



T.C.

**BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ
FEN FAKÜLTESİ
İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ BÖLÜMÜ**

2023-2024 AKADEMİK YILI BAHAR DÖNEMİ

İstatistiksel Paket Programlama Veri Analizi II

**ÖDEV KONUSU: KONU 7 [(B) 0,01 anlamlılık düzeyinde]
İkiden fazla bağımsız gruba konum parametresi bakımından karşılaştır. Gerekli ise
farklılığın nelerden kaynaklandığını tespit et.**

**Seyfettin Budak
BİLECİK 2024**

Aşağıda Youtube üzerinden alınan yıllara göre (2020-2021-2022-2023) tarihlerinde en çok izlenen videoların izlenme sayılarını 0.01 anlamlılık düzeyinde karşılaştırdım.

(Her Yıla Ait En Çok İzlenen İlk 50 Video)

Faktör Değişkeni : Yıllar, Nitel, Sınıflama, Nominal

Faktör Düzeyleri $\begin{cases} 2020 \text{ Yılı} & (1. \text{ grup}) \\ \vdots \\ 2023 \text{ Yılı} & (4. \text{ grup}) \end{cases}$ gruplar bağımsız $k=4$

Frekans Değişkeni : İzlenme Sayıları, Nicel, Oranlama, Scale

İlk önce verilerin normal olup olmadığına bakalım.

H_0 : Yıllara göre youtube izlenme sayılarının dağılımı normal dağılıma uymaktadır.

H_1 : Yıllara göre youtube izlenme sayılarının dağılımı normal dağılıma uymamaktadır.

Hipotezlerin doğruluğunu test etmek için SPSS programında aşağıdaki adımları sırasıyla uyguladım.

1. SPSSFinalÖdevi2.sav [DataSet0] - IBM SPSS Statistics Data Editor

2. Explore

3. Explore: Statistics

4. Explore: Plots

5. Sonuçlar

Tests of Normality							
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	yıllar	Statistic	df	Sig.	Statistic	df	Sig.
izlenme	2020 Yılı	.340	50	<.001	.492	50	<.001
	2021 Yılı	.338	50	<.001	.556	50	<.001
	2022 Yılı	.267	50	<.001	.660	50	<.001
	2023 Yılı	.217	50	<.001	.792	50	<.001

a. Lilliefors Significance Correction

→ P değerleri

Burada tüm k değerleri (Yıllar) için toplu yorum yapabiliriz.

Tüm K (Yıl) değerleri için $P(0,001) < \alpha(0,01)$ olup H_0 hipotezleri reddedilir.

Yorum

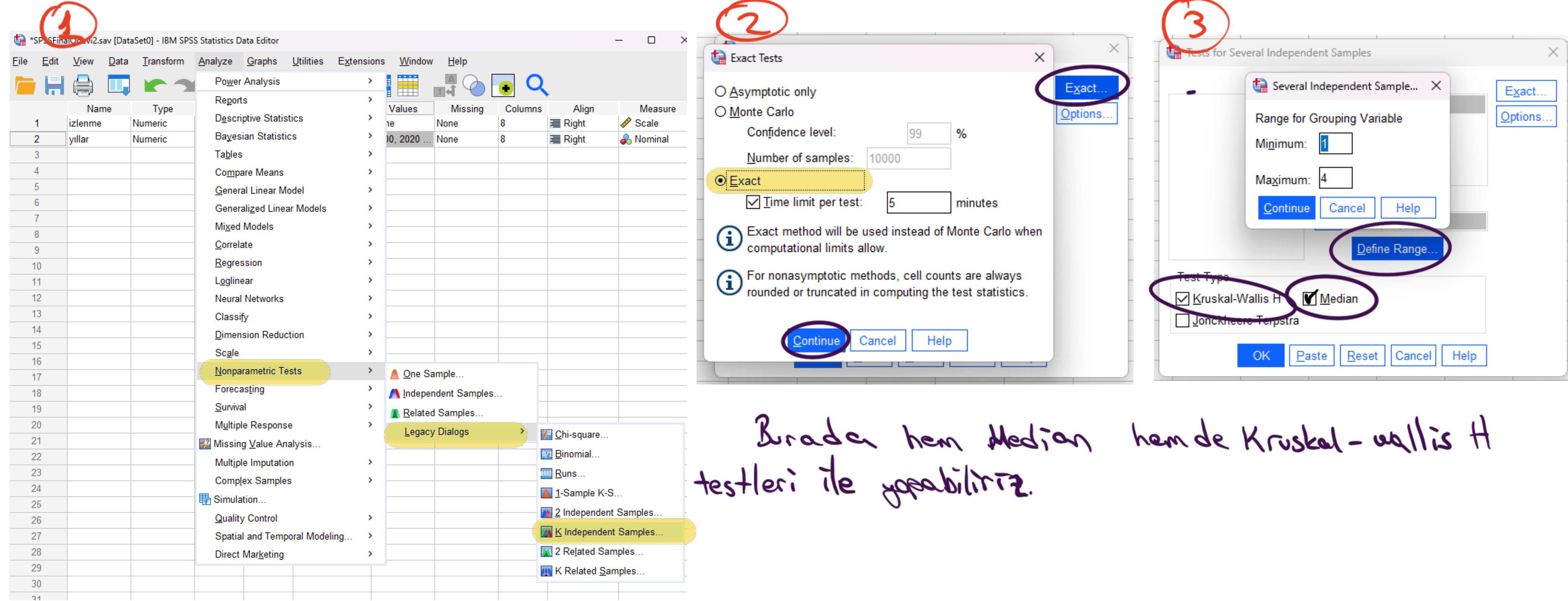
% 99 olasılıkla tüm yıl (K) değerleri için youtubedaki en çok izlenen ilk 50 videonun izlenme sayılarının dağılımları normal dağılıma uymamaktadır.

