

# Klasik Test Teorisi(KTT)

Madde Analizi



Dr. Kübra Atalay Kabasakal

# Paketler

```
1 library("dplyr") # veri düzenleme
2 library("car")   # veri düzenleme
3 library("skimr") # veri inceleme
4 library("DataExplorer") # veri inceleme
5 library("ggcorrplot")   # veri inceleme
6 library("psych")        #KTK analizler
7 library("CTT")          #KTK analizler
8 library("ShinyItemAnalysis") #KTK analizler
9 # devtools::install_github("zief0002/QME")
10 library("QME") #KTK analizler
11 library("difR") #KTK analizler
```

# Veriyi Okuma

```
1 nfc <- read.csv("nfc_data.csv", header = TRUE)
2 matris <- dplyr::select(nfc,
3                           starts_with("nfc"))
4
5 head(matris)
```

	nfc01	nfc02	nfc03	nfc04	nfc05	nfc06	nfc07	nfc08	nfc09	nfc10	nfc11	nfc12
nfc13												
1	5	7	5	1	6	2	2	5	2	1	2	2
6												
2	5	5	3	2	5	3	3	3	2	4	3	2
3												
3	5	6	5	1	7	3	2	3	1	3	3	1
6												
4	5	5	4	2	5	5	2	2	2	2	4	3
2												
5	2	3	3	5	3	5	6	1	6	6	6	6
2												
6	6	6	6	1	6	3	1	1	1	1	1	1
6												
nfc14	nfc15	nfc16										
1	5	2	1									

# Ters Kodlama

```
1  nfc_key <- c(1,1,1,-1,1,-1,-1,-1,-1,-1,-1,-1,1,1,-1,-1)
2
3  ters_matris <- psych::reverse.code(
4    keys = nfc_key,
5    items = matris,
6    mini = 1,
7    maxi = 7)
```

# Korelasyon

```
1 cor_ters_matris <- psych::polychoric(ters_matris)$rho
2 colnames(ters_matris) <- colnames(matris)
3 ters_matris <- as.data.frame(ters_matris)
4
5 head(ters_matris)
```

	nfc01	nfc02	nfc03	nfc04	nfc05	nfc06	nfc07	nfc08	nfc09	nfc10	nfc11	nfc12
nfc13												
1	5	7	5	7	6	6	6	3	6	7	6	6
6												
2	5	5	3	6	5	5	5	5	6	4	5	6
3												
3	5	6	5	7	7	5	6	5	7	5	5	7
6												
4	5	5	4	6	5	3	6	6	6	6	4	5
2												
5	2	3	3	3	3	3	2	7	2	2	2	2
2												
6	6	6	6	7	6	5	7	7	7	7	7	7
6												
	nfc14	nfc15	nfc16									
1	5	6	7									

# Madde analizi

- Madde analizi
  - her bir madde için betimleyici istatistikler (örn. ortalama, standart sapma, minimum ve maksimum),
  - madde düzeyinde istatistikler (örn. güçlük ve ayırt edicilik) ve bunların ölçek düzeyinde istatistiklerle (örn. güvenilirlik) ilişkileri
- R'de, ikili (yani, 0 veya 1) ve sıralı (örneğin, Likert ölçeği maddeleri) maddeler üzerinde madde analizi yapmak için çeşitli paketler bulunmaktadır.

# CTT (Willse 2018): itemAnalysis()

```
1 itemanalysis_ctt <- CTT::itemAnalysis(items = ters_matris, pBisFlag = .2,  
2 bisFlag = .2)  
3 itemanalysis_ctt$itemReport
```

	itemName	itemMean	pBis	bis	alphaIfDeleted	lowPBis	lowBis
1	nfc01	5.535153	0.5830007	0.6324858	0.8603155		
2	nfc02	4.944582	0.5983857	0.6212552	0.8587924		
3	nfc03	4.435897	0.5120628	0.5278773	0.8626385		
4	nfc04	5.526882	0.4275829	0.4678823	0.8665634		
5	nfc05	5.818859	0.4657545	0.5132235	0.8646917		
6	nfc06	4.254756	0.3164060	0.3254804	0.8717944		
7	nfc07	5.299421	0.6637826	0.7012827	0.8562648		
8	nfc08	4.660050	0.5697411	0.5895616	0.8598588		
9	nfc09	5.549214	0.4916520	0.5349076	0.8636576		
10	nfc10	4.837055	0.5032326	0.5251349	0.8631153		
11	nfc11	5.254756	0.5772984	0.6088347	0.8596070		
12	nfc12	5.554177	0.5069643	0.5468089	0.8628642		
13	nfc13	4.104218	0.5384126	0.5524339	0.8614377		
14	nfc14	3.570720	0.4585722	0.4710690	0.8651300		
15	nfc15	5.752688	0.4678062	0.5152649	0.8645620		
16	nfc16	5.524400	0.4585389	0.4943813	0.8650091		

