre sınıflanarak sunulmuştur. Not kategorilerinin oluşturulmasının nedeni, her bir not için bir sütun ayrıldığında ortaya çıkan karmaşıklık gidermek içindir. Bu şekilde bakarak sınıfta kalanlar ve geçenler ayrımı rahatlıkla yapılabileceği gibi, notların yoğunlaştığı kategoriler de ortaya çıkar.

4.1.1. Dağılım

Dağılım, her bir birimin (vakanın) veya birim kümelerinin yinelenmelerinin (frekanslarının) özetini verir. Her not veya not kategorisi bir birimdir. Her not grubuna giren birimin kaç kere yinelendiği dağılımı ortaya çıkarır. Benzer şekilde, belli bir gelir dilimi veya belli bir ağırlık ölçüsü de olabilirdi. Kimi dağılım, yüzdeler şeklinde verilir. Yukarıdaki şekildeki tüm sayı 94 olduğu için bunun 100 üzerinden oranları da benzer bir şekil oluşturacaktır. Dağılım yüzde olarak verilirse, buna *yüzde dağılımı* denir. Eğer birim sayımız artarsa kullanmamız gereken dağılım yüzde dağılımıdır. Örneğin 94 öğrenci yerine 3,563 öğrencinin notunun dağılımını almak isteseydik, yüzde dağılımını kullanmamız gerekir.

4.1.2. Merkezi Eğilim

Bir dağılımın merkezi eğilimi, değerlerin dağılımının “merkezinin” ölçümüdür. Merkezi eğilimin şu üç ana ölçüm türü bulunmaktadır:

- Ortalama (İngilizce mean)
- Ortanca (İngilizce median)
- Tepedeğer (İngilizce Mode, bazıları Türkçe’de “tarz” veya “mod” der)

4.1.3. Ortalama

Aritmetik ortalama olarak da tanımlanan ortalama, merkezi eğilimi tanımlamak için en çok kullanılan yöntemdir. Hesaplamak için yapmanız gereken bütün değerleri toplayıp birim sayısına bölmektir. Örneğin yukarıdaki öğrencilerin ortalaması tüm not değerlerinin toplamının öğrenci sayısı olan 94’e bölünmesiyle bulunur. Varsayalım ki, 13 öğrencinin girdiği sınavda aşağıdaki notlar alınmıştır. Bunun ortalaması bütün değerlerin toplamının (915) 13’e bölünmesiyle bulunur. Öğrencilerin ortalaması 70.3’tür.

Ortalamanın en sakıncalı yanı uç noktalardaki sayılardan etkilenmesidir. Bu uç noktalardaki sayılar, ortalamayı yukarı veya aşağı doğru çeker. Oysa büyük bir çoğunluk daha ortada yer almaktadır. Aşağıdaki kutudaki gibi 10 kişiden dokuzunun aylık geliri 100 milyon Türk Lirası, sonuncunun aylık geliri 2 Milyar TL ise ortalama gelir 290 milyon olarak çıkar. Bu oldukça saptırıcı bir özetdir. Ortalamanın bu sakıncası

özellikle birim sayısının 30'un altında olduğu durumlarda kendini belli eder dense de, birimlerin fazlalığı ortalamanın anlamlılığını garantilemez.

100 100 100 100 100 100 100 100 100 100 2,000

Ortalama= 290 (Milyon TL)

4.1.4. Ortanca

Ortancayı bulmak için bütün rakamlar en düşükten en yükseğe kadar sıralanır. Aşağıdaki örnekte 13 öğrencinin notlarının en azdan en çoğa sıralanmış durumunu görüyorsunuz. Bu sıralamanın ortancası en yüksekten altı, en çok tarafından altı rakam çıkartıldığında en ortada kalan 78'dir. Eğer öğrenci sayısı 14 gibi çift basamaklı olsaydı ortanca nasıl saptanırdı? Örneğin 14 öğrencinin olduğunu varsayarsak, hem en azdan hem de en sondan başlayarak ortada iki rakam kalana kadar atarız. Ortada kalan iki rakamın ortalamasını alırız (iki sayıyı toplar ikiye böleriz). Böylece ortada kalan iki rakam tek rakama indirgenmiş olur.

