



**Mimar Sinan Güzel Sanatlar Üniversitesi**

**Fen-Edebiyat Fakültesi**

**İstatistik Bölümü**

**Risk Analizi Projesi**

**H. Simay Özgül – 20201101061**

**Ataman Önel Ük – 20201101058**

## İÇİNDEKİLER

- Veri Setinin İncelenmesi
- Bir Çalışan Neden İşten Ayrılmak İster?
- Cinsiyete Göre İşten Ayrılma Oranları
- Histogramlar
- Regresyon Analizi
- Sonuç
- Kaynakça

## VERİ SETİNİN İNCELENMESİ

Bu proje kapsamında üzerinde duracağımız veri setinin linki aşağıda verilmiş olup, Julia programı ile detaylı bir şekilde incelemesi yapılmıştır.

- <https://www.kaggle.com/davinwijaya/employee-turnover>

Veri seti, bir çalışanın işten ayrılma riskinin hangi kriterler tarafından ne kadar etkilenebileceğini ifade etmektedir. Bu analiz için elimizde 16 değişken ve 1129 çalışan mevcut olup, aşağıda veri setinin bir çıktısı verilmiştir.

```
using CSV
using DataFrames
data = CSV.read("turnover.csv", DataFrame)
```

1129x16 DataFrame														1104 rows omitted	
Row	stag	event	gender	age	Industry	profession	traffic	coach	head_gender	greywage	way	extraversion	Independ	selfcontrol	
	Float64	Int64	String1	Float64	String15	String31	String15	String7	String1	String7	String7	Float64	Float64	Float64	
1	7.0308	1	m	35.0	Banks	HR	rabrecNERab	no	f	white	bus	6.2	4.1	5.7	
2	22.9651	1	m	33.0	Banks	HR	empjs	no	m	white	bus	6.2	4.1	5.7	
3	15.9343	1	f	35.0	PowerGeneration	HR	rabrecNERab	no	m	white	bus	6.2	6.2	2.6	
4	15.9343	1	f	35.0	PowerGeneration	HR	rabrecNERab	no	m	white	bus	5.4	7.6	4.9	
5	8.41068	1	m	32.0	Retail	Commercial	youjs	yes	f	white	bus	3.0	4.1	8.0	
6	8.9692	1	f	42.0	manufacture	HR	empjs	yes	m	white	bus	6.2	6.2	4.1	
7	8.9692	1	f	42.0	manufacture	HR	empjs	yes	m	white	bus	6.2	6.2	4.1	
8	120.444	1	f	28.0	Retail	HR	referral	no	m	white	bus	3.8	5.5	8.0	
9	8.6078	1	f	29.0	Banks	HR	empjs	no	f	white	bus	8.6	6.9	2.6	
10	4.43532	1	f	30.0	Consult	Marketing	youjs	yes	m	white	bus	5.4	5.5	3.3	
11	67.3511	1	f	40.0	manufacture	HR	rabrecNERab	no	m	white	car	8.6	4.1	1.8	
12	20.9261	1	m	23.0	Banks	HR	referral	yes	m	white	bus	3.0	6.9	4.9	
13	20.9261	1	f	22.0	Consult	HR	rabrecNERab	my head	f	white	foot	5.4	8.4	4.9	
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1118	11.0062	1	m	41.0	Consult	HR	youjs	no	m	white	bus	3.8	4.1	6.4	
1119	6.80082	0	m	25.0	manufacture	HR	friends	no	f	white	bus	5.4	4.8	6.4	
1120	26.6776	0	f	19.0	IT	HR	youjs	yes	f	white	bus	7.0	2.6	6.4	
1121	11.5647	0	m	41.0	IT	IT	youjs	no	m	white	foot	5.4	4.1	4.9	
1122	53.3552	0	f	27.0	IT	HR	youjs	no	m	white	bus	3.0	5.5	6.4	
1123	12.5175	0	m	48.0	manufacture	Marketing	friends	no	m	white	bus	1.4	6.9	6.4	
1124	6.6694	0	f	27.0	Banks	HR	youjs	no	f	white	bus	9.4	5.5	3.3	
1125	10.6119	0	f	41.0	Banks	HR	rabrecNERab	my head	m	white	bus	8.6	3.4	2.6	
1126	10.6119	0	f	41.0	Banks	HR	rabrecNERab	my head	m	white	bus	8.6	3.4	2.6	
1127	118.801	0	f	34.0	Telecom	Accounting	KA	no	f	white	bus	4.6	5.5	7.2	
1128	49.4127	0	f	51.0	Consult	HR	empjs	no	m	grey	bus	3.8	7.6	5.7	
1129	24.8378	0	f	29.0	Retail	HR	youjs	no	f	white	car	9.4	1.2	4.1	