Bu yüzden parametrik olmayan testlere başvurmanız gerekiyor.

Parametrik olmayan testleri kullandığımız için M (Medyan) değeriyle hipotez kuruyoruz.

$$H_0: M_1 = M_2 = M_3 = M_4 = M$$

H_1 : En az bir M_i diğerlerinden farklı



Burada hem Median hem de Kruskal-Wallis H testleri ile yapabiliriz.

Kruskal-Wallis Test				
Ranks				
	Yıllar	N	Mean Rank	
izlenme	2020 Yılı	50	96.62	
	2021 Yılı	50	71.86	
	2022 Yılı	50	118.20	
	2023 Yılı	50	115.32	
Total		200		

Test Statistics ^{a,b,c}	
	izlenme
Kruskal-Wallis H	20.421
df	3
Asymp. Sig.	<.001

a. Kruskal Wallis Test
b. Grouping Variable: yıllar
c. Some or all exact significances cannot be computed because there is insufficient memory.

Median Test				
Frequencies				
	Yıllar			
	2020 Yılı	2021 Yılı	2022 Yılı	2023 Yılı
izlenme > Median	22	16	30	32
izlenme <= Median	28	34	20	18

Test Statistics ^a	
	izlenme
N	200
Median	9270872.500
Chi-Square	13.120 ^b
df	3
Asymp. Sig.	.004
Exact Sig.	.005
Point Probability	.000

a. Grouping Variable: yıllar
b. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 25.0.

ikisinin de P değerleri < 0.01 olup

Kruskal-Wallis $P(0.001) < \alpha(0.01)$

Median $P(0.005) < \alpha(0.01)$

olup H_0 hipotezleri reddedilmektedir

Yorum

%99 olasılıkla yıllara göre youtube en çok izlenen ilk 50 videonun izlenme sayıları ile ilgili verilerde anlamlı fark vardır.