Ortancanın en önemli özelliği (ortalamanın tersine) uç noktadaki aşırılıklardan etkilenmemesidir. Örneğin yukarıda gelire ilgili 10 kişiyi yeniden ele alalım. Bu rakamları en azdan en çoğa sıralayıp ortancasını bulduğumuzda 100 Milyon TL çıkacaktır. Ortanca, en sağdaki 2 Milyarı dışarda bırakmaktadır.

32 32 32 63 65 70 78 78 85 91 94 95 100

↑

ORTANCA (ME) =78

Dikkat edilirse bu örnekte, çift basamaklı bir sayı (10) söz konusudur. Böyle bir durumda ortada kalan iki rakamın ortalaması alınır. Ancak yukarıdaki örnekte buna gerek yoktur çünkü ortada kalan her iki birim de 100'dür. Zaten ortalamasını alsak da aynı sayıyı yeniden buluruz (100+100/2). Bu örnekte de görüldüğü gibi özellikle birim sayısı azaldıkça, ortalama yerine ortanca çok daha anlamlı sonuçlar vermektedir.

4.1.5. Tepedeğer

Tepedeğer (Tarz/Mode): Adından da anlaşılacağı gibi, çoğunluk birimlerde en fazla yinelenen sayıyı temel alır. Sınav notları belli olan 13 öğrenciyi yeniden ele alalım.

32 32 32 63 65 70 78 78 85 91 94 95 100

↑

3 ADET

↑

2 ADET

TEPEDEĞER (MO) =32

Bu sayılarda 32 ve 78 dışında yinelenen sayı yoktur. Birincisinden üç, ikincisinden iki birim sayılmaktadır. Bu nedenle çoğunluk değeri 32'dir. Gelire ilgili örnekte de 100 Milyon TL'den dokuz birim, 2 Milyar

TL'den bir birim vardır. Dokuz kez yinelenen 100 Milyon TL çoğunluk değeridir. Eğer iki farklı değer birbirine eşit sayıda yinelenseydi hangisi tepedeğer olarak seçilecekti? Diyelimki 32'den iki birim, 78'den de iki birim. Hangisi tepedeğerdir? Yanıt "Her ikisi de" olacaktır. Böyle çoğunluk değerlerine iki tepedeğerli (İngilizce bimodal) denir.

Tablo : Ölçü Türlerine Göre Geçerli İstatistiki İşlemler

Merkezi Eğilim Ölçüsü	Ölçek Türü			
	İsimsel	Kademeli	Mesafe li	Oransal
Ortalama	Hayır	Hayır	Evet	Evet
Ortanca	Hayır	Evet	Evet	Evet
Tepedeğer	Evet	Evet	Evet	Evet

Öte yandan örneklerde ele aldığımız sınav notları ve gelir kümelerinin her birinin birimlerinin (vakalarının) düşüklüğü ortalama ölçüsünün de anlamlılığını düşürmektedir. Bununla birlikte son bir not daha iletmemiz gerekir. Eğer bir kümenin dağılımı "normal dağılımsa", ortalama, ortanca ve çoğunluk (tepedeğer) aynı sonucu verir. Dolayısıyla mükemmel bir normal dağılım durumunda merkezi ölçümlerin her biri aynı sonucu verir. Eğer bir eğrinin normal dağılım olduğunu biliyorsanız, her üç ölçüm de kullanılabilir demektir. Bununla birlikte normal dağılım çok yüksek sayıda birim (vaka) kullanılmasını gerektirir. Özellikle ölçülen birim sayısı 200'ün altına doğru düştükçe, normal dağılımın elde edilmesi oldukça güçtür.

