



İSTATİSTİK

6. Varyans Analizi (ANOVA)

ANOVA

- ▶ ANOVA, iki yada daha fazla grup ortalamalarının birbirinden farklı olup olmadığını anlamak için kullanılır.
- ▶ Araştırmacının temel amacı farklı gruptaki (bağımsız değişken) elemanların, ortalama skorları (bağımlı değişken) arasında anlamlı bir farklılığın olup olmadığını anlamaktır. Bir başka ifadeyle, bağımlı değişkendeki değişim deneklerin farklı grupta olmasından mı kaynaklanmaktadır sorusunun cevabı bulunmaya çalışılmaktadır.
- ▶ Bağımlı değişken sayısal, bağımsız değişken ise kategoriktir.
- ▶ ANOVA'da amaç grup ortalamalarının farklı olup olmadığını belirlemektir, ancak bunu gerçekleştirmek için varyanslar hesaplanır.

ANOVA örnekleri

- ▶ Üç farklı marka otomobilin benzin tüketimi birbirinden farklı mıdır?
- ▶ Farklı yaş gruplarındaki insanların (çocuk, genç, orta yaşlı, yaşlı) günlük TV seyretme süreleri farklı mıdır?
- ▶ Farklı yaş (çocuk, genç, orta yaşlı, yaşlı) ve eğitim (ilköğretim, lise, üniversite) gruplarındaki insanların günlük TV seyretme süreleri farklı mıdır?

ANOVA ve T testi

- ▶ T testi iki grup ortalamasının eşit olup olmadığını hipotezini test eder.
- ▶ ANOVA ise iki ya da daha fazla ortalamamanın eşit olup olmadığını test eder.
- ▶ ANOVA hangi grupların birbirinden farklı olduğunu söylemez, sadece bir farklılık bulması halinde grupların bazılarının diğerlerinden farklı olduğunu ifade eder. Hangi gruplar arasında farklılık olduğunu anlamak için ANOVA sonrası farklı testler yapmamız gerekir.

Neden bir çok T testi kullanmıyoruz?

- ▶ Gruplar arası ortalamaların farklılığını test etmek için ikişerli gruplar halinde t testi kullanılması düşünülebilir. Ancak bu durumda I. Tip hata yapma olasılığımız artar.
- ▶ Çünkü, her bir test Tip I (α) hataya sahiptir ve k karşılaştırma olması halinde toplam Tip I hata $1-(1-\alpha)^k$ olur
- ▶ Örneğin, 5 grup ve $\alpha=0.05$ olması durumunda, 10 adet ikişerli karşılaştırma yapmak gerekir. Bu nedenle, toplam Tip I hata $1-(1-0,05)^{10}=0,40$ olacaktır. Bunun anlamı, bu 10 karşılaştırmaların en az birisinde 1. tip hata yapma olasılığı 0,40 olduğudur ve bu kabul edilebilir bir durum değildir.
- ▶ Bu yüzden ikili bir çok karşılaştırma yerine I. Tip hata olasılığı 0.05 olarak belirlenmiş tek bir test yapmak daha uygundur ve bu ANOVA'dır.

Varyans Analizinin Varsayımları

- ▶ Her gruptaki değerler kendi içinde normal dağılım göstermelidir.
- ▶ Grup varyansları eşit olmalıdır. Gruplara deneklerin rastgele atanmış olmaları bu varsayımın sağlanmasına yardımcı olacaktır.
- ▶ Gruplar içinde ve gruplar arasında elde edilen gözlemler bağımsız olmalıdır.