# psych (Revelle 2021): alpha()

```
1 itemanalysis_psych <- psych::alpha(x = ters_matrix)
2 itemanalysis_psych
```

Reliability analysis

Call: psych::alpha(x = ters\_matrix)

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.87	0.87	0.88	0.3	6.8	0.0054	5	0.82	0.29

95% confidence boundaries

	lower	alpha	upper
Feldt	0.86	0.87	0.88
Duhachek	0.86	0.87	0.88

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
nfc01	0.86	0.86	0.87	0.29	6.2	0.0059	0.0098	0.27	
nfc02	0.86	0.86	0.87	0.29	6.2	0.0059	0.0098	0.27	



# ShinyItemAnalysis (Martinkova, Hladka, and Netik 2021):

```
1 itemanalysis_shiny <- ShinyItemAnalysis::ItemAnalysis(Data = ters_matris)
2
3 itemanalysis_shiny
```

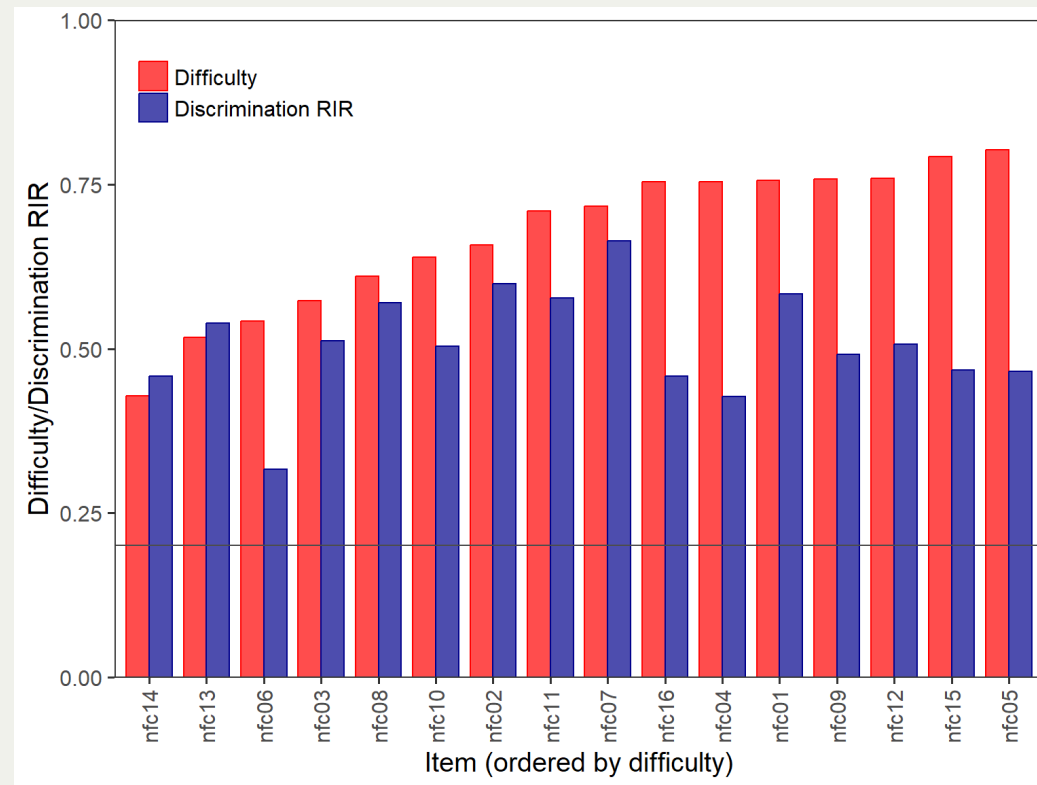
	Difficulty	Mean	SD	Cut.score	obs.min	Min.score	obs.max	Max.score	Prop.max.score	RIR
nfc01	0.7558588	5.535153	1.188645	NA	1	1	7	7	0.17866005	0.5830007
nfc02	0.6574304	4.944582	1.390390	NA	1	1	7	7	0.12406948	0.5983857
nfc03	0.5726496	4.435897	1.396922	NA	1	1	7	7	0.04962779	0.5120628
nfc04	0.7544803	5.526882	1.473097	NA	1	1	7	7	0.27791563	0.4275829
nfc05	0.8031431	5.818859	1.227220	NA	1	1	7	7	0.34491315	0.4657545
nfc06	0.5424593	4.254756	1.499872	NA	1	1	7	7	0.04797353	0.3164060
nfc07	0.7165702	5.299421	1.318365	NA	1	1	7	7	0.17038875	0.6637826
nfc08	0.6100083	4.660050	1.554549	NA	1	1	7	7	0.10918114	0.5697411
nfc09	0.7582024	5.549214	1.515188	NA	1	1	7	7	0.33995037	0.4916520
nfc10	0.6395092	4.837055	1.524876	NA	1	1	7	7	0.11579818	0.5032326
nfc11	0.7091260	5.254756	1.448205	NA	1	1	7	7	0.21339950	0.5772984
nfc12	0.7590295	5.554177	1.399119	NA	1	1	7	7	0.29859388	0.5069643
nfc13	0.5173697	4.104218	1.420891	NA	1	1	7	7	0.03639371	0.5384126
nfc14	0.4284533	3.570720	1.471642	NA	1	1	7	7	0.02564103	0.4585722
nfc15	0.7921147	5.752688	1.344924	NA	1	1	7	7	0.35649297	0.4678062
nfc16	0.7540667	5.524400	1.402319	NA	1	1	7	7	0.27956989	0.4585389