4.2. İKİ DEĞİŞKENLİ ÇÖZÜMLEME

4.2.1. Bağntı (Korelasyon)

Tarihsel gelişimi ve uyarılar

Bağntı kavramını ilk kullanan çocukların boylarıyla babaların boyları arasındaki ilişkiyi bulmaya çalışan 19. YY matematikcisi Galton'dur (Ellenberg:311-312). Galton'a göre aritmetik ortalamaya yaklaşma (regresyon), üzerinde çalışılmakta olan olgunun şans eseri etkileyen güçlerin bulunması durumunda olmaktadır. Acaba bu güçler, babalardan devranılan genetik mirasa ne kadar etkide bulunuyorlardı? Yanıtlamak istediği soru budur. Kağıt ve kalemle çalışan Galton, yaptığı denemeler sonucunda dağılım grafiğiri bulan kişi olmuştur. Eğer babanın ve çocuğun boyunu aktaran noktalar neredeyse bir daire gibi olursa, bütünüyle şans güçleri devrede demektir. Ama daire yavaş yavaş elipse dönüyorsa durumun değerlendirilmesi gerektiği ortaya çıkacaktı. Mükemmel bir ilişki, elipsin neredeyse çizgiye dönüştüğü görünüm olurdu.

Galton bu ölçünün adını bağntı (İng. Correlation) olarak tanımladı. Galton'un elipsi neredeyse daire gibiyse bağntı 0'a yakındı. Elips zayıflayıp bir çizgiyi andırdığında bağntı 1 oluyordu. Bağntının artı olması durumu bir değişkendeki artışın diğerinde de artışa yol açtığını gösterir. Eksi olması durumuysa bir

değişken düşerken diğeri artarsa (örneğin gelir düştükçe, bir siyasi partiye verilen oyların artması) ortaya çıkar.

Aslında Galton'dan önce, Paris Polisi için çalışan Bertillon suçluların yakalanması için bir sistem geliştirmişti. 1883'te Paris Polis Müdürlüğü'nün sistemi kabul etmesinden sonraki yıllarda tüm dünyada yayılan bu sisteme göre, suçluların çeşitli fiziksel özellikleri bir araya getiriliyor ve hangi fiziksel özelliklere sahip olanların suç işlemiş olduklarına dayanarak, şüpheliler arasında elemeye gidilebiliyordu. Bunu bilten Galton, aslında onlarca fiziksel özellik yerine sadece bir tek fiziksel ölçünün suçlulukla ilişkisinin olup olmadığını merak ediyordu. Kendi kullandığı boy verisinde, babanın boyu yerine, ölçülen kişinin orta parmağından dirseğine olan uzunluk ile boy uzunluğu arasında bir bağıntı olup olmadığına baktı. Evet vardı!

Galton'un geliştirdiği bağıntı kavramıyla, herhangi bir değişkenin diğerinin nedeni olmamasına karşın, iki değişkenin bir arada değişme ilişkisi olduğuna ilişkin sonuçlara varmak mümkün oluyordu. Aslında Galton, kendi sorunlarını çözmek için babanın uzunluğu/çocuğun uzunluğu veya ortaparmaktan dirseğe kadar uzunluğu/boy uzunluğu değişkenleri dışında çok fazla değişkenle çalışmamıştı. Bunu herhangi iki veya daha fazla değişkenle yapılması yöntemini genelleştiren Karl Pearson oldu.