Veri setimizde bulunan değişkenleri sırasıyla inceleyelim;

- **STAG:** Çalışma süresi
- **EVENT:** Çalışanın işten ayrılıp ayrılmama durumu (0= hayır, 1= evet)
- **GENDER:** Çalışanın cinsiyeti
- **AGE:** Çalışanın yaşı
- **INDUSTRY:** Kişinin çalıştığı endüstri
- **PROFESSION:** Kişinin çalıştığı departman
- **TRAFFIC:** Kişinin şirketi nereden bulduğu

- **COACH:** Staj / deneme sürecinde çalışanın eğitmeni olup olmaması
- **HEAD GENDER:** Yöneticinin cinsiyeti
- **GREYWAGE:** Alınan maaş türü (white= asgari ücret , grey= diğeri)
- **WAY:** İşe giderken kullanılan ulaşım türü
- **EXTRAVERSION:** Çalışanın dışadönüklük skoru
- **INDEPEND:** Çalışanın özgürlük / rahatlık skoru
- **SELFCONTROL:** Çalışanın otokontrol skoru
- **ANXIETY:** Çalışanın anksiyete skoru
- **NOVATOR:** Çalışanın üreticiliğini gösteren skor

Not: Yukarıda verilen değişkenlere ilişkin çalışan skorları 1 ile 10 arasında değer almaktadır.

*Yapılan incelemeler sonucu veri setinde herhangi bir kayıp gözleme rastlanmamış ve dağılım incelenmiş olup, veri setinin normal dağıldığı görülmüştür.*

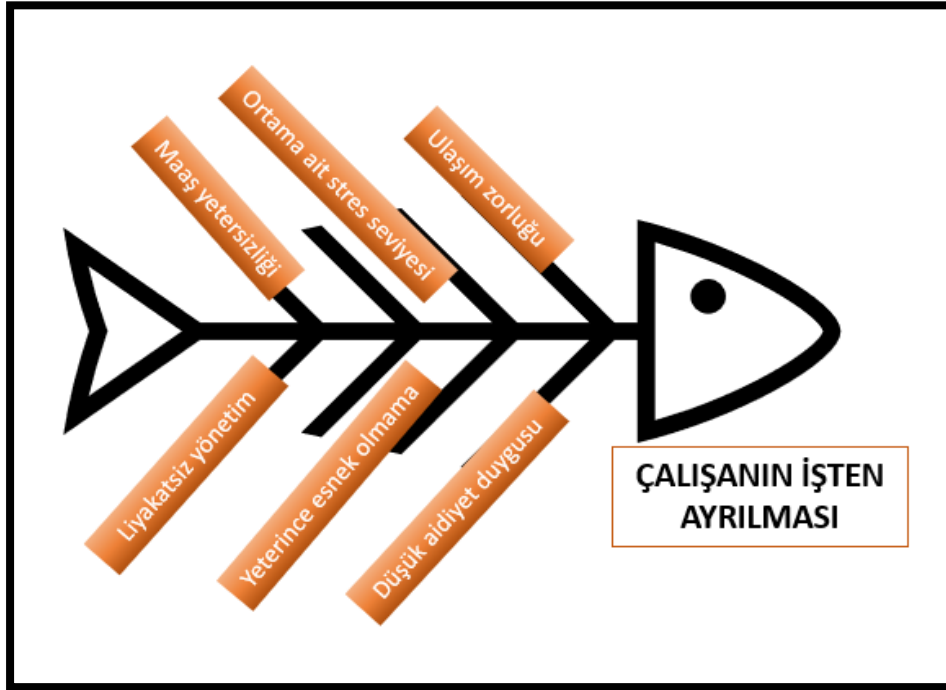
16x7 DataFrame							
Row	variable	mean	min	median	max	nmissing	eltype
	Symbol	Union...	Any	Union...	Any	Int64	DataType
1	stag	36.6275	0.394251	24.345	179.45	0	Float64
2	event	0.505757	0	1.0	1	0	Int64
3	gender		f		m	0	String1
4	age	31.067	18.0	30.0	58.0	0	Float64
5	industry		HoReCa		transport	0	String15
6	profession		Accounting		manage	0	String31
7	traffic		KA		youjs	0	String15
8	coach		my head		yes	0	String7
9	head_gender		f		m	0	String1
10	greywage		grey		white	0	String7
11	way		bus		foot	0	String7
12	extraversion	5.59238	1.0	5.4	10.0	0	Float64
13	independ	5.47803	1.0	5.5	10.0	0	Float64
14	selfcontrol	5.59725	1.0	5.7	10.0	0	Float64
15	anxiety	5.66563	1.7	5.6	10.0	0	Float64
16	novator	5.87963	1.0	6.0	10.0	0	Float64