ANOVA Türleri

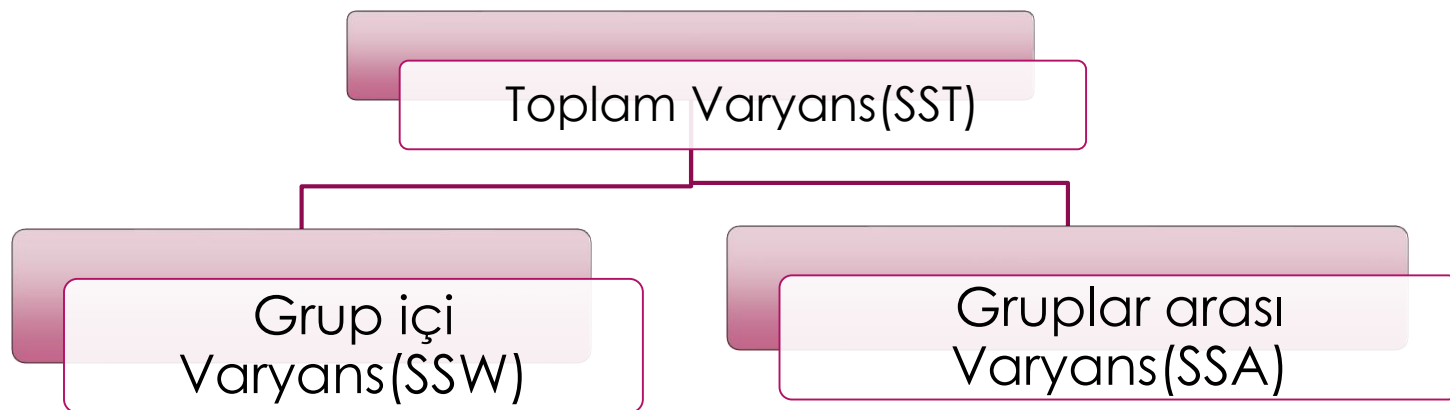
- ▶ ANOVA analizi ölçüm şekline, bağımlı ve bağımsız değişken sayısına göre farklı şekillerde adlandırılabilir. Bu bağlamda,
 1. Ölçüm türlerine göre
 1. Bağımsız ölçümler ANOVA testleri: Farklı gruplar birer kez ölçülür
 2. Tekrarlı ölçümler ANOVA testleri: Aynı denekler birden fazla ölçülür
 3. Karışık tasarım ANOVA testleri: Denekler bazı değişkenler açısından birden fazla ölçülürken bazı değişkenler açısından bir kere ölçülür
 2. Bağımsız değişken sayısına göre
 1. Tek yönlü varyans analizi (one-way ANOVA)
 2. Çok yönlü varyans analizi (m-way ANOVA $m=2,3,4,\dots$)
 3. Bağımlı değişken sayısına göre
 1. ANOVA (Bir bağımlı değişken)
 2. MANOVA (Birden fazla bağımlı değişken)

Tek Yönlü Varyans Analizi

- ▶ K tane kitlenin ortalamaları arasında önemli bir fark olup olmadığı problemi ile ilgileniyorsak $\mu_1, \mu_2, \dots, \mu_k$ söz konusu kitlelerin ortalamalarını göstermek üzere test edeceğimiz H_0 ve H_a hipotezleri;
- ▶ H_0 : Grupların ortalamaları arasında fark yoktur.
- ▶ H_a : En az bir grup ortalaması diğerlerinden farklıdır.
 - ▶ $H_0: \mu_1 = \mu_2 = \dots = \mu_k$
 - ▶ $H_a : \mu_i \neq \mu_j$

ANOVA testi

- ANOVA testinde gruplar arası ve gruplar içi değişime (varyansa) bakılır. Toplam değişkenlik bu iki kısımdan oluşur. Gruplar arası varyansın, grup içi varyansa oranı hesaplanarak bir F istatistiği elde edilir. Bu istatistiğin anlamlı olması halinde H_0 hipotezi reddedilir ve en az bir grubun ortalamasının diğer grup ortalamalarından farklı olduğu sonucuna varılır.



ANOVA hesaplamaları

- ▶ Gruplar arası değişim, her bir grup ortalamasının genel ortalamadan olan farklılığından elde edilir. Karşılaştırılan grupların ortalaması birbirine çok yakınsa, gruplar arası değişim küçük olacaktır. Grup sayısı c ile gösterildiğinde, gruplar arası değişime ilişkin serbestlik derecesi $c-1$ olacaktır.
- ▶ Grup içi değişim her bir gruptaki bireylerin ait oldukları grup ortalamalarından farklılıklardan kaynaklanır. Gruplardaki gözlemler birbirine yakın değerler alıyorsa grup içi değişim de küçük olacaktır.
- ▶ Grup içi değişim için serbestlik derecesi $= n-c$ olarak gösterilir.