Bağıntıyla ilgil olarak da çeşitli problem bulunur. Örneğin HDL olarak bilinen iyi kolesterolün yüksek olmasının on yıllardır kalp-damar sağlığındaki riskleri azalttığı bilinir (Ellenberg:532). Ayrıca bilinir ki bazı ilaçlar HDL düzeylerini yükseltir ki bunlardan biri niacin (B vitamini biçimi) Eğer HDL kalp-damar hastalıklarındaki ölüm riskini azaltıyorsa niacin almak akıllıca olur diye düşünülür. Ama bunun işleyip işlemeyeceği açık değildir. Niacin eklerini almak küçük tıbbi araştırmalarda oldukça iyimser sonuçlar verdiler. Ama büyük düzeyde yapılan ve devlet tarafından gerçekleştirilen deneyler 2011 yılında, sonuçların çok zayıf olması nedeniyle devam ettirilmedi. Hastalar niacin hapını daha fazla aldılar ama herkes kadar kalp krizi geçirdiler. Bu nasıl olur? Bunun nedeni bağıntının geçişli olmamasıdır. Niacin yüksek düzeyde HDL ile bağıntılıdır ve yüksek HDL daha az kalp krizi demektir ama bu, Niacin'in kalp krizini önleyeceği anlamına gelmemektedir.

Bir başka bilinmesi gereken unsur bağıntının olmaması değişkenlerin ilişkili olmasını engellemez (Ellenberg:344-345). İki değişken arasında bağıntı varsa, bu iki değişken bir biçimde bir biriyle ilişkilidir. Eğer bağıntı yoksa? Değişkenlerin bütünüyle ilişkisiz olduğunu söyleyebilir miyiz? Galton'dan başlayarak bağıntı doğrusal (lineer) olarak anlaşılmıştır. Her iki değişken bir arada artar (pozitif bağıntı) veya iki değişkenden biri düşerken diğeri artar (netagif bağıntı). Ama hayatta her eğri düzçizgi değildir ve her ilişki doğrusal değildir.

Eğer evli olmayanlar arasında sigara içme oranı, evli olanlara kıyasla daha yüksektir türünden bir bağıntı bulundu diyelim. Böylesi bir durumda tüm toplumda sigara içenlerin oranının, evli olmayıp da sigara içenlerin oranıyla tıpa tıp aynı oranda olması şansı oldukça düşüktür. Böylesi bir küçük ihtimal olmadıkça, evlilik ve sigara alışkanlığı her durumda bağıntı içindedir! Ya pozitif ya da negatif. Sonuçta her şey her şeyle bağıntı içinde olacaktır demeniz mümkündür. Bunun denmesinin bir anlamı olmayacağı için bağıntının ancak "güçlü olduğu" durumlarda raporlanması gerekir.

Aslında cümlemizdeki basit deęişiklikler önemlidir. Őu cümleleri bakalım:

- Eęer sigara içiyorsanız, muhtemelen evli deęilsinizdir
- Eęer sigara içiyorsanız, muhtemelen evlenmeyeceksiniz.

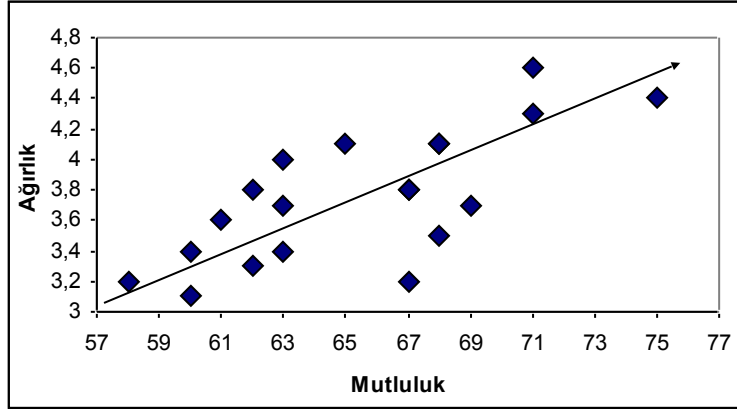
Yukarıdaki cümlelerden birincisi bir baęıntıya işaret eder. Ancak ikinci cümle baęıntıya deęil, neden-sonuç ilişkisine işaret etmektedir. Birinci cümle arařtırmamızın baęıntı sonucunu vermesi nedeniyle kurulabilir. Sigara içenlerin, dięerlerine daha az evlilik yapıyor olması, sigarayı bıraktınca karřımıza bir eř çıkaracak olması anlamına gelmez. Baęıntı ve neden-sonuç anlayıřımız oldukça kaygan bir zemine çeker bizleri. Sezgisel olarak bazılarının olabileceęini bazılarının olamayacaęını pek çok durumda anlayabiliriz. Ama bazı durumlarda bunu sezgisel olarak anlamamız mümkün olmaz. Örneęin iyi kolesterol, daha az kalp krizi riskiyle baęıntı içindedir dendięinde olgusal bir ifadede bulunuyoruz: Eęer yüksek düzeyde iyi kolesterol varsa, kalp krizi geçirme riskiniz daha düşüktür. Ne yazık ki, iyi kolesterolü arttıran hapi kullanmanız durumunda, kalp krizi geçirmeyeceęiniz garantisi bulunmaz. Bunun olması için, baęıntıdan bir adım öteye neden-sonuç ilişkisinin bulunmuř olması gerekir. Çünkü iyi kolesterol dıřında invan bedeninin karmařık yapısının sonucu olarak başka bir etki nedeniyle de kalp kirizi geçirebilirsiniz.