	RIT	Corr.criterion	ULI	gULI	Alpha.drop	Index.rel	Index.val	Perc.miss	Perc.nr
nfc01	0.6409212		NA	0.2726639	NA	0.8603155	0.7618279	NA	0
nfc02	0.6639133		NA	0.3391863	NA	0.8587924	0.9230985	NA	0
nfc03	0.5880748		NA	0.3049003	NA	0.8626385	0.8214949	NA	0
nfc04	0.5166594		NA	0.2873591	NA	0.8665634	0.7610891	NA	0

# ShinyItemAnalysis (Martinkova, Hladka, and Netik 2021):

```
1 ShinyItemAnalysis::DDplot(Data = ters_matrix, discrim = "RIR")
```



# Güvenirlilik

- **psych** paketindeki `splitHalf()` fonksiyonu ile bu işlem çok daha kolaydır. Bu fonksiyon, belirli bir veri kümesi için tüm (ya da en azından çok sayıda) olası split-half güvenirlik değerlerini hesaplar. Bu fonksiyonu kullanmak için sadece analiz edilecek veri setini belirtmek yeterlidir.

```
1 psych::splitHalf(r = ters_matris)
```

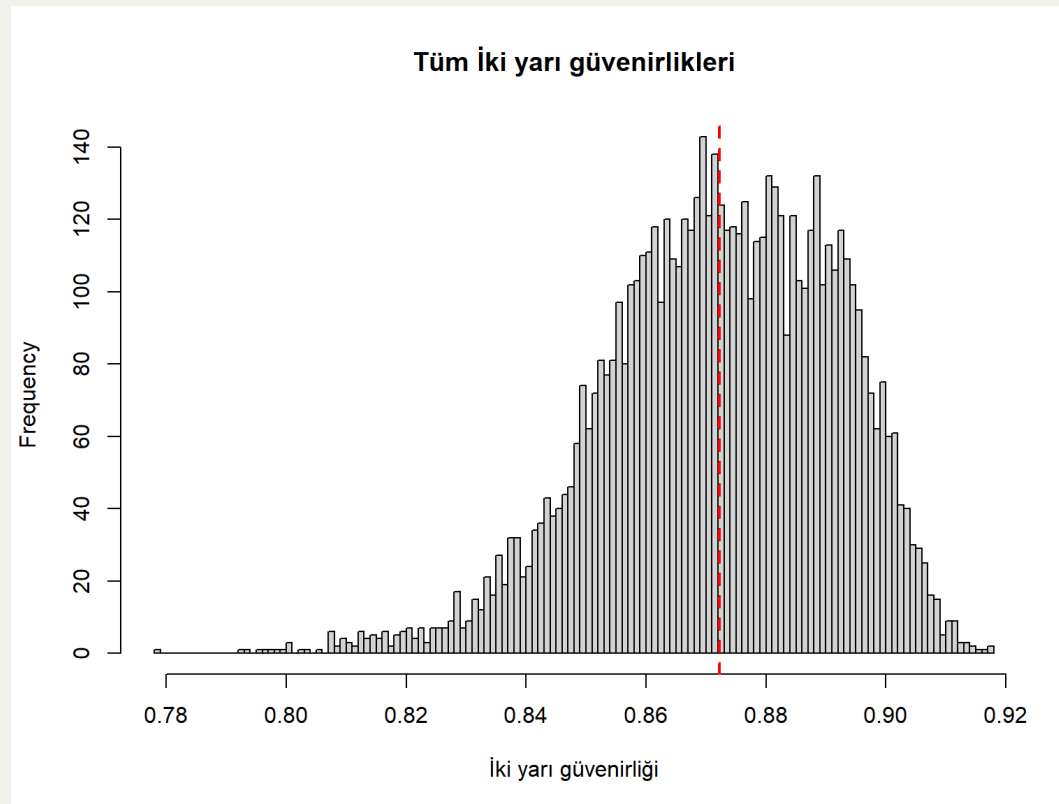
```
Split half reliabilities
```

```
Call: psych::splitHalf(r = ters_matris)
```

```
Maximum split half reliability (lambda 4) = 0.92  
Guttman lambda 6 = 0.88  
Average split half reliability = 0.87  
Guttman lambda 3 (alpha) = 0.87  
Guttman lambda 2 = 0.88  
Minimum split half reliability (beta) = 0.78  
Average interitem r = 0.3 with median = 0.29
```

# İki yarı

```
1 sp_rel <- psych::splitHalf(r = ters_matris, raw = TRUE)
2 hist(x = sp_rel$raw,
3     breaks = 101,
4     xlab = "İki yarı güvenirliği",