Kikare tablosu değerleri üzerinden de aynı sonuçları bulabiliriz

$\chi^2_h > \chi^2_T$ olup red bölgesinde olduğundan H_0 reddedilir

$$\chi^2_h = 13.120$$

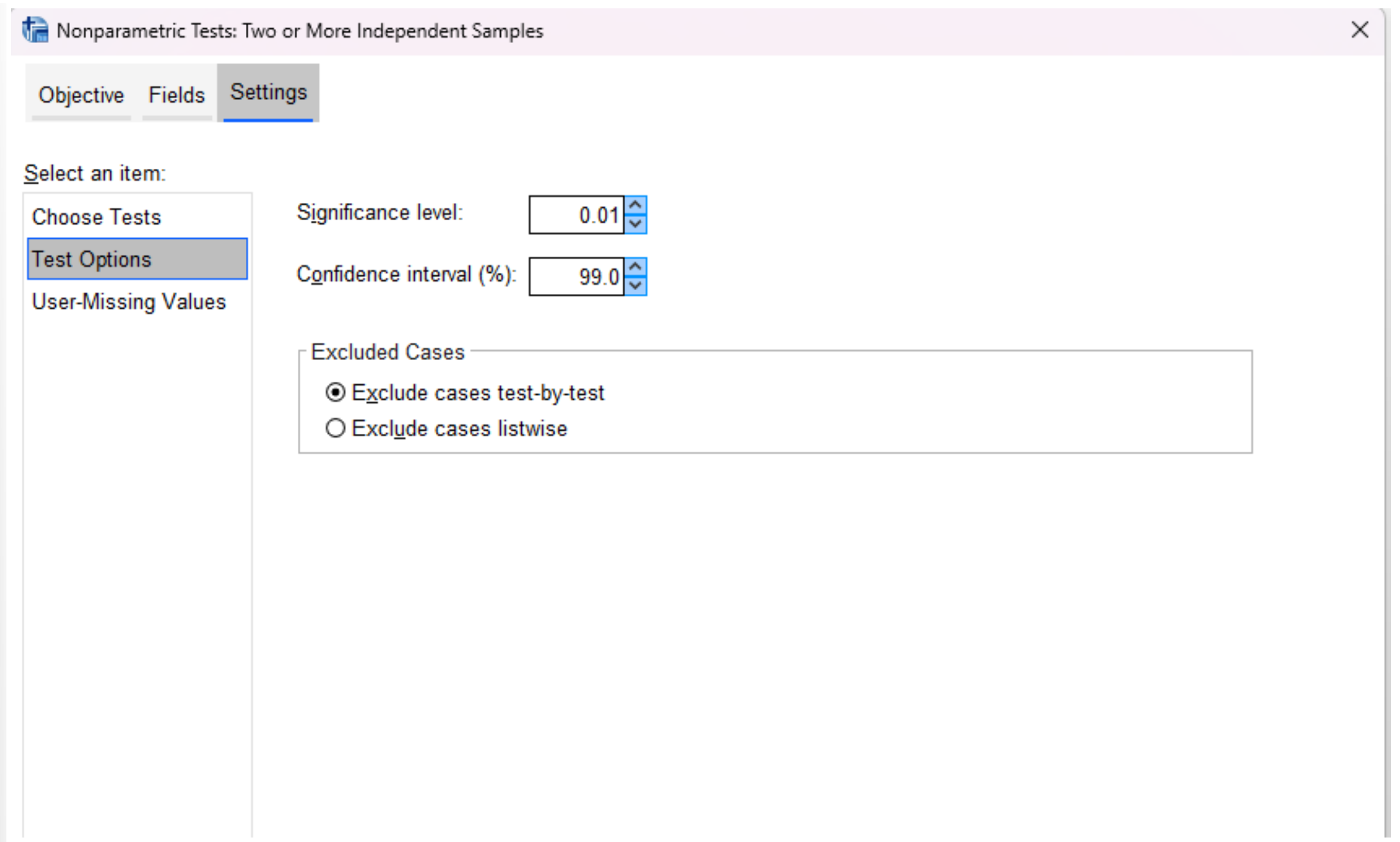
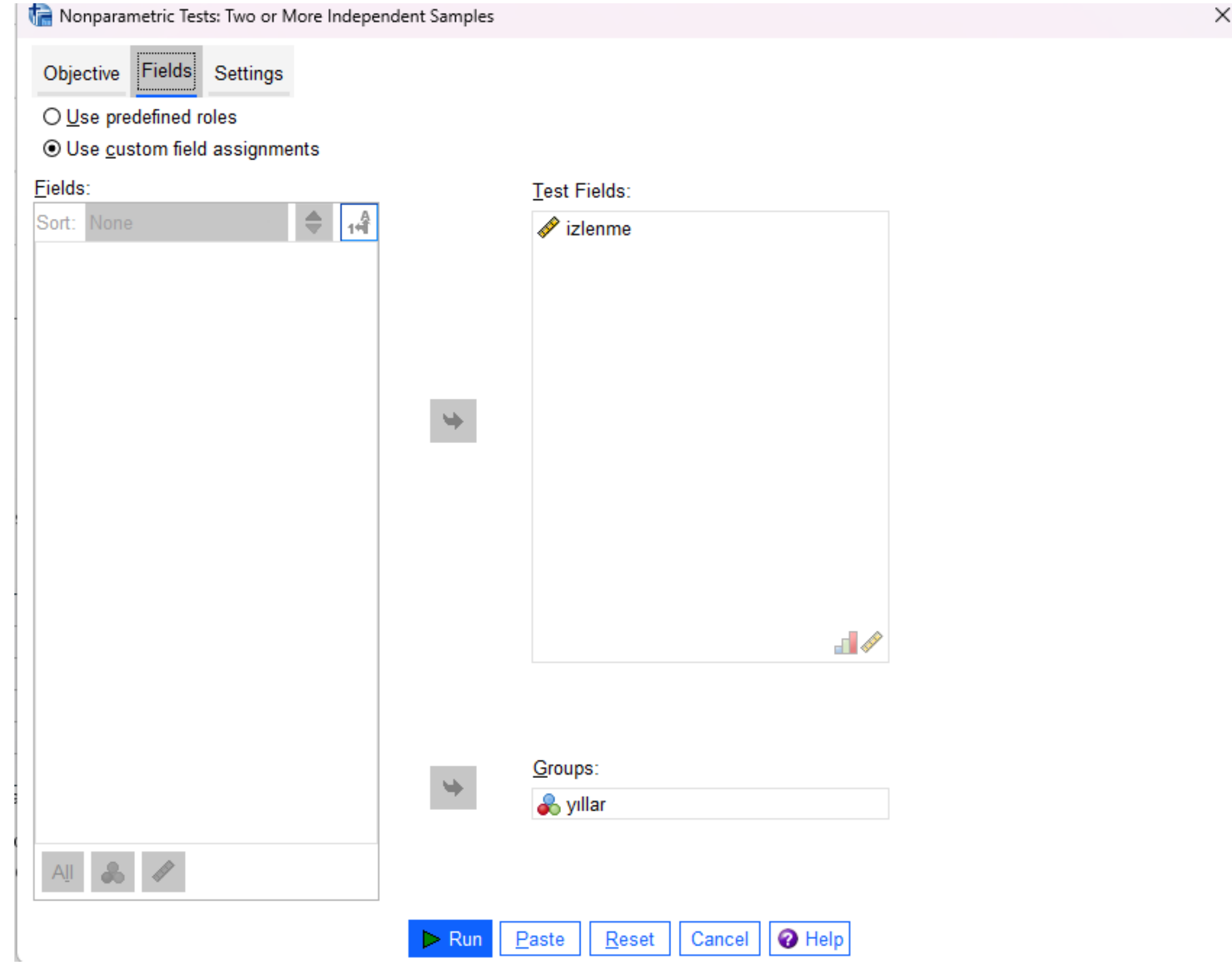
$$\chi^2_T \frac{n-1}{3}, 0.01 = 11.3448$$

