# 

***Altıncı Bölüm***

***Değişkenler Arasındaki İlişkiler***

******

***Hedefler***

Bu üniteyi çalıştıktan sonra;

* Çapraz Tablo Analizini öğrenecek
* Phi Coefficient’ı hesaplamayı ve yorumlamayı öğrenecek
* İki sürekli değişken arasındaki ilişkiyi öğrenecek
* Uç değer analizi yapmayı öğreneceksiniz

## *Anahtar Kavramlar*

Çapraz Tablo

Phi Coefficient

Correlation Coefficient

Scatter Plots

Pearson Correlation Coefficient

Spearman Correlation Coefficient

Outliers

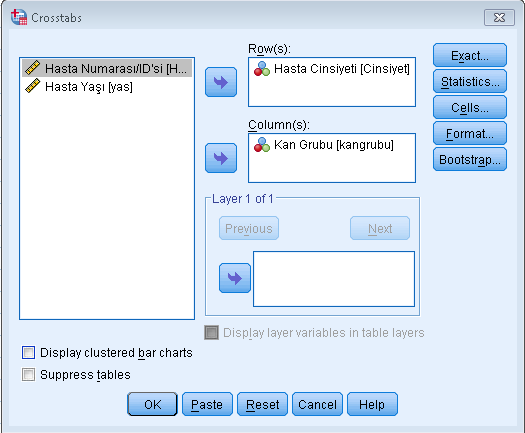
Saplı Kutu grafiği

***İçindekiler***

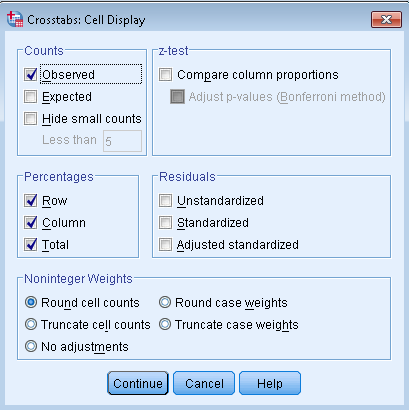
1. Çapraz tablo Analizi
2. Phi Coefficient
3. İki Sürekli Değişken arasındaki ilişki
4. Uç değer analizi
5. **Çapraz Tablo (Crosstab) Analizi**

İki veya daha fazla değişken hakkındaki analizler frekans tabloları ve çapraz tablolar aracılığıyla daha net ortaya konulabilmektedir. İki veya daha fazla seçeneğe sahip, iki değişken arasındaki ilişki çapraz tablolar ile ortaya konabilir. Örneğin daha önce yaptığımız örneklerden olan hasta kaydı örneğinde hastaların cinsiyetleri ile kan grupları arasındaki ilişkiyi araştırmak istediğimizi düşünelim. Yani hangi kan grubundaki hastaların, hangi cinsiyete sahip olduğunu ortaya koymak isteyelim. Bunun için önce hasta kaydi.sav dosyasını çağırmamız gerekecektir.

File > Open > Data menüsünü takip ederek kaydettiğimiz yerden hasta kaydi.sav dosyasını çağıralım. Çapraz Tablo analizi yapmak için Analyze > Descriptive Statistics > Cross Tab menüsünü takip edelim. Aşağıda gösterildiği gibi değişkenlerimizden biri olan hasta Cinsiyeti [Cinsiyet] değişkenini sağda bulunan Row(s) kutucuğuna, Kan grubu [Kangrubu] değişkenini Column(s) kutucuğuna sol ok tuşu yardımı ile atalım.



Daha sonra sağ tarafta bulunan Cells kutucuğuna tıklayalım. Açılan diyalog kutusundan Percentages grubu altında Row, Column ve Total kutucuklarını aşağıda görüldüğü gibi tıklayalım.



Sonra Continue tuşuna ve bir sonraki ekranda da OK tuşuna basalım. Output penceresinde aşağıda yer alan tablo ortaya çıkacaktır.

| **Hasta Cinsiyeti \* Kan Grubu Crosstabulation** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Kan Grubu | | | | Total |
| 0 | A | AB | B |
| Hasta Cinsiyeti | Erkek | Count | 2 | 1 | 1 | 2 | 6 |
| % within Hasta Cinsiyeti | 33,3% | 16,7% | 16,7% | 33,3% | 100,0% |
| % within Kan Grubu | 50,0% | 33,3% | 100,0% | 100,0% | 60,0% |
| % of Total | 20,0% | 10,0% | 10,0% | 20,0% | 60,0% |
| Kadın | Count | 2 | 2 | 0 | 0 | 4 |
| % within Hasta Cinsiyeti | 50,0% | 50,0% | ,0% | ,0% | 100,0% |
| % within Kan Grubu | 50,0% | 66,7% | ,0% | ,0% | 40,0% |
| % of Total | 20,0% | 20,0% | ,0% | ,0% | 40,0% |
| Total | | Count | 4 | 3 | 1 | 2 | 10 |
| % within Hasta Cinsiyeti | 40,0% | 30,0% | 10,0% | 20,0% | 100,0% |
| % within Kan Grubu | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |
| % of Total | 40,0% | 30,0% | 10,0% | 20,0% | 100,0% |