5     main = "Tüm İki yarı güvenirlikleri")
6 abline(v = mean(sp_rel$raw), col = "red", lwd = 2, lty = 2)
```



# İç tutarlık

- İç tutarlılık, bir ölçüm aracındaki maddeler arasındaki homojenlik derecesidir.

```
1 itemanalysis_ckt
```

```
Number of Items
```

```
16
```

```
Number of Examinees
```

```
1209
```

```
Coefficient Alpha
```

```
0.87
```

- Benzer şekilde, **psych** içinde `alpha()` fonksiyonunu kullanarak madde analizi yaptığımızda, iç tutarlılık sonuçlarını da elde ettik. Tekrar `itemanalysis_psych` yazdırdığımızda aşağıdaki sonuçları görebiliriz:

```
1 itemanalysis_psych
```

```
Reliability analysis
```

```
Call: psych::alpha(x = ters_matris)
```

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r
0.87	0.87	0.88	0.3	6.8	0.0054	5	0.82	0.29

```
95% confidence boundaries
```

```
lower alpha upper
```

```
Feldt      0.86 0.87 0.88
```

Duhachek 0.86 0.87 0.88

Reliability if an item is dropped:

	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha se	var.r	med.r
nfc01	0.86	0.86	0.87	0.29	6.2	0.0059	0.0098	0.27
nfc02	0.86	0.86	0.87	0.29	6.2	0.0059	0.0091	0.28
nfc03	0.86	0.86	0.88	0.30	6.4	0.0058	0.0094	0.29
nfc04	0.87	0.87	0.88	0.31	6.6	0.0056	0.0098	0.30
nfc05	0.86	0.87	0.88	0.30	6.5	0.0057	0.0100	0.30
nfc06	0.87	0.87	0.89	0.31	6.9	0.0054	0.0084	0.31
nfc07	0.86	0.86	0.87	0.29	6.0	0.0060	0.0092	0.27
nfc08	0.86	0.86	0.88	0.29	6.3	0.0059	0.0098	0.28