— Anlamlı fark oluşturan grubu bulmak için parametrik olmayan bir teste daha başvurabiliriz.

— Mann Whitney U testi bu sayede anlamlı fark hangi gruptan kaynaklanıyor görebiliriz

$H_0: \mu_1 = \mu_2 = \mu$ (K yerine yıl değerleri gelecektir. 2020-2021 gibi)

$H_1: \mu_1 \neq \mu_2$



Pairwise Comparisons of yillar					
Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig. ^a
2021 Yılı-2020 Yılı	24.760	11.576	2.139	.032	.195
2021 Yılı-2023 Yılı	-43.460	11.576	-3.754	<.001	.001
2021 Yılı-2022 Yılı	-46.340	11.576	-4.003	<.001	.000
2020 Yılı-2023 Yılı	-18.700	11.576	-1.615	.106	.637
2020 Yılı-2022 Yılı	-21.580	11.576	-1.864	.062	.374
2023 Yılı-2022 Yılı	2.880	11.576	.249	.804	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .010.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Burada P değerlerine bakarak olursak

2021-2022 ve 2021-2023 yılı değerleri
 $P(0,001)$ $P(0,001)$

$P(0,001) < \alpha(0,01)$ olup H_0 reddedilir

2020-2023, 2020-2023, 2020-2022, 2022-2023 yıllarına ait P değerleri:
 $P(0,32)$ $P(0,106)$ $P(0,062)$ $P(0,804)$

$P_K > \alpha(0,01)$ olup H_0 hipotezleri kabul edilir

Yorum 1

%99 olasılıkla 2021-2022 ve 2021-2023 yıllarındaki youtube'daki en çok izlenen ilk 50 videonun izlenme sayısıyla alakalı verilerde anlamlı fark vardır.

Yorum 2

%99 olasılıkla 2020-2023, 2020-2023, 2020-2022, 2022-2023 yıllarındaki youtube'daki en çok izlenen ilk 50 videonun izlenme sayısıyla alakalı verilerde anlamlı fark yoktur.

Genel Yorum

%99 olasılıkla (2020, 2021, 2022, 2023) yıllarına ait youtube en çok izlenen ilk 50 videolarının izlenme sayıları ile ilgili verilerde anlamlı fark vardır.

