# Numerik Veri İkiden Fazla Grup

Amaç: Bu konu sonunda okuyucunun ikiden fazla bağımsız gruptan elde edilen numerik verilerin ortalamalarının karşılaştırılmasında kullanılan tek yönlü varyans analizi (ANOVA) ve bunun nonparametrik alternatifi olan Kruskal Vallis testi hakkında bilgi sahibi olması ve bu testleri SPSS ile yapabilmesi amaçlanmıştır.

Hedefler:

* Numerik veri, ikiden fazla bağımsız grubu tanımlayabilmeli
* ANOVA varsayımlarını sayabilmeli
* SPSS’te ANOVA testi yapabilmeli
* ANOVA için Post Hoc analiz yapabilmeli
* ANOVA testi SPSS çıktısını yorumlayabilmeli
* SPSS’te Kruskal Vallis testi yapabilmeli
* Kruskal Vallis testi SPSS çıktısını yorumlayabilmeli

Diyabet hastalarımızın eğitim durumlarına göre boylarının farklılık gösterip göstermediğini araştırmak istediğimizi varsayalım (*H0*: Eğitim durumları farklı olan diyabet hastalarının boyları arasında fark yoktur). Ölçmek istediğimiz değişken (boy) numerik sürekli bir değişkendir. Eğitimsiz, ilkokul, ortaokul ve lise olmak üzere dört grubumuz var.

Yukarıdaki hipotezimizi test etmek için uygulayabileceğimiz test tek yönlü varyans analizidir (Analysis of Variance; ANOVA). Bu testi uygulayamamız halinde nonparametrik alternatifi olan Kruskal Vallis testini yapabiliriz.

## ANOVA

### Varsayımları

Tek yönlü varyans analizinde gruplar tek bir faktörle belirlenir (örn. babanın mesleğinin öğrencinin ÖSS sınav notuna etkisi). Birden fazla faktörün birbiriyle ilişkisinin incelendiği durumlarda (örn. babanın mesleğinin VE okul türünün öğrencinin ÖSS sınav notuna etkisi) çok yönlü varyans analizi uygulanır.

* İncelenen toplumda değişken, grupların hepsinde normal dağılmalıdır.
* Değişkenin gruplar arasındaki varyansı eşit olmalıdır.
* Örneklem sayısı bu varsayımları kontrol edecek kadar büyük olmalıdır.

### Mantığı

Bu test grup içi ve gruplar arası varyansların oranına dayanmaktadır. Sıfır hipoteezinin doğru olması halinde değişkenin grupların genelindeki varyasyonu ile gruplar arasındaki varyasyonu aynı olmalıdır. Testin adının varyans analizi olması da buradan gelmektedir.

Şimdi, diyabet.sav verisetimizde “height” değişkenimizin eğitim grupları açısından durumunu incelemek için ANOVA testini uygulayalım:

1. **Sıfır hipotezi (H0) ve alternatif hipotezin (H1) tanımlanması:**

*H0*: Eğitim durumları farklı olan diyabet hastalarının boyları arasında fark yoktur.

*H1*: Eğitim durumları farklı olan diyabet hastalarının boyları arasında fark vardır.

1. **Verilerin toplanması:**

Veriler diyabet.sav verisetinde “height” ve “educat” değişkenlerinin altında kaydedildi.

Örneklem sayısı yeterlidir. Değişkenimizin gruplar açısından histogram grafiğine bakıldığında normal dağıldığı görülmektedir:



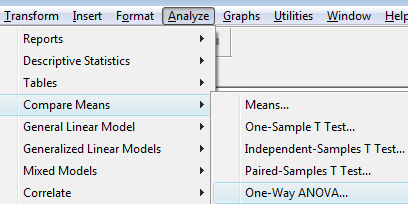
Varyansların homojenliği varsayımı ANOVA testi sırasında incelenecektir.

1. **İlgili sıfır hipotezi için *test istatistiğinin* hesaplanması:**

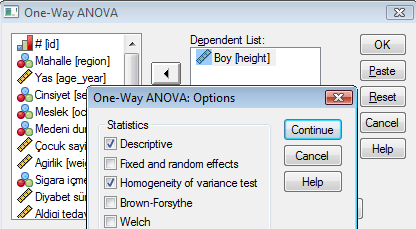
ANOVA testi grup için ve gruplar arası varyansların birbirine oranlanmasından (*F* oranı) oluşur. Bu oran *F* dağılımına uymaktadır.