# İç tutarlık

- raw\_alpha: Maddeler arasındaki kovaryanslara dayalı alfa katsayısı ( $\geq 0,7$  değerleri “kabul edilebilir” güvenilirliği gösterir)
- std.alpha: Maddeler arasındaki korelasyonlara dayalı standartlaştırılmış alfa katsayısı
- G6: Guttman’ın Lambda 6 güvenilirliği
- average\_r: Ortalama öğeler arası korelasyonlar
- S/N: Sinyal/Gürültü oranı, burada  $s/n = nr/(1 - r)$
- ase: Alfa katsayısının standart hatası
- Ortalama: Maddelerin ortalaması alınarak veya toplanarak oluşturulan ölçek ortalaması
- sd: Toplam puanın standart sapması
- median\_r: Medyan öğeler arası korelasyonlar

# İç tutarlık

- Bu değerler arasında, iç tutarlılığı yorumlamak için raw\_alpha değerini kullanırız. NFC Ölçeği için yüksek iç tutarlılığa işaret etmektedir. Çıktının “Bir madde çıkarılırsa güvenilirlik” başlığı altındaki ikinci kısmı, her bir madde araçtan çıkarıldıktan sonra iç tutarlılığın nasıl değiştiğini göstermektedir. Örneğimizde, ham\_alfa sütunu iç tutarlılığın ya aynı kaldığını ya da 0,86’ya düştüğünü göstermektedir, bu da maddelerin hiçbirini çıkarmanın güvenilirliği artırmamıza yardımcı olmayacağını ve bu nedenle orijinal ölçeği kullanmamız gerektiğini göstermektedir.



# İç tutarlık

- NFC Ölçeği için iç tutarlılığı hesaplamak üzere **QME** paketini (Brown et al. 2016) kullanacağız. Alfa katsayısı ve Guttman'ın lambda güvenilirlik istatistiklerine ek olarak, **QME** paketi Feldt-Gilmer ve Feldt-Brennan güvenilirlik istatistiklerini de hesaplar.

```
1 reliability_qme <- QME::analyze(test = ters_matris, id = FALSE, na_to_0 = FALSE)
2
3 reliability_qme$test_level
```

\$descriptives

	Value
Minimum Score	21.0000000
Maximum Score	112.0000000
Mean Score	80.6228288
Median Score	82.0000000
Standard Deviation	13.1904670
IQR	17.0000000
Skewness (G1)	-0.6471103
Kurtosis (G2)	0.7557418

\$reliability

	Estimate	95% LL	95% UL	SEM
Guttman's L2	0.8731550	0.8624677	0.8833207	4.697826
Guttman's L4	0.8932339	0.8842384	0.9017905	4.309996
Feldt-Gilmer	0.8717610	0.8609563	0.8820384	4.723569
Feldt-Brennan	0.8712988	0.8604551	0.8816133	4.732074
Coefficient Alpha	0.8704009	0.8594817	0.8807874	4.748551

# Spearman-Brown formülü

- İlk olarak, NFC ölçeğinin uzunluğunu 0,5 kat kısaltmanın güvenirliği nasıl değiştirileceğini inceleyeceğiz (yani, uzunluğu 16 maddeden 8 maddeye düşürmek)

```
1 CTT::spearman.brown(r.xx = 0.87, input = 0.5, n.or.r = "n")
```

```
$r.new
```

```
[1] 0.7699115
```

- Çıktı, NFC Ölçeğinin uzunluğunu 16 maddeden 8 maddeye düşürmek zorunda kalsaydık, iç tutarlılığın kabaca 0,77 olacağını göstermektedir.

# Spearman-Brown formülü

- iç tutarlılığı 0.87'den (mevcut değer) 0.90'a (hedef değer) çıkarmak için kaç maddeye daha ihtiyacımız olacağını göreceğiz.

```
1 n <- CTT::spearman.brown(r.xx = 0.87, input = 0.90, n.or.r = "r")  
2 round(n$n.new * 16, digits = 0)
```

```
[1] 22
```

- Yukarıdaki sonuç, 0,90'lık iç tutarlılık düzeyine ulaşmak için 22 maddeye (yani orijinal NFC maddeleriyle benzer özelliklere sahip 6 ek maddeye) ihtiyacımız olduğunu göstermektedir.

