

Cramér's V

İki kategorik(nominal) değişken arasındaki ilişkinin gücünü ölçmek için kullanılan bir istatistiksel ölçüdür. φ_c ile belirtilir.

0 ile 1 arasında değerler alır.

- 0, iki değişken arasında ilişki olmadığını gösterir.
- 1, iki değişken arasında güçlü bir ilişki olduğunu gösterir

Aşağıdaki gibi hesaplanır:

$$V_c = \sqrt{\frac{\varphi^2}{\min(k-1, r-1)}} = \sqrt{\frac{X^2/n}{\min(k-1, r-1)}}$$

Φ : Phi katsayısı

X^2 : Pearson'un ki-kare testinden türetilmiştir.

N: toplam örneklem büyüklüğü

k: sütun sayısı

r: satır sayısı

Varsayımlar:

- 1.İki değişken de kategorik olmalıdır.
2. Kategori başına iki veya daha fazla benzersiz(unique) değer olmalıdır.

Cramér's V, popülasyon karşılığı için oldukça yanlış bir tahmin edici olabilir ve ilişkinin gücünü abartma eğilimi vardır. Bunun için bir yanlışlık düzeltmesi yapılabilir.

Örnek:

Sigara kullanma durumunun felç geçirme ile ilişkili olup olmadığını inceleyelim.

Stroke Prediction Dataset

smoking_status: "formerly smoked", "never smoked", "smokes"

stroke: 1 if the patient had a stroke or 0 if not

Analyze > Descriptive Statistics > Crosstabs

>> Statistics > Nominal: Phi and Cramer's V

smoking_status * stroke Crosstabulation

Count

		stroke		
		0	1	Total
smoking_status	formerly smoked	815	70	885
	never smoked	1802	90	1892
	smokes	747	42	789
Total		3364	202	3566

Yandaki tablodan verinin nasıl dağıldığını görebiliyoruz.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,057	,003
	Cramer's V	,057	,003
N of Valid Cases		3566	

Cramer's V için $p = 0.003$

Degrees of freedom	Small	Medium	Large
1	0.10	0.30	0.50
2	0.07	0.21	0.35
3	0.06	0.17	0.29
4	0.05	0.15	0.25
5	0.04	0.13	0.22

df = 1 için:

$p < 0.10$ olduğundan iki değişken arasında bir ilişki olmadığını söyleyebiliriz.

(Formülden de görülebileceği gibi serbestlik derecesi $\min(k-1, r-1)$ 'dir.)

+ Tetrakorik Korelasyon(Tetrachoric Correlation):

İkili(binary) kategorik değişkenler arasındaki korelasyonu hesaplamak için kullanılır.

+ Polikorik Korelasyon(Polychoric Correlation):

Sıralı(ordinal) kategorik değişkenler arasındaki korelasyonu hesaplamak için kullanılır.

ETA Correlation Ratio (ETA Korelasyon Oranı)

Sadece Correlation Ratio da denebilir. Varyans analizinde (ANOVA) yaygın olarak kullanılan kareler toplamı kavramına dayanır. η ile belirtilir.

Korelasyon oranının karesi, eta kare (η^2) olarak bilinir. Gruplar arası kareler toplamının, toplam kareler toplamına bölünmesiyle hesaplanır.

$$\eta^2 = \frac{SS_{between}}{SS_{total}}$$

İkili olmayan kategorik(nominal) bir değişken ve nicel bir değişken arasındaki ilişkinin gücünü ölçmek için de kullanılabilir.

Örnek:

Öğrencinin bulunduğu etnik grubun, okuma sınavında aldığı notla ilişkili olup olmadığını inceleyelim.

Students Performance in Exams

EthnicGroup: Ethnic group of the student (group A to E)

(Veri sentetik olduğu için etnik grupların bir karşılığı yok.)

ReadingScore: reading test score(0-100)

Analyze > Compare Means > Means >

>> Options > Statistics for First Layer: Anova table and eta

Report

ReadingScore					
EthnicGroup	Mean	N	Std. Deviation	Minimum	Maximum
group A	64,67	89	15,544	23	100
group B	67,35	190	15,177	24	97
group C	69,10	319	13,997	17	100
group D	70,03	262	13,895	31	100
group E	73,03	140	14,874	26	100

Yandaki tablodan verinin nasıl dağıldığını görebiliyoruz.

ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
ReadingScore * EthnicGroup	Between Groups (Combined)	4706,287	4	1176,572	5,622	,000
	Within Groups	208246,152	995	209,293		
	Total	212952,439	999			

Gruplar arası, her grup ortalamasının genel ortalamadan sapması ile bilgi vermektedir.

Gruplar arası varyasyonların, grup içi varyasyonlara kıyasla daha büyük olması, F değerinin daha büyük olmasına neden olur.

H0: Tüm grupların ortalamaları eşittir.

$p < \alpha = 0.05$, H0 kabul edilmez.

Measures of Association

	Eta	Eta Squared
ReadingScore * EthnicGroup	,149	,022

$0.01 < \text{Eta Squared} < 0.06$

İki değişken arasındaki ilişkinin küçük olduğunu söyleyebiliriz.

Eta kare değeri 0 ile 1 arasında değişir. 1'e yakın değerler, modeldeki belirli bir değişkenle açıklanabilen varyansın oranının yüksek olduğunu gösterir.

Eta kare değeri pratik olarak aşağıdaki gibi yorumlanabilir:

• 0.01: Küçük etki büyüklüğü, • 0.06: Orta etki büyüklüğü, • 0.14 veya üzeri: Büyük etki büyüklüğü

+ Sınıf İçi Korelasyon Katsayısı(Intraclass Correlation Coefficient, ICC):

Eta yalnızca sayısal değişken açısından gruplar arasındaki farkı ölçer. ICC aynı zamanda gruplar içindeki sayısal değerler arasındaki uyumu da ölçer.