SPSS ile yapacak olursak:

*Analyze>Compare Means>One-Way ANOVA>[“height” değişkenini “Dependent List” alanına, “educat” değişkenini de “Factor” alanına geçirelim*



*>Options >[“Descriptive” ve “Homogeneity of variance test” kutucuklarını işaretleyelim]>Continue>ok.*

****

Aşağıdaki çıktıları elde ederiz:



**Test of Homogeneity of Variances**

Boy

|  |  |  |  |
| --- | --- | --- | --- |
| Levene Statistic | df1 | df2 | Sig. |
| ,434 | 3 | 397 | ,729 |

**ANOVA**

Boy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 7252,244 | 3 | 2417,415 | 35,938 | ,000 |
| Within Groups | 26704,754 | 397 | 67,266 |  |  |
| Total | 33956,998 | 400 |  |  |  |

Birinci tablo, değişkenin gruplara göre tanımlayıcı istatistiklerini vermektedir.

İkinci tabloda varyansların homojenliği test edilmektedir. *p* (Sig.) >0,05 olduğundan eğitim gruplarına göre varyanslar arasında fark olmadığı sonucunu çıkarabiliriz.

Üçüncü tablo F istatistiğini veriyor. Gruplar arası varyansların kareleri ortalamasının (2417,415) gruplar içi varyansların kareleri ortalamasına (67,266) *F* değerini vermektedir.

1. **Test istatistiğinden elde edilen değerin bilinen bir olasılık dağılımı ile karşılaştırılması:**

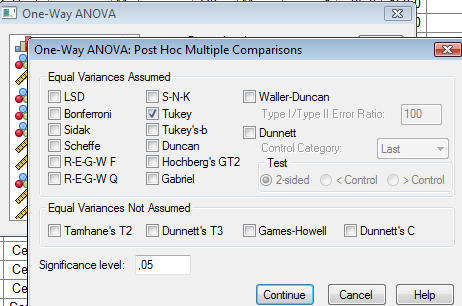
*F* oranı için appendix A5’e baktığımızda p değerinin 0,05’ten küçük olduğunu görürüz. Zaten SPSS’te *p*’yi sıfıra yakın bir değer olarak vermiştir.

1. ***P değerinin* ve sonuçların yorumlanması**

*P* değeri 0,05’ten küçük olduğu için *H0* hipotezini reddederiz. Dolayısıyla diyabetik erkeklerin VKİ değerlerinin diyabetbireylerin eğitim durumlarına göre boylarının farklılık gösterdiği sonucuna varabiliriz.

ANOVA bize incelenen gruplar arasında bağımlı değişken açısından fark olduğunu söylemektedir. Grupların hepsinin birbirinden farklı olması gerekmez. Örneğin, sadece eğitimsiz bireylerle lise mezunları arasında bir fark olabilir. Farkın hangi gruplar arasında olduğunu anlamak için ANOVA’da Post Hoc analiz yapabiliriz. Bunun için yakarıdaki işlemi şu iekilde değiştirmeliyiz:

*Analyze>Compare Means>One-Way ANOVA>[“height” değişkenini “Dependent List” alanına, “educat” değişkenini de “Factor” alanına geçirelim>Options >[“Descriptive” ve “Homogeneity of variance test” kutucuklarını işaretleyelim]>Post Hoc>[“Tukey” kutucuğunu işaretleyelim]>Continue>ok.*

**

Önceki çıktıya ek olarak aşağıdaki tablo oluşacaktır:



Bu tablodan görüldüğü gibi esas fark eğitimsiz bireylerle diğer gruplar arasındadır. Diğer grupların kendi aralarında fark yoktur.

## Varsayımlar karşılanamıyorsa

ANOVA testi normallikten sapmalara kısmen dayanıklı olmakla beraber, varyansların homojen olmaması durumuna dayanıklı değildir. Verilerimizin varsayımları karşılamaması halinde veri dönüşümü uygulayabilir veya bu testin nonparametrik versiyonundan yararlanmamız gerekir.

### Kruskal Vallis Testi

Bu test ANOVA’nın nonparametrik alternatifidir. Gruplar arasında sıraların toplamını dikkate alır. ANOVA varsayımlarını karşılamayan bir değişken için yapacak olursak.