Tabloda yüzdelikler Cells kutucuğunu tıklayarak tüm satır, sütün ve toplam değerlerin yüzdelerini aldığımız için çıkmıştır. Tabloyu yorumlayacak olursak;

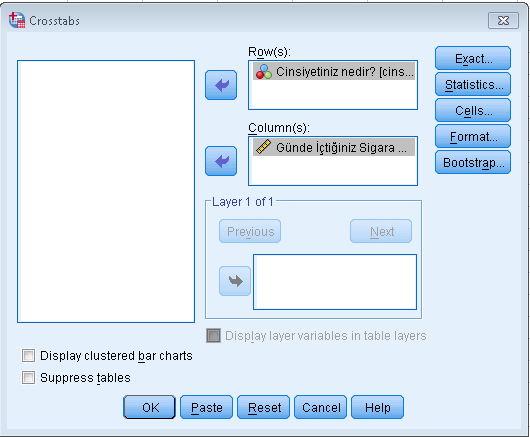
* İlk satır toplam veri seti içerisinde uygun kritere göre kaç örneğin veya vakanın olduğunu göstermektedir. Toplam 6 tane erkek örneği içerisinde 2 tanesi 0, 1 tanesi A, 1 tanesi AB ve 2 tanesi B grubu kana sahiptir. Toplam 4 tane kadın hastanın 2 tanesi 0 grubu ve diğer 2 tanesi A grubu kana sahiptir. AB ve B grubu kana sahip hasta yoktur.
* İlk yüzde satırı hasta cinsiyeti içerisindeki oranı göstermektedir. Erkek hastaların %33,3’ü 0 grubu, %16,7’si A grubu, %16,7’si AB grubu ve %33,3’ü B grubu kana sahiptir. Kadın hastaların %50’si 0 grubu ve geri kalan diğer yarısı (%50’si) de A grubu kana sahiptir.
* İkinci yüzde satırı kan grupları içerisindeki yüzdeleri vermektedir. 0 kan grubuna sahip 4 hasta vardır, bunların 2 tanesi (%50) erkek, 2 tanesi (%50) kadındır. A kan grubuna sahip 3 tane hasta vardır. Bunların 1 Tanesi (%33,3) erkek, 2 tanesi (%66,7) kadındır. AB kan grubuna sahip 1 hasta vardır ve o da Erkektir (%100). B kan grubuna sahip 2 hasta vardır ve bunun tamamı (%100) erkektir.
* Üçüncü yüzde satırı toplam içerisindeki oranlarını göstermektedir. Toplam örnek büyüklüğümüz 10 kişidir. Bunun 4’ü (%40) 0 grubu, 3’ü (%30) A grubu, 1’i (%10) AB grubu ve 2’si (%20) B grubu kana sahiptir.

Çapraz Tablo analizi iki değişkeni iki boyutlu düzlemde karşılaştırma yapmamız açısından oldukça güzel ve önemli bir araçtır.

Örnek: Aşağıda cinsiyetlerin göre hastaların günde içtikleri sigara sayısı verilmiştir. Hastaların Cinsiyetleri ve sigara tüketimleri arasında ne tür bir ilişki vardır?

|  |  |  |  |
| --- | --- | --- | --- |
| **Cinsiyet** | **Sigara Sayısı** | **Cinsiyet** | **Sigara Sayısı** |
| E | 12 | E | 14 |
| K | 8 | K | 5 |
| K | 0 | E | 13 |
| E | 6 | K | 8 |
| K | 7 | E | 0 |
| E | 10 | K | 4 |
| E | 8 | K | 10 |

SPSS’te Kadın 1, Erkek 2 olarak kodlanacaktır ve sigara sayısı da sürekli bir değişken olarak girilecektir.,



Cells kutucuğunun altında column, row ve total değerlerinin percent’leri de işaretlenir. SPSS çıktısı aşağıdaki gibi olacaktır.