Baęıntı nasıl hesaplanır?

Özetlersek, istatistikte en çok kullanılan yöntemlerden biri olan baęıntı, iki deęişken arasında ilişkiyi tanımlayan bir sayıdır. Nasıl hesaplandığını ve nasıl kullanıldığını bir örnek aracılıęıyla verelim. Diyelim ki aęırlıkla (Kilogram) mutluluk arasındaki ilişki var, veya aęırlık arttıkça mutluluęun azaldığı yönünde bir denenceniz var (bařlangıç olarak nedensellięin yönü üzerinde fazla düşünmeyelim. Yani mutluluęun aęırlığı arttırmadığı ortada gibi... Bakacaęımız beden aęırlığı arttıkça, mutluluęun da artıp artmadığı). Toplam 20 kişilik bir küme üzerinde veri topladık. Bu kişilerin aęırlıklarını ölçtük ve kaydettik. Sonra da bir anket uygulayıp mutlu olup olmadıklarına ilişkin 1 ile 5 arasında bir sayı elde ettik.

Karıřtırılmaması için iki şeyi birbirinden ayıralım. *Yanıt ölçeęi* ile tutum/psikolojik durum gibi *test ölçeęi* birbirinden farklıdır. Yanıt ölçeęinin örneęini bu bölümün bařında ele almıřtık. Bu kişilere yanıt ölçeęi sormadık. Mutlu musunuz? Birini seçiniz: Çok mutsuzum (1); mutsuzum (2); ne mutluyum ne mutsuzum (3); mutluyum (4); çok mutluyum (5). Bunun yerine bir dizi soru kullandık, bunlardan her biri için yanıt ölçeęi (1-5 arasında) kullanmış olabiliriz. Soruların her birinin mutluluk tanımlaması açısından bir deęeri var. Bu deęerlerin tamamından bir mutluluk ölçeęi yarattık. Tıpkı zeka ölçeęi (IQ Testi) gibi. Bunun üzerinde neden duruyoruz? Çünkü yanıt ölçekleri kademeli/rütbesel ölçü olarak kabul edilir. Mutluluk endeksi gibi 1-5 arasında oluřturulan endeks ise mesafeli ölçektir.

Bu verilerin daęılımını gösteren şekillere bakalım ve inceleyelim. Birinci şekilde kişilerin aęırlıkları ve her aęırlığın yinelenme sıklıkları görülmektedir. Aęırlıkların yer aldığı eksen "X" eksen; yinelenme sıklığının yer aldığı eksen "Y" eksenidir.