1. **Sıfır hipotezi (H0) ve alternatif hipotezin (H1) tanımlanması:**

*H0*: Medeni durumları farklı olan diyabet hastalarının boyları arasında fark yoktur.

*H1*: Medeni durumları farklı olan diyabet hastalarının boyları arasında fark vardır.

1. **Verilerin toplanması:**

Veriler diyabet.sav verisetinde “height” ve “marital” değişkenlerinin altında kaydedildi.

Örneklem sayısı bazı gruplarda 30’un altındadır. Değişkenimizin gruplar açısından histogram grafiğine bakıldığında da bazı gruplarda normal dağılmadığı görülmektedir:



ANOVA yapıldığında da varyansların homojen dağılmadığı görülmektedir:

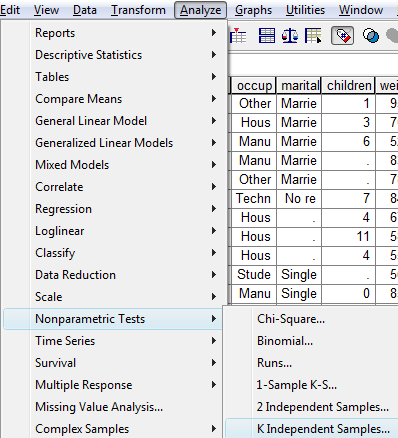
**Test of Homogeneity of Variances**

Boy

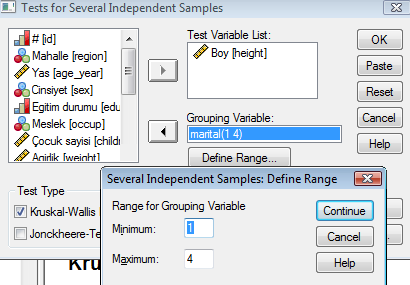
|  |  |  |  |
| --- | --- | --- | --- |
| Levene Statistic | df1 | df2 | Sig. |
| 4,852 | 3 | 392 | ,003 |

1. **İlgili sıfır hipotezi için *test istatistiğinin* hesaplanması:**

*Analyze>Nonparametric Tests >K Independent Samples>[“height” değişkenini “Test Variable List” alanına, “marital” değişkenini de “Grouping Variable” alanına geçirelim*



*>Defiine Range >[Bu değişken için 1’den 4’e kadar 4 grubumuz olduğundan “Minimum” alanına 1, “Maximum” alanına 4 girelim]>Continue>ok.*



Aşağıdaki çıktıları elde ederiz:

**Ranks**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Medeni durum | N | Mean Rank |
| Boy | Single | 22 | 236,02 |
| Married | 313 | 213,72 |
| Divorced | 20 | 117,35 |
| Widow | 41 | 101,73 |
| Total | 396 |  |

**Test Statistics(a,b)**

|  |  |
| --- | --- |
|  | Boy |
| Chi-Square | 47,341 |
| Df | 3 |
| Asymp. Sig. | ,000 |

a Kruskal Wallis Test

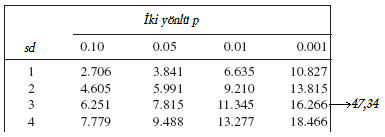
b Grouping Variable: Medeni durum

Birinci tablo, değişkenin gruplara göre sıralanmış verilerinin sıra ortalamalarını vermektedir.

İkinci tabloda kruskal Vallis testi sonuçları görülmektedir.

1. **Test istatistiğinden elde edilen değerin bilinen bir olasılık dağılımı ile karşılaştırılması:**

Kruskal-Wallis testi *Ki Kare* dağılımı göstermektedir. 47,3 değerini Tablo A3’ten bakacak olursak:

**

SD=3 için (serbestlik derecesi burada grup sayısı – 1 olarak hesaplanır) iki yönlü *p* değerinin 0,001’den küçük olduğunu görürüz. Zaten SPSS’te *p*’yi sıfıra yakın bir değer olarak vermiştir.

1. ***P değerinin* ve sonuçların yorumlanması**

*P* değeri 0,05’ten küçük olduğu için *H0* hipotezini reddederiz. Dolayısıyla medeni durumları farklı olan diyabet hastalarının boyları arasında fark olduğu sonucuna varabiliriz.