Varyans Analizi Tablosu

Değişim Kaynakları	Sd	Kareler Toplamı	Varyans	F
Gruplar arası	c-1	SSA	$MSA=SSA/c-1$	$F= MSA/MSW$
Grup içi	N-c	SSW	$MSW=SSW/N-c$	
Toplam	N-1	$SST=SSA+SSW$		

- Hesapla bulunan F değeri, tablodan elde edilen c-1 pay ve N-c payda serbestlik dereceli F kritik değerinden büyükse H_0 hipotezi reddedilir.

Örnek

- ▶ 4 farklı şehirde, 2'şer kişiden oluşan 25'er aileden 6 ay boyunca sarf ettikleri elektriğin toplam maliyeti hakkında veri toplanıyor. Farklı şehirlerdeki aynı büyüklükteki bu ailelerin ortalama elektrik tüketimleri arasında anlamlı bir farklılık olup olmadığını anlamak istiyoruz.
- ▶ $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
- ▶ H_A : En az iki şehirdeki ortalama elektrik tüketimi birbirinden farklıdır

1

[illegible]

Variable View



1 : sehir 1

	sehir	maliyet
1	Adana	54
2	Adana	47
3	Adana	44
4	Adana	57
5	Adana	46
6	Adana	38
7	Adana	45
8	Adana	57
9	Adana	52
10	Adana	47
11	Adana	53
12	Adana	58
13	Adana	46
14	Adana	62
15	Adana	42
16	Adana	48
17	Adana	46
18	Adana	50
19	Adana	50

- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Quality Control
- ROC Curve...