Birkaç değişkeni buradan yorumlayacak olursak;

- Veri setindeki çalışanların yaşının ortalama 31 olduğu söylenebilir.
- Çalışanların kişisel özelliklerine ilişkin skorları veren “extraversion”, “independ”, “selfcontrol”, “anxiety” ve “novator” değişkenlerinin hemen hemen benzer ortalamalara sahip olduğu görülmektedir.
- Çalışanların büyük çoğunluğu asgari ücret kazanmaktadır. (white)

### **BİR ÇALIŞAN NEDEN İŞTEN AYRILMAK İSTER?**

İnceleyeceğimiz konu, bir çalışanın işten neden ayrılabilirliğine ilişkin analiz yapmak olacağından Sebep Sonuç Analizi Metodu'ndan (FISH BONE) yararlanacağız.



Yukarıdaki şemada görüldüğü üzere bir çalışanın işten ayrılma nedenleri kabaca maaş yetersizliği, ortama ait stres seviyesi, işe giderken kullanılan ulaşım araçlarında yaşanabilecek zorluklar, iş yerinde liyakatsiz yönetim, çalışan için yeterince esnek davranamama durumu ve düşük aidiyet duygusu olabilir.

Bizim veri setimizde yukarıdaki maddeler ile paralel değişkenler olan anksiyete, çalışanın kendini kontrol edebilme durumu, sosyalliğinde artış azalış gibi durumları temsil eden skorlara ilişkin değişkenleri incelememiz mantıklı olacaktır.

Merak ettiğimiz bir diğer konu ise veri setinde cinsiyet rolünün bu noktada ne kadar etkili olduğudur. Dolayısıyla öncesinde "gender" ve "event" değişkenlerine ilişkin bir crosstabs analizi yapıp, sonrasında birtakım tablo incelemeleri ve regresyon analizi ile konuya ilişkin bir hipotez kurulup ele aldığımız konu test edilecektir.

### CİNSİYETE GÖRE İŞTEN AYRILMA ORANLARI

Cinsiyete bağlı olarak işten ayrılma ve ayrılmama durumları incelenmiş olup aşağıda tablo çıktısı rapora eklenmiş ve yorumlanmıştır.

gender * event Crosstabulation					
			event		Total
			hayır	evet	
gender	f	Count	417	436	853
		% within event	74,7%	76,4%	75,6%
	m	Count	141	135	276
		% within event	25,3%	23,6%	24,4%
Total	Count	558	571	1129	
	% within event	100,0%	100,0%	100,0%	