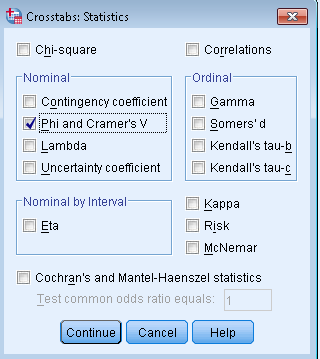
| **Cinsiyetiniz nedir? \* Günde İçtiğiniz Sigara Sayısı Nedir? Crosstabulation** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Günde İçtiğiniz Sigara Sayısı Nedir? | | | | | | | | | | Total |
| ,00 | 4,00 | 5,00 | 6,00 | 7,00 | 8,00 | 10,00 | 12,00 | 13,00 | 14,00 |
| Cinsiyetiniz nedir? | Kadın | Count | 1 | 1 | 1 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 7 |
| % within Cinsiyetiniz nedir? | 14,3% | 14,3% | 14,3% | ,0% | 14,3% | 28,6% | 14,3% | ,0% | ,0% | ,0% | 100,0% |
| % within Günde İçtiğiniz Sigara Sayısı Nedir? | 50,0% | 100,0% | 100,0% | ,0% | 100,0% | 66,7% | 50,0% | ,0% | ,0% | ,0% | 50,0% |
| % of Total | 7,1% | 7,1% | 7,1% | ,0% | 7,1% | 14,3% | 7,1% | ,0% | ,0% | ,0% | 50,0% |
| Erkek | Count | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 7 |
| % within Cinsiyetiniz nedir? | 14,3% | ,0% | ,0% | 14,3% | ,0% | 14,3% | 14,3% | 14,3% | 14,3% | 14,3% | 100,0% |
| % within Günde İçtiğiniz Sigara Sayısı Nedir? | 50,0% | ,0% | ,0% | 100,0% | ,0% | 33,3% | 50,0% | 100,0% | 100,0% | 100,0% | 50,0% |
| % of Total | 7,1% | ,0% | ,0% | 7,1% | ,0% | 7,1% | 7,1% | 7,1% | 7,1% | 7,1% | 50,0% |
| Total | | Count | 2 | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | 14 |
| % within Cinsiyetiniz nedir? | 14,3% | 7,1% | 7,1% | 7,1% | 7,1% | 21,4% | 14,3% | 7,1% | 7,1% | 7,1% | 100,0% |
| % within Günde İçtiğiniz Sigara Sayısı Nedir? | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% | 100,0% |
| % of Total | 14,3% | 7,1% | 7,1% | 7,1% | 7,1% | 21,4% | 14,3% | 7,1% | 7,1% | 7,1% | 100,0% |

Yukarıdaki çapraz Tablo’dan kadınların genellikle daha az sayıda sigra içtikleri erkeklerin kadınlara göre daha fazla sigara günde tükettikleri ortaya çıkmaktadır. Günde 12, 13 ve 14 sigara içen bayan yokken, 1’er erkek günde 12, 13 ve 14 sigara tüketmektedir. Ayrıca günde 4 ve 5 adet sigara içen erkek mevcut değildir. Tablo’dan çıkabilecek bir diğer sonuçta en fazla günde tüketilen sigara sayısı 8’dir ve bu 3 kişi toplam katılımcıların %21,4’üne denk gelmektedir.

1. **Phi Coefficient**

Phi Coefficient iki X İki’lik bir çapraz tablodaki ilişkinin indeks derecesini veren bir katsayıdır. Phi Katsayısı -1 ile +1 arasında yer alır. -1 ve +1’e yakın olan Phi değerleri yüksek ilişki seviyesini, 0’a yakın olan Phi değerleri düşük ilişki seviyesini gösterir.

Yukarıdaki örneğimizde kan grubu ve cinsiyet arasındaki ilişkinin derecesine bakacak olursak. Analyze > Descriptive Statistics > Crosstab menüsü çağrılır. Hasta cinsiyeti Row(s)’a, Kan grubu Cloumn(s) hücresine atanır. Sonra Crosstab menüsünde sağ tarafta görülen Statistics menüsü tıklanır ve aşağıda görülen Statistics diyalog kutusu açılır. Çapraz Tablo analizinde kullandığımız iki değişkende nominal yani kategorik değişkenler olduğu için Nominal grubu içerisindeki Phi ve Cramer’s V kutucuğu tıklanır.



Continue ve OK tuşlarına basılır. Output ekrarında aşağıdaki Tablo elde edilir.

| **Symmetric Measures** | | | |
| --- | --- | --- | --- |
|  | | Value | Approx. Sig. |
| Nominal by Nominal | Phi | ,553 | ,383 |
| Cramer's V | ,553 | ,383 |
| N of Valid Cases | | 10 |  |

İki nominal değişken olan cinsiyet ve kan grubu arasındaki ilişkide Phi ve Cramer’s V değerleri 0,553 çıkmış ve anlamlılık derecesi Approx. Sig. 0,383 olarak hesaplanmıştır. Buradaki hesaplanan index değeri olan 0,553’ü yorumlamak için anlamlılık derecesine bakmak gerekmektedir. Anlamlılık dereceleri eğer 0,05’in altında ise o değer istatistiksel olarak %5 anlamlılık seviyesinde anlamlıdır diye tanımlanır. Eğer değer 0,01’de düşük ise de o değer %1 anlamlılık seviyesinde anlamlıdır denir. Hesaplanan 0,383 değeri 0,05’ten doğal olarak da 0,01’de oldukça büyüktür. Dolayısıyla anlamsızdır. Bu durumda örneğimizdeki hastaların kan grupları ile cinsiyetleri arasında bir ilişki yoktur dememiz gerekmektedir ki bu da beklenen bir durumdur. İnsanların cinsiyetleri ile kan grupları orasında herhangi bir ilişki olmadığı ve dolayısıyla istatiksel olarak da bir ilişki aranamayacağı açıktır.