Visible: 2 of 2 Variables

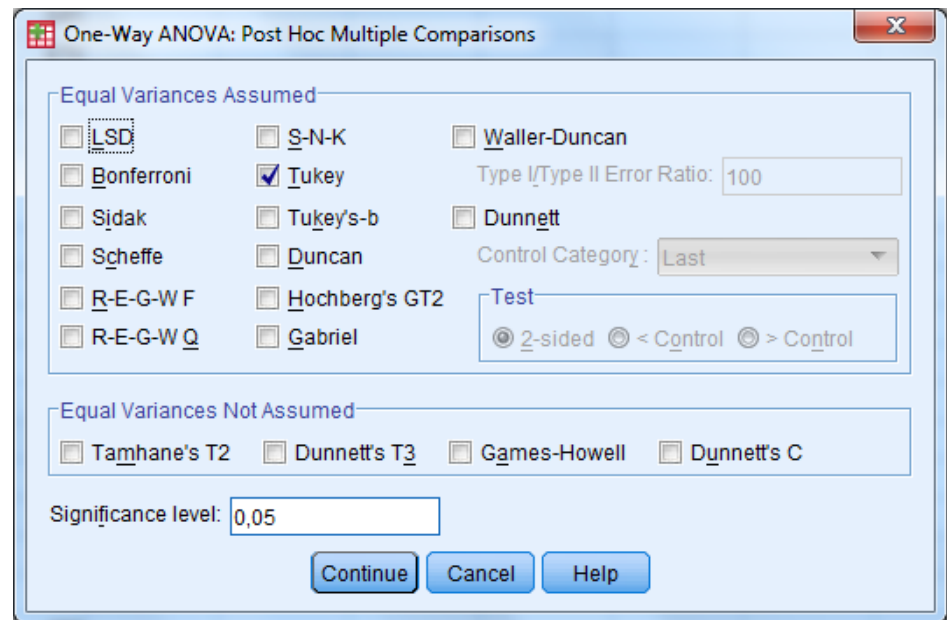
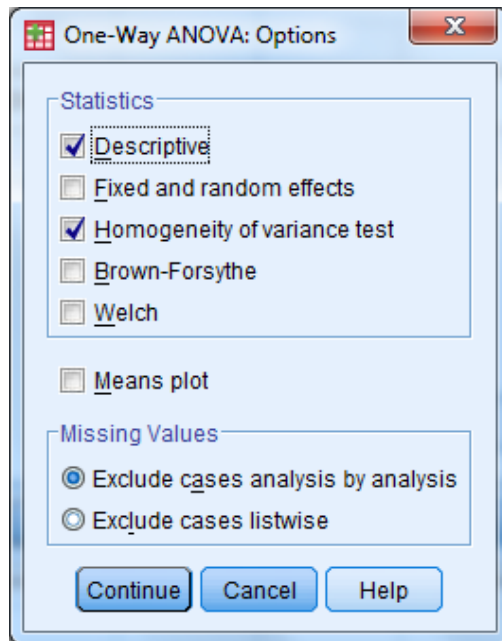
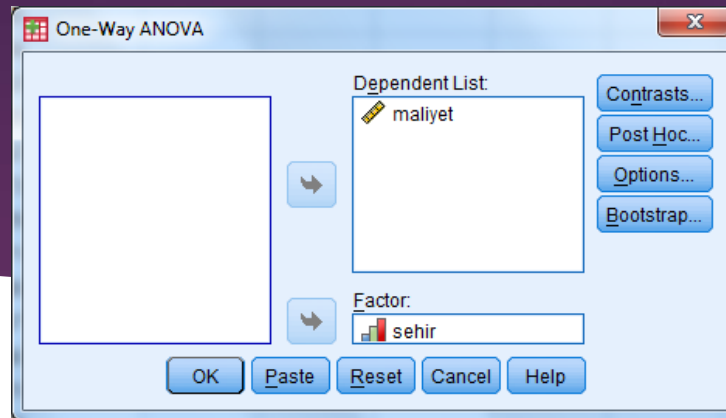
- Means...
- One-Sample T Test...
- Independent-Samples T Test...
- Paired-Samples T Test...
- One-Way ANOVA...

Data View

Variable View

One-Way ANOVA...

PASW Statistics Processor is ready



Descriptives

maliyet

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Adana	25	497,28	56,628	11,326	473,91	520,65
Gaziantep	25	515,84	56,529	11,306	492,51	539,17
Ankara	25	531,20	63,976	12,795	504,79	557,61
İstanbul	25	555,12	72,576	14,515	525,16	585,08
Total	100	524,86	65,385	6,539	511,89	537,83

Test of Homogeneity of Variances

maliyet

Levene Statistic	df1	df2	Sig.
,817	3	96	,488

ANOVA

maliyet

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	44947,000	3	14982,333	3,802	,013
Within Groups	378299,040	96	3940,615		
Total	423246,040	99			

Multiple Comparisons

maliyet

Tukey HSD

(I) şehir	(J) şehir	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Adana	Gaziantep	-18,560	17,755	,723	-64,98	27,86
	Ankara	-33,920	17,755	,231	-80,34	12,50
	İstanbul	-57,840*	17,755	,008	-104,26	-11,42
Gaziantep	Adana	18,560	17,755	,723	-27,86	64,98
	Ankara	-15,360	17,755	,823	-61,78	31,06
	İstanbul	-39,280	17,755	,127	-85,70	7,14
Ankara	Adana	33,920	17,755	,231	-12,50	80,34
	Gaziantep	15,360	17,755	,823	-31,06	61,78
	İstanbul	-23,920	17,755	,535	-70,34	22,50
İstanbul	Adana	57,840*	17,755	,008	11,42	104,26
	Gaziantep	39,280	17,755	,127	-7,14	85,70
	Ankara	23,920	17,755	,535	-22,50	70,34

*. The mean difference is significant at the 0.05 level.