# Ölçüt bağlantılı geçerlik

- Ölçüt bağlantılı geçerlik, bir ölçme aracının aynı (veya benzer yapıyı) ölçen başka bir aracın sonucunu ne kadar iyi tahmin ettiğini gösterir.
- *eşzamanlı geçerlik*
- *yordama geçerlik*
- Öz-Kontrol Ölçeği'nin kısa formundan alınan öz-kontrol puanları (Bertrams and Dickhäuser 2009),
- Yetişkin Mizaç Anketi (Wiltink, Vogelsang, and Beutel 2006) Çabalı Kontrol Ölçeğinden alınan çabalı kontrol puanları ve
- Eylem Kontrol Ölçeğinden (Kuhl 1994) alınan eylem yönelimi puanları.

Önceki araştırmaların bulgularına dayanarak, NFC ile diğer üç yapı (öz kontrol, çaba gerektiren kontrol ve eylem kontrolü) arasında pozitif, küçük ila orta düzeyde ilişkiler bulmayı bekliyoruz.

# Ölçüt bağlantılı geçerlik

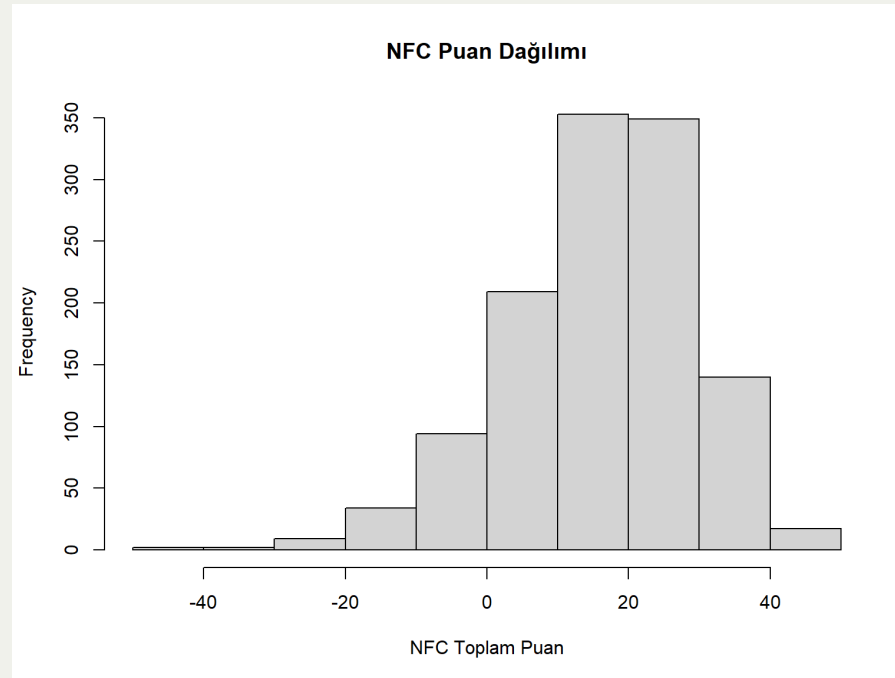
İlk olarak, NFC Ölçeği için 1 ila 7 arasındaki madde puanlarını -3 ila +3 olarak yeniden kodlama **car** paketindeki (Fox, Weisberg, and Price 2020) **recode()** fonksiyonu ile yapılabilir.

```
1 recode_nfc <- function(x) {  
2   car::recode(x, "1=-3; 2=-2; 3=-1; 4=0; 5=1; 6=2; 7=3")  
3 }  
4 ters_matris <- apply(ters_matris,  
5                       2, # her bir sutuna uygulanir  
6                       recode_nfc)  
7 nfc_score <- rowSums(ters_matris)  
8 summary(nfc_score)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-43.00	9.00	18.00	16.62	26.00	48.00

# Ölçüt bağlantılı geçerlik