**3. İki Sürekli Değişken Arasındaki İlişki**

Bu bölümde iki sürekli değişken arasındaki ilişkiye ve bu ilişkinin gücünü hesaplamaya çalışacağız. İlişki grafik olarak veya sayısal olarak ortaya konulup, hesaplanabilir. Grafik teknik Dağılım grafiği (Scatter Plot), sayısal teknik korelasyon katsayısı’dır (Correlation Coefficient) ve r ile gösterilir.

Bireylerin çocuk sayıları ile gelirleri arasında bir ilişki olup, olmadığını araştırdığımız varsayalım. İnsanlara bir anket yapıp kaç çocuğa sahip olduklarını ve TL olarak aylık ortalama gelirlerini sorduğumuzu düşünelim. Burada her iki değişkende sürekli birer değişken olarak düşünülmüş ve 10 farklı kişi için aşağıdaki gibi belirlenmiş olsun.

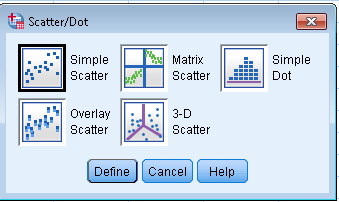
|  |  |
| --- | --- |
| Aylık Ortala Gelir (TL) | Çocuk Sayısı |
| 1,500 | 1 |
| 2,500 | 2 |
| 800 | 0 |
| 3,800 | 3 |
| 2,600 | 2 |
| 4,200 | 3 |
| 6,300 | 4 |
| 5,250 | 4 |
| 850 | 1 |
| 1,000 | 1 |

File > New > Data seçeneğini seçerek yeni bir veri seti açalım ve yukarıda yer alan verileri daha önce açıklandığı gibi SPSS’e data Editörünü kullanarak girelim. Gelir değişkenine gelir, Çocuk Sayısı değişkenine de cocuksay ismini verelim. Label seçeneğine uzun uzun isimlerini yazalım ve ölçütlerini scale olarak belirleyelim.

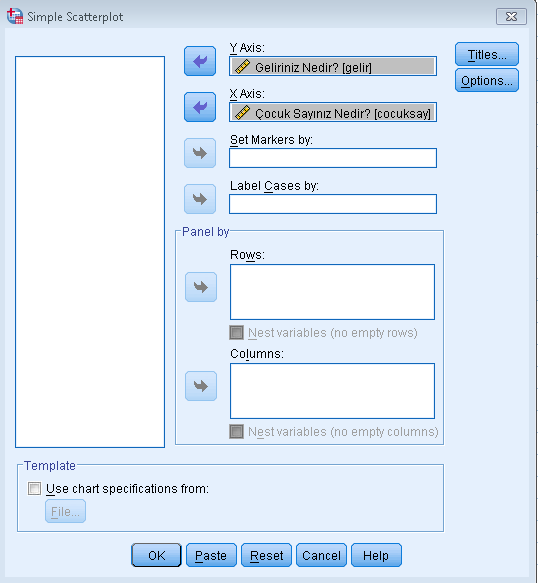
Dağılım Grafikleri (Scatter Plots)

Dağılım grafikleri iki sayısal değişken arasındaki ilişkinin yönünü belirlemek için kullanılan görsel bir araçtır. Veri x ekseni ve y ekseni üzerinde yer alır. Eksenler üzerinde her bir nokta gözlemlenen bir değeri ifade eder. Yani örneğimizde gelir ve çocuk sayısı ikilisini.

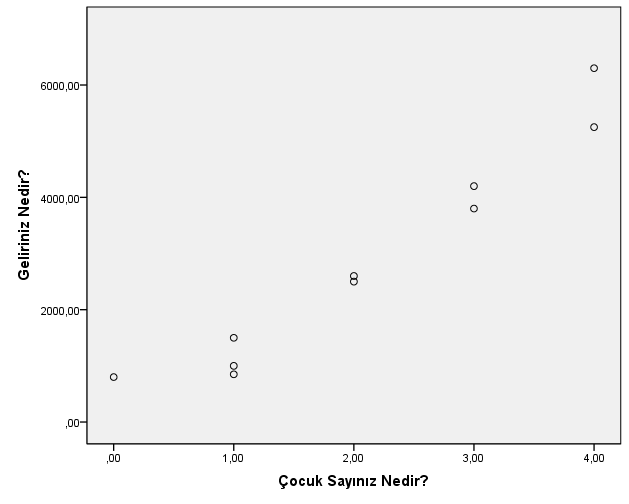
Dağılım Grafiği elde etmek için Ana menüden Graphs menüsü seçilir. Aşağıya düşen menüden Legacy Graphs menüsü açılır ve son olarak da oradan Scatter/Dots grafiği seçilir. Açılan Scatter/Dot diyalog Menüsünden aşağıda gösterildiği gibi Simple Scatter seçilir ve define tuşuna basılır.