Farklılık 2021-2022 ve 2021-2023 yıllarına ait verilerden kaynaklanmaktadır.

2020 YILI YOUTUBE EN ÇOK İZLENEN İLK 50 VİDEO İZLENME SAYILARI

[illegible]

2021 YILI YOUTUBE EN ÇOK İZLENEN İLK 50 VİDEO İZLENME SAYILARI

[illegible]

2022 YILI YOUTUBE EN ÇOK İZLENEN İLK 50 VİDEO İZLENME SAYILARI

[illegible]

2023 YILI YOUTUBE EN ÇOK İZLENEN İLK 50 VİDEO İZLENME SAYILARI

[illegible]

KAYNAK: Verileri Google' un sunmuş olduğu Youtube API üzerinden aşağıda kullandığım kodlarla aldım.

```
JS youtube.js > getMostPopularVideosFromYear > filteredVideos > top50Videos.map() callback
1 // API anahtarınızı buraya girin
2 const apiKey = "AIzaSyB..."; → Herkese özel gizli bir şifre (Youtube API)
3
4 async function getMostPopularVideosFromYear(year) {
5   try {
6     // Arama tarihlerini belirleyin
7     const publishedAfter = `${year}-01-01T00:00:00Z`;
8     const publishedBefore = `${year}-12-31T23:59:59Z`;
9
10    // Belirli bir tarih aralığındaki videoları arayın
11    const searchApiUrl = `https://www.googleapis.com/youtube/v3/search?part=snippet&type=video&publishedAfter=${publishedAfter}&publishedBefore=${publishedBefore}&regionCode=TR&maxResults=50&key=${apiKey}`;
12
13    let response = await fetch(searchApiUrl);
14    if (!response.ok) {
15      throw new Error("Network response was not ok " + response.statusText);
16    }
17
18    let data = await response.json();
19    let videoIds = data.items.map((item) => item.id.videoId);
20
21    // Video detaylarını alın
22    const videosApiUrl = `https://www.googleapis.com/youtube/v3/videos?part=snippet,contentDetails,statistics&id=${videoIds.join(
23      ","
24    )}&key=${apiKey}`;
25    response = await fetch(videosApiUrl);
26    if (!response.ok) {
27      throw new Error("Network response was not ok " + response.statusText);
28    }
29
30    data = await response.json();
31
32    // Videoları izlenme sayılarına göre sırala
33    let sortedVideos = data.items.sort((a, b) => {
34      return (
35        parseInt(b.statistics.viewCount) - parseInt(a.statistics.viewCount)
36      );
37    });
38
39    // En çok izlenen 50 videoyu al
40    let top50Videos = sortedVideos.slice(0, 50);
41
42    // Videoları filtrele ve geri döndür
43    const filteredVideos = top50Videos.map((video) => {
44      // ...
45    });
46
47    return filteredVideos;
48  } catch (error) {
49    console.error("Fetching error:", error);
50  }
51 }
52
53 }
```

```
JS main.js JS twitch.js JS twitter.js {} twitch2022.json
JS main.js > main
1 import { GetPlaylist } from "../spotify.js";
2 import { JSONtoCSV } from "../jsontocsv.js";
3 import { saveJSON } from "../savejson.js";
4 import { getMostPopularVideosFromYear, displayVideos } from "../youtube.js";
5 import { getToken, getTopVideos } from "../tiktok.js";
6 // import { getTweets } from "../twitter.js";
7 import { getTWData } from "../twitch.js";
8
9 import data from "../JSON/spotify2020.json" assert { type: "json" };
10
11 async function main() {
12
13   const data = await getMostPopularVideosFromYear(2020);
14   console.log(data);
15   saveJSON(data, "youtube2020.json");
16
17 }
18 main();
19
20
```

```
JS savejson.js > saveJSON
1 // save obj code to json file
2 import fs from "fs";
3
4 async function saveJSON(json, path) {
5   fs.writeFileSync(path, JSON.stringify(json));
6   console.log(`JSON file saved to ${path}`);
7 }
8
9 export { saveJSON };
10
```

```
main.py > ...
1 import pandas as pd
2
3 dp=pd.read_json('youtube2020.json')
4 dp.to_csv('youtube2020.csv', index=False)
5
```

Sırasıyla Javascript dili kullanarak API yöntemi ile verileri aldım.

Sonrasında verileri filtreleyip JSON formatında kaydettim.

En son ise Python dili kullanarak pandas kütüphanesi ile bu verileri CSV formatına çevirdim.