```
1 hist(nfc_score,  
2      xlab = "NFC Toplam Puan",  
3      main = "NFC Puan Dağılımı")
```



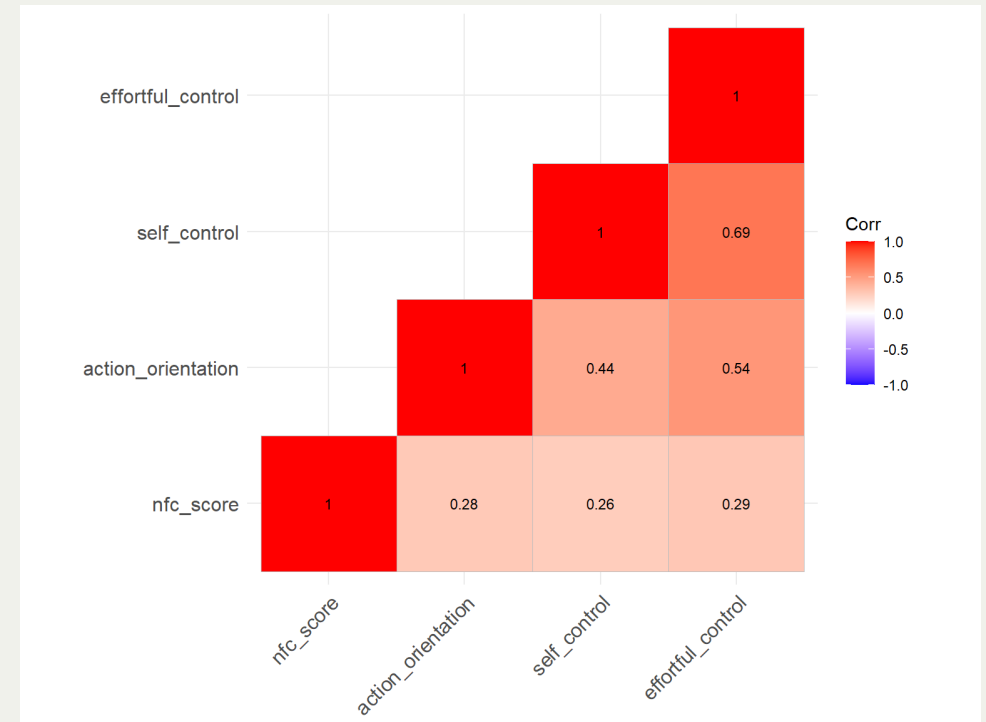
# Ölçüt bağlantılı geçerlik

- NFC Ölçeği puanlarını diğer ölçeklerden gelen puanlarla tek bir veri kümesinde birleştirebiliriz.

```
1 scores <- cbind(nfc_score, nfc[, c("action_orientation",  
2 "self_control", "effortful_control")])
```

# Ölçüt bağlantılı geçerlik

```
1 cormat_scores <-  
2 cor(scores, use = "pairwise.complete")  
3  
4 ggcorrplot::ggcorrplot(corr = cormat_scores,  
5 type = "lower",  
6 show.diag = TRUE,  
7 lab = TRUE,  
8 lab_size = 3)
```





# Ölçüt bağlantılı geçerlik

- korelasyon matrisi grafiği, NFC Ölçeği puanları ile diğer ölçeklerden alınan puanlar arasında gerçekten de pozitif, küçük ila orta düzeyde bir korelasyon olduğunu göstermektedir. Bunların sadece ölçek puanları arasındaki ham korelasyonlar olduğunu unutmayın. Yani, her bir ölçekte yer alan ölçüm hatasından kaynaklanan zayıflama için **düzeltilmemiştir**. Zayıflatılmış korelasyonları elde etmek için **psych** paketinden `correct.cor()` fonksiyonunu kullanacağız. Bu fonksiyonu kullanmak için, ölçek puanlarının ham korelasyon matrisini ve güvenilirlik (yani iç tutarlılık) tahminlerinin bir vektörünü listelememiz gerekir. NFC Ölçeği için güvenilirlik tahmini 0.87'dir ve üç kriter ölçümü için güvenilirlik tahminleri (Grass et al. 2019) şeklindedir:
- Eylem Kontrol Ölçeği: 0.791
- Özdenetim Ölçeği: 0.817
- Çabalı Kontrol Ölçeği: 0.783

# Ölçüt bağlantılı geçerlik

- Bu bilgiyi kullanarak, zayıflatılmış korelasyonları aşağıdaki gibi hesaplayabiliriz:

	nfc_score	action_orientation	self_control	effortful_control
nfc_score	0.87	0.34	0.31	0.35
action_orientation	0.28	0.79	0.54	0.69
self_control	0.26	0.44	0.82	0.87
effortful_control	0.29	0.54	0.69	0.78

- yeni korelasyon matrisinde, üst köşegen kısmı zayıflama için düzeltilmiş dört ölçek puanı arasındaki korelasyonları, köşegen kısmı dört ölçek için güvenilirlik tahminlerini ve alt köşegen kısmı dört ölçek puanı arasındaki orijinal (yani ham) korelasyonları göstermektedir. Zayıflama için düzeltme uygulandıktan sonra, yeni korelasyon değerleri ham değerlerinden daha yüksek hale gelmiştir. Örneğin, NFC ölçeğinden alınan puanlar Çabalı Kontrol Ölçeğinden alınan puanlarla korelasyona sahipken, iki ölçek arasındaki düzeltilmiş korelasyon 'tir.

# Madde Geçerlik İndeksi

- NFC Ölçeğinden alınan toplam puanlar ile ölçüt ölçümleri arasındaki ilişkiye ek olarak, NFC Ölçeğindeki her bir madde ile ölçüt ölçümlerinden alınan puanlar arasındaki ilişkiyi kontrol etmek için madde geçerlilik indeksini (IVI) de hesaplayabiliriz. IVI -0,5 ile 0,5 arasında değişebilir ve büyük değerler (mutlak büyüklük olarak) daha yüksek geçerliliğe işaret eder.
- `ivi()` fonksiyonu orijinal olarak **hemp** paketinden gelmektedir.