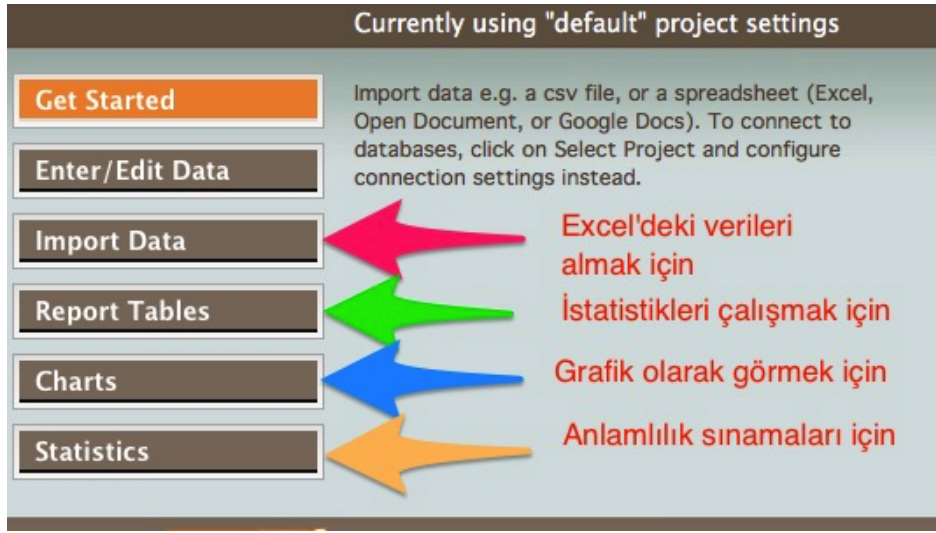


Son olarak, iki değişkenli dağılım grafiği yaratarak durumu inceleyelim. Şekle bakıldığında iki değişken arasındaki ilişkinin olumlu (pozitif) olduğunu görmek mümkündür. Eğer noktalar arasında tek bir doğru çizgisi çekersek, oklu çizgiyi elde ederdik. Buradaki eğim olumlu eğimdir. O çizginin üzerindeki herhangi bir noktayı aldığımızda ağırlık arttıkça, mutluluk endeksinin yükseldiğini görecektik. Örneğimizde bu duruma oldukça yakındır. Bağntı, ilişkinin varlığını ortaya koyan sayısal bir veri olduğu için, olumlu bağntı beklentimiz vardır.

Bağntının nasıl hesaplandığı ancak istatistikçi olmak isteyenler açısından önem taşır. Bağntının uluslararası düzeyde istatistiki simgesi “r”dir. Matematikçilerin uygulamaları “sayesinde” r, -1,0 ve +1,0 arasında değişen bir sayıdır. Bağntı olumsuzsa, eksi işaret; bağntı olumluysa + işaret ortaya çıkar. Şu ana kadar ele aldığımız kümenin kişi sayılarından söz ettik. Bu sayıyı belirtmenin yönteminin de “N” harfi olduğunu hatırlayınız. Eğer 20 kişilik bir küme sözkonusuysa N=20 yazılır. Örneğimizdeki küme için bağntıyı bilgisayara hesaplattığımızda, +0,73 çıkmaktadır. Bu oldukça güçlü bir olumlu bağntı olduğu anlamına gelir. Eğer +1,0 çıksaydı ne anlama gelirdi? Yukarıdaki şekildeki oklu çizginin aynısının olduğu ortaya çıkardı.