Açılan Simple Scatter diyalog kutusundan Y Axis’e Geliriniz Nedir?[Gelir] ve X Axis’e de Çocuk Sayınız Nedir?[cocuksay] değişkenleri aşağıda gösterildiği gibi sağ ok kullanarak konulur. Burada unutulmaması gereken sol kutucukta bulunan bir değişkeni sağ tarafa atmadan önce sağ ok’a basmadan, fare ile üzerine tıklanıp, seçilmesi gerekmektedir.



Ok tuşuna basılarak aşağıda gösterilen Scatter Plot output menüsünde yerini alır.



Grafikten görüldüğü gibi hiç çocuğu olmayan ve 800 TL geliri olan ve 4 çocuğu olan ve 6,300 TL geliri olan bireyler ve tüm diğer 8 anket katılımcısı Scatter Plot’da gösterilmiştir. Grafikten açıkça görüleceği gibi örneğimizdeki ailelerin gelirleri ile sahip oldukları çocuk sayısı arasında pozitif bir ilişki vardır. Yani çok gelire sahip olan aileler daha çok, az gelire sahip olan aileler ise daha az sayıda çocuğa sahip olmaktadırlar.

SPSS çıktı (Output) ekranında Dağılım Grafiği üzerine iki kere tıklandığında Chat editörü açılmaktadır. Aşağıda görülen Chart editörünü kullanarak X ve Y eksenlerinin ölçülerini ve eksenlerin isimlerini, grafiğimizi başlığını ve diğer pek çok grafikle ilgili unsuru değiştirip, yeniden ayarlama yapabiliriz.

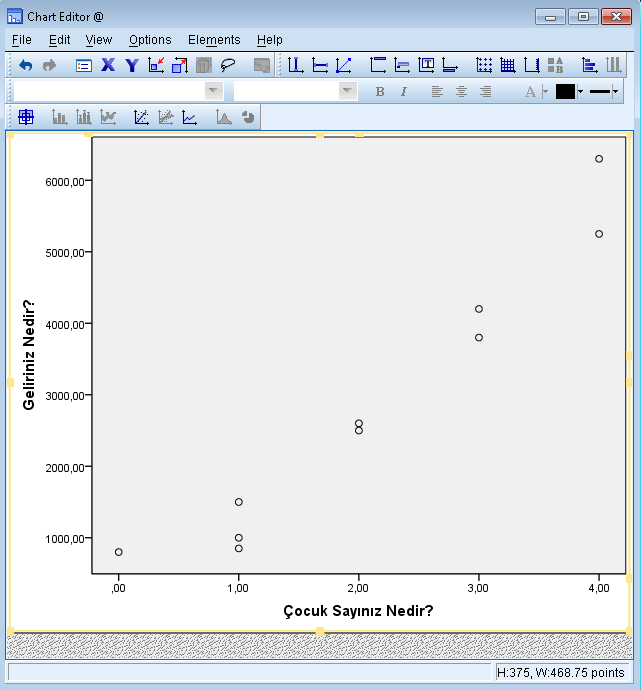
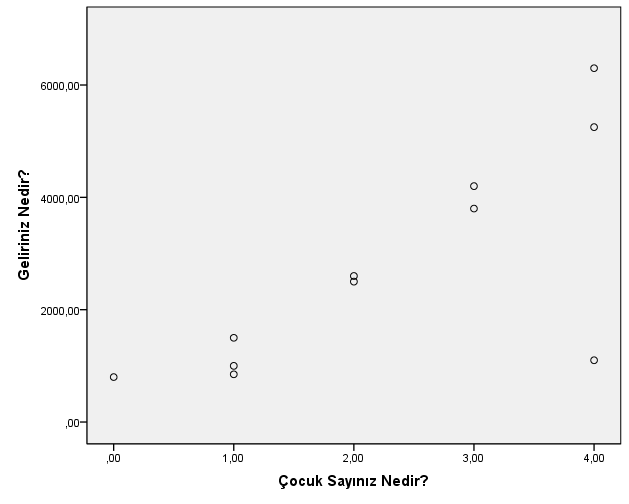


Chart Editör açıkken eksenlerin üzerine veya eksenlerdeki başlıklara tıklayarak gerekli menüleri açarak istenilen değişiklikleri yapmak mümkündür.

Scatter Plot ayrıca bize veri setimizde alışılagelmiş veya beklenen dışındaki verilerin durumunu da göstermektedir. Örneğin veri setimize 11. Aile olarak geliri 1,100 TL olan ve 4 çocuğa sahip bir aileyi de ekleyelim ve yeniden Scatter Plot’ı çizelim.