Tablolara bakacak olursak “Risk Estimate” tablosunda bulunan Odds Ratio değeri 0.916 olarak hesaplanmıştır. Bu değer 1’e çok yakın olması nedeniyle araştırdığımız durumun cinsiyet etkeni ile azaltıcı veya artırıcı bir etkisinin olduğunu söyleyemeyiz.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,404 <sup>a</sup>	1	,525		
Continuity Correction <sup>b</sup>	,321	1	,571		
Likelihood Ratio	,404	1	,525		
Fisher's Exact Test				,534	,286
N of Valid Cases	1129				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 136,41.  
b. Computed only for a 2x2 table

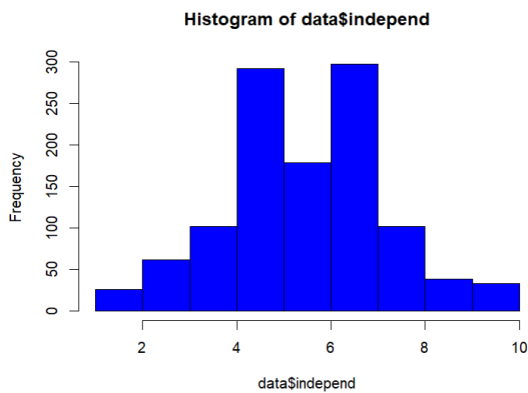
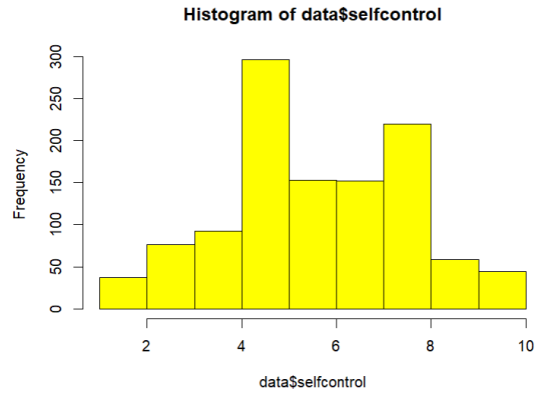
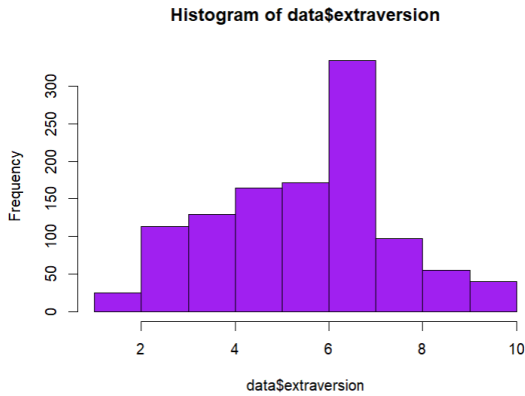
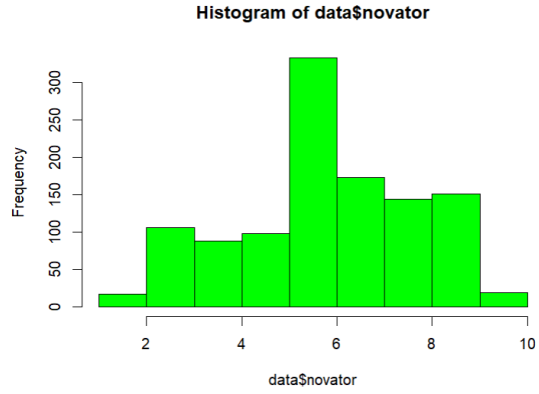
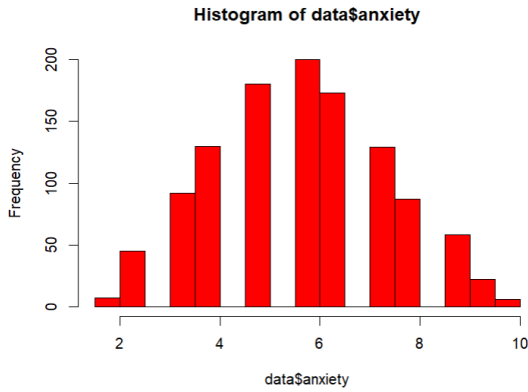
  