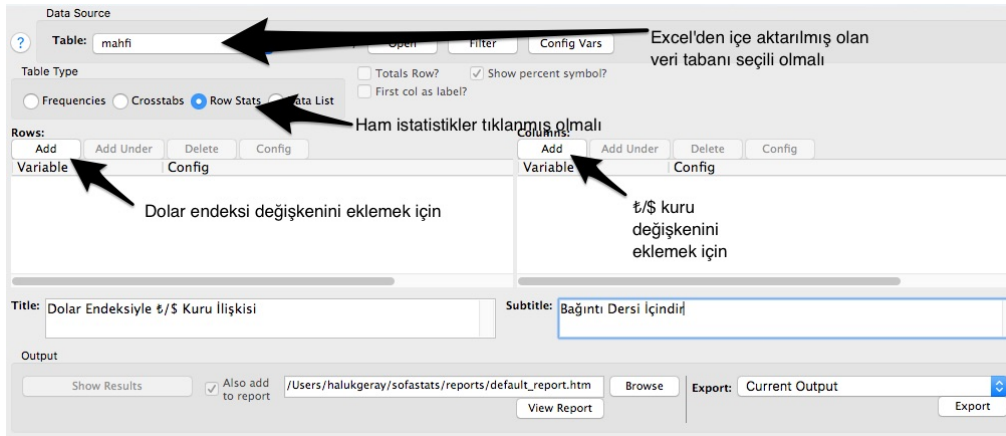
Örneğimizde mutluluk endeksi mesafeli bir ölçektir. Çünkü, sorular takımından yaratıldığı sayılandır bu ölçekte 1,3 ile 1,5 arasındaki fark 0,2 olduğu gibi; 2,3 ile 2,5 arasındaki fark da 0,2’dir. Bunu vurgulamamızın nedeni “yanıtlama ölçeği” (hiç katılmıyorum, kesinlikle katılıyorum gibi) ile “tutum ölçeğinin” birbirinden farklı olmasıdır. Eğer tek bir soruyla mutluluğu ölçmek isteseydik ve 1 ile 5 arasında (kesinlikle mutsuzum, mutsuzum, kararsızım, mutluyum, kesinlikle mutluyum) yanıt ölçeği kullansaydık, ortaya çıkan değer kademeli (ordinal) olurdu. Bu durumda Pearson tekniğini kullanmazdık. Çünkü bu teknik mesafeli (interval) değer söz konusuysa kullanılır. Eğer iki kademeli (ordinal) değerler söz konusuysa ‘Spearman Rank Order Correlation (rho)’ veya ‘Kendall Rank Order Correlation (tau)’ kullanılır. Bir değişken mesafeli diğeri isimsel (iki kategorili, kadın/erkek gibi) ise ‘Point-Biserial’ yöntemi kullanılır.

Bir başka örnekle devam edelim ve bu kez SOFA Stats isimli özgür, açık, kaynak kodlu (ÖAKK) yazılımı kullanalım. ÖAKK yazılımlar bilindiği gibi serbestçe indirilir ve parasızdır. Yukarıdaki SPSS programı günümüzde oldukça pahalı lisanslarla satılıyor. Konumuz T24 dijital gazetesinde bir yazarın yaptığı yorum. Bu yorum 2016 Güzünden başlayarak ABD Dolarının Türk Lirası karşısındaki yüksek değer kaybını açıklamayı hedefliyor (Eğilmez, 2017).



Yazar, 2016 yılında şubat-aralık ayları arasında ABD Dolar Endeksi (birinci değişken) ile ABD Dolar/Türk Lirası kuru (2. değişken) arasındaki ilişki olup olmadığını bağıntı tekniğiyle anlatmayı hedefliyor. Bunu bizim de yapabilmemiz için kendisinin tabloda verdiği verileri hesap tablosu yazılımına yüklememiz gerekiyor. Çünkü bu tabloyu SOFA yazılımına aktaracağız. Bunu yapmak için SOFA yazılımının arayüzündeki “aktar” (import) işlevini kullanmalıyız. Diğer işlevlerin kullanımı da şöyledir:

- Report tables: İstatistik analizler yapmak için kullanılan temel araçtır. Burada istenen türde tabloları içeren istatistik raporları oluşturulabilir.
- Charts: Bu araç, istatistik analize ilişkin grafikler oluşturmak ve onlar aracılığıyla inceleme yapmak içindir.
- Statistics: Bu araç anlamlılık sınamalarının gerçekleştirilmesini sağlayan araçtır.