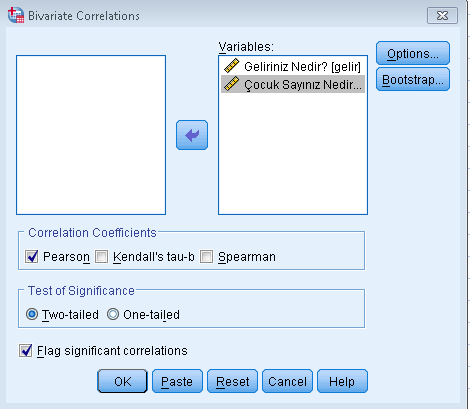


Yukarıdaki Grafikten kolayca gözlemleneceği gibi son girdiğimiz değer bir outlier (aykırı değer) olarak veri setinde olduğunu Dağılım Grafiği ile kolayca belirleyebiliriz. Araştırmacıya düşen bunun neden olduğunu bulmaktır. Bunun sebebi basit bir veri girişi hatası olabileceği gibi, üzerinde düşünülmesi gereken çok önemli bir sosyal olay da olabilir.

Korelasyon Katsayısı (r)

Pearson Korelasyon katsayısı iki sürekli değişken arasındaki ilişkinin derecesini belirlemek için kullanılan bir araçtır. Korealsyon katsayısı -1 ile +1 arasında yer alır. Pozitif değerler (0,20, 0,40, 0,80 gibi) pozitif yani doğrusal bir ilişkiyi; negatif değerler (-0,20, -0,30, -0,40, -0,80) negatif yani ters bir ilişkiyi gösterir. Değerler 0’ya yaklaştıkça ilişkinin kuvveti azalır ve 0 korelasyon katsayısı iki değişken arasında hiçbir ilişki olmadığının ifadesidir. Değerler -1 ve +1’e yaklaştıkça ilişkinin gücü artar. -1 tam negatif ilişki, +1 tam pozitif ilişkiyi ifade eder.

Korelasyon katsayısı hesaplamayı yine ailelerin aylık geliri ve çocuk sayısı örneğinden faydalanarak göstereceğiz. Ana menüden Analyze menüsü ardında Corrolate menüsü ve bivariate korelesyon seçeneği seçilir. Bivariate Correlation diyalog kutusu açıldığında aşağıda gösterildiği gibi Geliriniz Nedir?[gelir] ve Çocuk Sayınız nedir?[cocuksay] değişkenleri ok yardımı ile sağ tarafta yer alan Variable(s) kutucuğuna atılır. Pearson Korelasyon Katsayısının seçili olduğuna emin olunur.



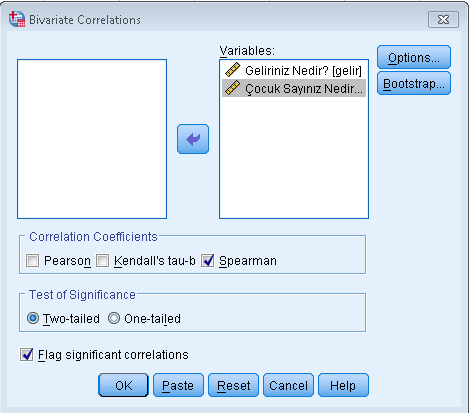
En olarak da OK tuşuna basılır ve aşağıda gösterilen korelasyon Tablosu elde edilir.

| **Correlations** | | | |
| --- | --- | --- | --- |
|  | | Geliriniz Nedir? | Çocuk Sayınız Nedir? |
| Geliriniz Nedir? | Pearson Correlation | 1 | ,740\*\* |
| Sig. (2-tailed) |  | ,009 |
| N | 11 | 11 |
| Çocuk Sayınız Nedir? | Pearson Correlation | ,740\*\* | 1 |
| Sig. (2-tailed) | ,009 |  |
| N | 11 | 11 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | |

Gelir ve Çocuk Sayısı arasındaki ilişki 11 örnek için (N) Pearson korelasyon Katsayısı ile hesaplanmış ve bu katsayı 0,740 çıkmıştır. Bunun anlamı gelirle, çocuk sayısı arasında oldukça güçlü ve pozitif yani doğrusal bir ilişkinin varlığıdır. Ayrıca bu katsayının anlamlılık değeri de hesaplanmış ve Significance (Sig. – 2 tailed) değeri 0,009 çıkmıştır. Bu değer 0,01’den küçük olduğu için %1 anlamlılık seviyesinde çocuk sayısı ve gelir arasındaki pozitif ilişki istatiksel olarak anlamlıdır diye ifade edilir.

Korelasyon nedensellik anlamına gelmemektedir. Yani iki değişken arasında yüksek korelasyon çıkması bu değişkenlerin birinin diğerine neden olduğu anlamında yorumlanmamalıdır.