Örnek: 2 Yönlü ANOVA

- Bir oyuncak üreticisi firma, ürettiği oyuncakların hangi tür mağazalarda daha çok satıldığını merak etmektedir. Firma, ürünlerinin farklı mağaza türlerindeki satışlarını karşılaştırmak istemektedir. 3 tür mağaza (süper market, oyuncak mağazası, alışveriş merkezi mağazası) vardır ve bu mağazalar ayrıca şehir merkezinde ve şehir dışında olmak üzere ayrıca sınıflandırılmaktadır. Bu bilgiler ışığında;
- 1. Mağazanın türü oyuncak satışlarını etkiliyor mu?
- 2. Mağazanın yeri oyuncak satışlarını etkiliyor mu?
- 3. Oyuncak satışlarına mağazanın etkisi mağazanın yerine göre değişiyor mu?

*ANOVA 2-Oyuncak.sav [DataSet1] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

4 : satış 7

Visible: 3 of 3 Variables

	yer		ar	var	var	var	var	var	var
1	şehir merkezi								
2	şehir merkezi								
3	şehir merkezi								
4	şehir merkezi								
5	şehir merkezi	Al							
6	şehir merkezi	Al							
7	şehir merkezi	Al							
8	şehir merkezi	Al							
9	şehir merkezi	Oyu							
10	şehir merkezi	Oyu							
11	şehir merkezi	Oyu							
12	şehir merkezi	Oyu							
13	şehir dışı								
14	şehir dışı								
15	şehir dışı								
16	şehir dışı								
17	şehir dışı	Al							
18	şehir dışı	Al							
19	şehir dışı	Al							

1

Data View Variable View

Univariate...

PASW Statistics Processor is ready

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control
ROC Curve...

Univariate...
Multivariate...
Repeated Measures...
Variance Components...

Univariate

Dependent Variable:

Fixed Factor(s):

Random Factor(s):

Covariate(s):

WLS Weight:

Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
Bootstrap...

OK Paste Reset Cancel Help

Univariate: Options

Estimated Marginal Means

Factor(s) and Factor Interactions:

Display Means for:

☐ Compare main effects

Confidence interval adjustment:

Display

☒ Descriptive statistics
☒ Estimates of effect size
☒ Observed power
☐ Parameter estimates
☐ Contrast coefficient matrix

☒ Homogeneity tests
☐ Spread vs. level plot
☐ Residual plot
☐ Lack of fit
☐ General estimable function

Significance level: Confidence intervals are 95,0 %

Continue Cancel Help

Between-Subjects Factors				
		Value Label	N	
yer	1	şehir merkezi	12	
	2	şehir dışı	12	
mağaza	1	Süper market	8	
	2	Alışveriş merkezi	8	
	3	Oyuncak mağazası	8	

Levene's Test of Equality of Error Variances ^a				
Dependent Variable: satış				
F	df1	df2	Sig.	
1,000	5	18	,446	

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + yer + mağaza + yer * mağaza

Descriptive Statistics

Dependent Variable: satış

yer	mağaza	Mean	Std. Deviation	N
şehir merkezi	Süper market	3,00	3,162	4
	Alışveriş merkezi	10,00	4,761	4
	Oyuncak mağazası	14,00	3,916	4
	Total	9,00	5,970	12
şehir dışı	Süper market	11,00	3,916	4
	Alışveriş merkezi	12,00	5,477	4
	Oyuncak mağazası	10,00	4,082	4
	Total	11,00	4,200	12
Total	Süper market	7,00	5,398	8
	Alışveriş merkezi	11,00	4,870	8
	Oyuncak mağazası	12,00	4,276	8
	Total	10,00	5,150	24

Tests of Between-Subjects Effects

Dependent Variable: satış

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	280,000 ^a	5	56,000	3,055	,036	,459
Intercept	2400,000	1	2400,000	130,909	,000	,879
yer	24,000	1	24,000	1,309	,268	,068
mağaza	112,000	2	56,000	3,055	,072	,253
yer * mağaza	144,000	2	72,000	3,927	,038	,304
Error	330,000	18	18,333			
Total	3010,000	24				
Corrected Total	610,000	23				

a. R Squared = ,459 (Adjusted R Squared = ,309)