Aktarma aracını kullanarak, hazırladığımız tabloyu SOFA'ya aktarırız. Report tables aracında bu dosya yani aktardığımız veri tabanı seçilmiş olmalıdır. Daha sonra Dolar Endeksi değişkenini (değişkenlerden biri birinci değişken olarak diğeri ikinci değişken olarak seçilir hangisinin bir hangisinin iki olduğunun önemi yoktur) ve Dolar/Türk Lirası Kur verilerini “ekleriz” (add komutunu tıklayınca çıkan kutudan tıklayarak seçeriz). Sonra aynı işlemi kur değişkeni için yaparız. Burada atlanmaması gereken nokta, “ham istatistikler (raw statistics) seçeneğinin tıklanmış olmasıdır. Bu seçeneğin dışında frekans (sıklık) veri listesi ve çaprazlama seçenekleri de kullanılabilir. Ancak biz bağıntıya bakmak istediğimiz için bunu seçiyoruz.

Ayrıca aşağıdaki kutucuğa başlık ve onun sağındaki kutucuğa alt başlık yazıyoruz. Daha sonra yapmanız gereken en aşağıda bulunan “view report” ikonunu tıklayarak raporunuzu görmeyizdir.

Bu işlemi yaptığımız zaman bağıntı değerinin (r değeri) -0,258 olduğunu görüyoruz. Biliyorsunuz, R değeri 1 olduğunda çok güçlü ilişki var demektir. Sıfıra yaklaştıkça bağıntı değeri düşer. Yazarımız da buradan hareketle bu bağıntının zayıf olduğunu ve ancak yüzde 25 etki yaptığını söylüyor. Ancak yazar, P-değerine bakmamış veya bize bilgisini aktarmıyor. O halde biz yapalım ve görelim. Bunu yapabilmek için SOFA arayüzünde “statistics” ikonunun tıklanması gerekiyor. Bu ikon tıklanınca, yazılım sizin için otomatik olarak hangi anlamlılık sınaması algoritmasının uygulanacağını seçer. Bunu yapabilmesi için dosyanızı SOFA'ya aktarırken değişkenlerinizin hangi ölçü türü olduğunu menüden tıklayarak seçmiş olmanız gerekir. Ancak eğitim amaçlı olarak, ikinci seçeneği tıklıyoruz. Burada karşımıza çıkan seçeneklerden birincisi 2 değişken arasındaki farklılık ilişkisini ölçmek içindir. Örneğin kadınlara kıyasla erkekler arasındaki verilerde kullanılır. Biz ise ikinciyi tıklıyoruz çünkü bir ilişki olup olmadığını öğrenmek istiyoruz.

SELECT A STATISTICAL TEST HERE

OR GET HELP CHOOSING BELOW

Tests that show if there is a difference
E.g. Do females have a larger vocabulary average than males?

2 groups 3 or more

Normal Not normal

Independent Paired

Tests that show if there is a relationship
E.g. Does wealth increase with age?

Names Only Ordered

Normal

Answering questions about your data

Groups Normality Data Type

Statistical Test

ANOVA

Chi Square

Correlation - Pearson's

Correlation - Spearman's

Kruskal-Wallis H

Mann-Whitney U

t-test - independent

t-test - paired

Wilcoxon Signed Ranks

Tips

The Pearson's R Correlation test is probably a good choice if you are testing linear correlation. Always use the scatterplot to decide if a linear relationship is appropriate. The Spearman's R Correlation test may be preferable in cases of its resistance to extreme outliers (isolated high or low values).

If your data is numerical, you can evaluate normality by clicking on the "Normality" button down the bottom left.

CONFIGURE TEST

En son burası tıklanır

İlişki gösteren, oransal/kademeli sayı türü seçenekleri tıklanmalı

İsimsel ölçü olmadıkça normal tıklanır

Hangi sınamayı yapacağını otomatik olarak seçer