Eğer verimizin normal dağılmadığını düşünüyorsak Spearman korelasyon katsayısını hesaplamalıyız. Spearman korealsyon katsayısı hesaplanırken önce veriler sıralanır ve sıralanan verilere Pearson korelasyon katsayısı uygulanır. Analyze > Correlate > Bivariate menüsü seçilir. Ardından aşağıda şekilde gösterildiği gibi gelir ve çocuk sayısı variables kutucuğu içerisine alınır ve Pearson iptal edilir ve Spearman korelasyon katsayısı seçilir.

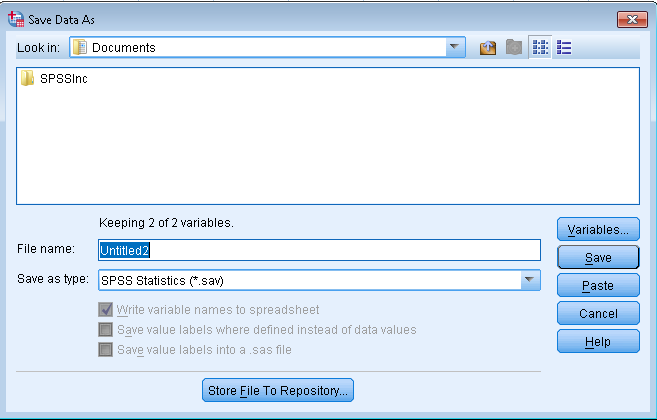


Spearman korealsyon katsayısının SPSS çıktısı aşağıdadır.

| **Correlations** | | | | |
| --- | --- | --- | --- | --- |
|  | | | Geliriniz Nedir? | Çocuk Sayınız Nedir? |
| Spearman's rho | Geliriniz Nedir? | Correlation Coefficient | 1,000 | ,782\*\* |
| Sig. (2-tailed) | . | ,004 |
| N | 11 | 11 |
| Çocuk Sayınız Nedir? | Correlation Coefficient | ,782\*\* | 1,000 |
| Sig. (2-tailed) | ,004 | . |
| N | 11 | 11 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | | |

Spearman Korealsyonu da %1 anlamlılık seviyesinde anlamlı çıkmış ve değeri 0,782 olarak hesaplanmıştır. Testimiz çift kuyruklu/taraflıdır ve örnek sayımız 11 adettir. Bu pozitif ve yüksek sayılabilecek bir ilişkinin gelir ve çocuk sayısı arasında olduğunu bize göstermektedir. Ancak bu ilişki tamamen istatistiksel olabilir ve bir nedensellik olmayabilir. Yani korelasyon bir nedensellik göstermeyebilir. Fazla veya az çocuk sahibi olmak üzerine etkili diğer değişkenlere de bakılmalıdır.

SPSS’i kapatmadan önce, yeni yarattığımız gelir ve coçuk sayısı arasındaki ilişkiyi gösteren veri setimizi kaydetme düğmesine  basarak kaydedelim. Kaydetme düğmesine batığımızda aşağıdaki gibi değişik kaydetme (Save As) menüsü açılacaktır.



Look in aşağıya düşen menüden bilgisayarımızda nereye kaydedeceğimiz belirleyebiliriz. Burada taşınabilir flash belleğimizi bilgisayarımıza takıp, aktif hale getirerek, genellikle F: olarak tanımlanan flash belleğimizi seçebiliriz. Flash bellekte yeni bir klasör yaratmak istediğimizi varsayalım. Save As diyalog kutusunda bunu yapmak oldukça basittir. “Look in” menüsü sırasında bulunan  yeni klasör yaratma düğmesine basmamız ve hafta 6 diye klasörümüze isim vermemiz yeterli olacaktır. Bu arada File Name kutucuğunda bulunan ve default yani varsayım olarak Untitled2 diye isimlendirilen veri setimize Gelir çocuk sayısı ismini verip SAVE düğmesine bastığımızda, veri setimiz belirlediğimiz klasöre gelir çocuk sayisi.sav olarak bir SPSS dosyası biçiminde kayıt edilecektir.

**4. Uç Değer Analizi**

Veri seti analiz edilirken bir diğer aşama da uç değer analizidir. Uç değer analizi iki nedenle önemlidir:

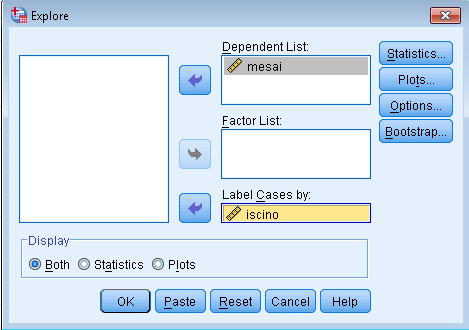
1. Uç değerler normal sonuçlar elde etmeyi engelleyeceği için tespit edilip veri setinden çıkartılmalıdır.
2. Uç değerler birer bilgi kaynağıdır. Tespit edilip nedenleri araştırılmalıdır

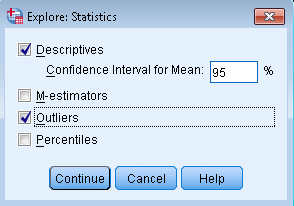
Aşırı değerler, çok aşırı değerler (extreme values) ve aşırı değerler (outlier values) olarak ikiye ayrılırlar. Bu değerler hatalı veri girişinden veya nadir olan gözlemlerden dolayı olabilirler.