Risk Estimate			
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for gender (f / m)	,916	,698	1,201
For cohort event = hayır	,957	,837	1,094
For cohort event = evet	1,045	,911	1,199
N of Valid Cases	1129		

Bunun dışında Odds değerimizin %95 güven aralığı sınırları içinde olduğunu söyleyebilmekteyiz. İşten ayrılmama durumu için odds oranı 0.957 iken, işten ayrılma durumuna ait odds oranı 1.045’tir.

## HİSTOGRAMLAR

```
1 data <- read.table(file="/Users/w10/OneDrive/Desktop/turnover.csv", header= TRUE, sep=",")
2 hist(data$extraversion, col= "purple")
3 hist(data$anxiety, col= "red")
4 hist(data$independ, col= "blue")
5 hist(data$selfcontrol, col= "yellow")
6 hist(data$novator, col= "green")
7
```



Skorlara ilişkin değişkenlerin histogram grafiklerini incelediğimizde farklı dağılımlı grafikler elde etmiş bulunmaktayız. Sırasıyla “anxiety” ve “novator” değişkenleri normal dağılımlı bir histogram grafiği verirken, “extraversion” sola çarpık bir histogram grafiği vermektedir.

“selfcontrol” ve “independ” değişkenlerine ait son iki grafik ise bize bimodal histogram grafiğini vermektedir. İki uçlu histogram, iki belirgin mod (zirve) içeren bir dağılımı ifade eder. Bu durum, veri setinde iki farklı grup veya alt grup olduğunu gösterebilir.

		Correlations				
		extraversion	independ	selfcontrol	anxiety	novator
extraversion	Pearson Correlation	1	-,200**	-,538**	-,135**	,297**
	Sig. (2-tailed)		,000	,000	,000	,000
	N	1129	1129	1129	1129	1129
independ	Pearson Correlation	-,200**	1	-,166**	-,427**	,024
	Sig. (2-tailed)	,000		,000	,000	,423
	N	1129	1129	1129	1129	1129
selfcontrol	Pearson Correlation	-,538**	-,166**	1	-,108**	-,566**
	Sig. (2-tailed)	,000	,000		,000	,000
	N	1129	1129	1129	1129	1129
anxiety	Pearson Correlation	-,135**	-,427**	-,108**	1	,247**
	Sig. (2-tailed)	,000	,000	,000		,000
	N	1129	1129	1129	1129	1129
novator	Pearson Correlation	,297**	,024	-,566**	,247**	1
	Sig. (2-tailed)	,000	,423	,000	,000	
	N	1129	1129	1129	1129	1129

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Yukarıdaki tabloda skor değişkenlerinin birbirleri ile ilgili lineer ilişkisi gözlemlenmek istenmiştir.

“Pearson Correlation” için katsayılar kısaca bize şu bilgiyi verir; yaklaşık olarak değerin 1 olması güçlü pozitif bir lineer ilişki olduğunu söylerken, – 1 olması ise güçlü negatif bir ilişki olduğunu belirtir. Benzer durumda katsayı yaklaşık 0 olur ise herhangi bir lineer ilişkiden söz edemeyiz.

Bunun için “Pearson Correlation” değerlerine bakıldığında değerlerin sırasıyla “extraversion” ile “independ” , “anxiety” ve “novator” ile arasında lineer bir ilişki olmadığını ancak “selfcontrol” ile negatif yönlü bir ilişkisi olduğunu söyleyebiliriz.

“Sig.” değerlerinin 0.05’ten küçük çıkması testin anlamlı olduğunu söylemektedir. Ancak “novator” ile “independ” için aynısını söyleyememekteyiz.



### REGRESYON ANALİZİ:

H<sub>0</sub>: anxiety, extraversion, independ, selfcontrol, novator faktörlerinin işten ayrılma ile ilişkisi vardır.