```
1 ivi <- function(item, criterion) {  
2   s_i <- sd(item, na.rm = TRUE)
```

```
                                nfc_ivi  
nfc01 0.37297915  
nfc02 0.21153487
```

```

3  r <- cor(item, criterion, use = "c
4  index <- s_i * r
5  return(index)
6  }
7
8  nfc_ivi <- apply(ters_matris,
9  2,
10 function(x) ivi(item = x,
11 criterion = nfc$action_orientation
12
13 nfc_ivi <- as.data.frame(nfc_ivi)
14 print(nfc_ivi)

```

```

nfc03 0.19981987
nfc04 0.14540417
nfc05 0.06209198
nfc06 0.20228579
nfc07 0.26048244
nfc08 0.25950684
nfc09 0.25584111
nfc10 0.44981819
nfc11 0.27994998
nfc12 0.24413937
nfc13 0.29590845
nfc14 0.29309729
nfc15 0.02070200

```

# Kaynaklar

- Bertrams, Alex, and Oliver Dickhäuser. 2009. "Messung Dispositioneller Selbstkontroll-Kapazität." *Diagnostica* 55 (1): 2–10. <https://doi.org/10.1026/0012-1924.55.1.2>.
- Brown, Ethan, Andrew Zieffler, Kyle Nickodem, Kory Vue, and Erik Anderson. 2016. *QME: Classic Test Theory Item Analysis*. <https://github.com/zief0002/QME>.
- Fox, John, Sanford Weisberg, and Brad Price. 2020. *Car: Companion to Applied Regression*. <https://CRAN.R-project.org/package=car>.
- Grass, Julia, Florian Krieger, Philipp Paulus, Samuel Greiff, Anja Strobel, and Alexander Strobel. 2019. "Thinking in Action: Need for Cognition Predicts Self-Control Together with Action Orientation." *PLOS ONE* 14 (8): 1–20. <https://doi.org/10.1371/journal.pone.0220282>.
- Kuhl, Julius. 1994. "Action Versus State Orientation: Psychometric Properties of the Action Control Scale (ACS-90)." *Volition and Personality: Action Versus State Orientation* 47 (56).
- Martinkova, Patricia, Adela Hladka, and Jan Netik. 2021. *ShinyItemAnalysis: Test and Item Analysis via Shiny*. <https://CRAN.R-project.org/package=ShinyItemAnalysis>.

- Revelle, William. 2021. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. <https://personality-project.org/r/psych/> <https://personality-project.org/r/psych-manual.pdf>.
- Willse, John T. 2018. *CTT: Classical Test Theory Functions*. <https://CRAN.R-project.org/package=CTT>.
- Wiltink, Jörg, Ute Vogelsang, and Manfred E Beutel. 2006. "Temperament and Personality: The German Version of the Adult Temperament Questionnaire (ATQ)." *GMS Psycho-Social Medicine* 3.