Örnek: Aşağıda 20 işçinin işyerinde yaptıkları fazla mesailer verilmiştir.

|  |  |  |  |
| --- | --- | --- | --- |
| **İşçi** | **Mesai** | **İşçi** | **Mesai** |
| 1 | 2 | 11 | 6 |
| 2 | 4 | 12 | 1 |
| 3 | 3 | 13 | 3 |
| 4 | 6 | 14 | 5 |
| 5 | 2 | 15 | 15 |
| 6 | 6 | 16 | 3 |
| 7 | 3 | 17 | 5 |
| 8 | 4 | 18 | 6 |
| 9 | 12 | 19 | 5 |
| 10 | 3 | 20 | 14 |

Uç değer analizinde hangi işçilerin daha fazla mesai yaptıklarını görebiliriz. Bunun için Analyze > Descriptive Statistics > Explore seçenekleri seçilir ve burada dependent kısmına “Mesai”, Label Cases kısmına “İşçi” değeri girilir. Ardından Statistics menüsünün altında Outliers seçilir.

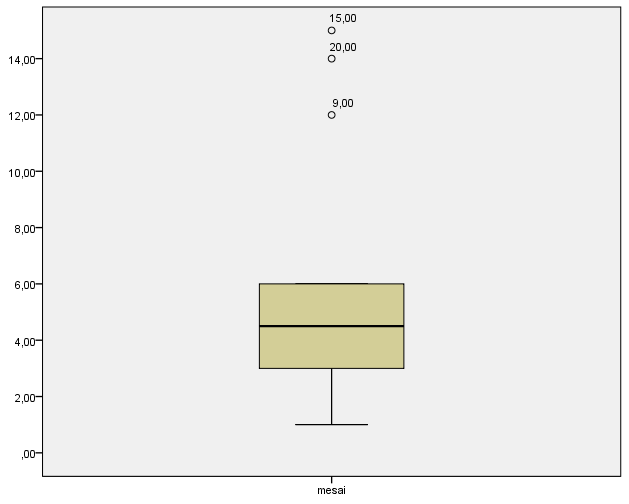




İlk önce en yüksek ve en düşük değerleri içeren analizi görelim.

| **Extreme Values** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | | Case Number | iscino | Value |
| mesai | Highest | 1 | 15 | 15,00 | 15,00 |
| 2 | 20 | 20,00 | 14,00 |
| 3 | 9 | 9,00 | 12,00 |
| 4 | 4 | 4,00 | 6,00 |
| 5 | 6 | 6,00 | 6,00a |
| Lowest | 1 | 12 | 12,00 | 1,00 |
| 2 | 5 | 5,00 | 2,00 |
| 3 | 1 | 1,00 | 2,00 |
| 4 | 16 | 16,00 | 3,00 |
| 5 | 13 | 13,00 | 3,00b |
| a. Only a partial list of cases with the value 6,00 are shown in the table of upper extremes.  b. Only a partial list of cases with the value 3,00 are shown in the table of lower extremes. | | | | | |

Yukarıdaki Tabloda mesaisi en yüksek olan 5 işçi ve en düşük olan 5 işçi ve bunların mesai değerleri verilmiştir. 15, 20, 9, 4 ve 6 nolu işçilerin mesaileri en fazladır. Bunun karşılığında 12, 5, 1, 16 ve 13 nolu işçilerin mesaileri en düşüktür. Sonra aşağıda yer alan saplı kutu grafiğini inceleyelim.



Saplı kutu grafiğinden görüldüğü gibi 15, 20 ve 9 nolu işçiler aşırı mesai yapmışlardır ve normal mesai saati 4 saatin biraz üzerinde iken bu işçilerde 12 saatin üzerine çıkmıştır. Bu yükseklik ve düşüklüklerin nedeni araştırılmalıdır.

**Değerlendirme Soruları**

1. Çapraz Tablo analizi nasıl yapılmaktadır ve nasıl yorumlanmaktadır. Tartışınız.
2. Phi Coefficient ne işe yaramaktadır ve tür bilgi bize vermektedir.
3. Korelasyon katsayının amacı nedir? Korelasyon katsayısı nedensellik gösterir mi?
4. Uç değer analizi neden önemlidir ve nasıl yapılmaktadır.

**Kaynakça**

Spiers, N., Manktelow, B. Ve Hewitt, M. J.(2009), Practical Statistics Using SPSS, National Institude for Health Research NHS, England.

Field, A. (2005), Discovering Statistics using SPSS, SAGE, London.

Kalaycı, Ş. (2006), SPSS Uygulamalaı Çok değişkenli İstatistik Teknikleri, Asil Yayın Dağıtım