H<sub>1</sub>: anxiety, extraversion, independ, selfcontrol, novator faktörlerinin işten ayrılma ile ilişkisi yoktur.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,083 <sup>a</sup>	,007	,002	,500

a. Predictors: (Constant), novator, independ, extraversion, anxiety, selfcontrol

Model Summary tablosuna baktığımızda R değerimizin 0.083 olduğunu görmekteyiz. Bu da bağımlı değişken ile bağımsız değişkenler arası ilişkinin % 0,8'lik kısmını açıklamakta olup, beklenilenden düşük bir değer vermektedir.

ANOVA <sup>a</sup>						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,922	5	,384	1,540	,175 <sup>b</sup>
	Residual	280,291	1123	,250		
	Total	282,213	1128			

a. Dependent Variable: event  
b. Predictors: (Constant), novator, independ, extraversion, anxiety, selfcontrol

Büyük bir F değeri, gruplar arasında anlamlı bir fark olduğunu gösterir. “Sig.” olarak gösterilen p-value değerimiz 0.05’ten büyük olduğu için ele alınan gruplar arasında anlamlı bir fark yoktur diyebiliriz.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,738	,211		3,494	,000
	anxiety	-,020	,011	-,069	-1,777	,076
	extraversion	-,007	,012	-,024	-,571	,568
	independ	,002	,012	,007	,187	,852
	selfcontrol	-,016	,012	-,062	-1,350	,177
	novator	-,001	,010	-,004	-,106	,916

a. Dependent Variable: event

$$Y = X_1 + X_2 + X_3 + X_4 + X_5 + \Sigma$$

Y: event X<sub>1</sub>: anxiety X<sub>2</sub>: extraversion X<sub>3</sub>: independ X<sub>4</sub>: selfcontrol X<sub>5</sub>: novator Σ: hata terimi

Coefficients tablosundan hareketle yukarıda yazdığımız denklemi düzenleyecek olursak son durum;

$$Y = -0.02x_{i1} - 0.007x_{i2} + 0.002x_{i3} - 0.016x_{i4} - 0.001x_{i5} + \Sigma \text{ halini alır.}$$

Yorumlamak istersek, örneğin çalışana ait “anxiety” değerinde bir birimlik artış işinde devam etme olasılığını yaklaşık % 2 azaltmaktadır.

### **Sonuç olarak:**

p-value değerimiz (“Sig.”) 0.05’ten büyük olduğu için H<sub>0</sub> kabul edilir. Yani çalışanların işten ayrılma nedenleri ile ele alınan ilgili faktörlerin ilişkisi vardır diyebiliriz.

İşveren için çalışanın iş ortamında nasıl hissettiği, ne gibi şeylerden memnun veya rahatsız olduğu önemlidir. Bunların karşılanmadığı veya yetersiz olduğu durumlarda çalışanı kaybetme riski artar. Bu da işvereni olumsuz etkiler.

Çalışanları mutlu etmek adına yapılabilecekler birkaç örnek verecek olursak,

- İş ve yaşam dengesine öncelik vermek
- Şeffaf ve dürüst olmak
- Olumlu bir çalışma ortamı sunmak
- İletişimi güçlü tutmak
- Çalışanları ödüllendirmek

şeklinde özetleyebiliriz.

## **Kaynakça**

- <https://www.linkedin.com/pulse/%C3%A7al%C4%B1%C5%9Fanlar%C4%B1n-i%C5%9Ften-ayr%C4%B1lmas%C4%B1n%C4%B1n-en-y%C3%BCKsek-10-nedeni-tu%C4%9Fba-okumu%C5%9F/?originalSubdomain=tr>
- <https://edwvb.blogspot.com/2017/10/employee-turnover-how-to-predict-individual-risks-of-quitting.html>
- <https://www.linkedin.com/pulse/zam-yapmadan-%C3%A7al%C4%B1%C5%9Fanlar%C4%B1-mutlu-etmenin-12-yolu-we-point-apan/?originalSubdomain=tr>