Varsayımlar (Varyansların Homojenliği)

- Gruplanmamış veriler için, varyansların homojenliği varsayımı, bir sürekli değişken için puanlardaki değişkenliğin, bir başka sürekli değişkenin tüm degerlerinde kabaca aynı olmasıdır.
- **Gruplandırılmış veriler** için, varyansların homejenliği varsayımı, bağımlı değişkendeki değişkenliğin gruplama değişkeninin tüm düzeylerinde yaklaşık olarak aynı olmasıdır
- Varyansların homojenliği normallik sayıltısı ile ilişkilidir. Çok değişkenli normallik sayıltısı karşılandığında, değişkenler arasındaki ilişkiler homojendir.
- Varyansların homojenliği varsayımının <u>ihlali</u>, değişkenlerden birinin normallikten sapması ya da bağımsız değişkenin bazı düzeylerinde yapılan büyük ölçme hatalarında kaynaklanır.
- Gruplanmamış verilerin analizi için hayati önemde değildir; iki değişkenli saçılma grafikleri ile incelenebilir. Değişkenler arasındaki varyans farklılıkları tahmin edilebiliyorsa grafikte bu farklılıklar gözlenebilir. Grafikte açıklanamayan farklılıklar varsa analiz zayıflar ve geçerliği düşer. Ancak homojenlik sağlanırsa daha fazla yordama gücüne erişilebilir.



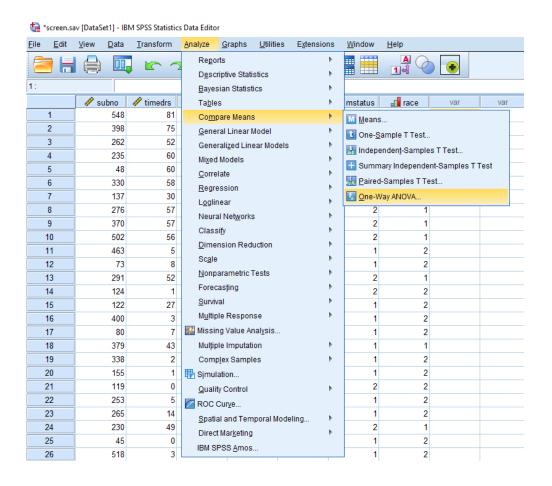
Varsayımlar (Varyansların Homojenliği)

- Varyansın homojenliği için testler bulunmaktadır, ancak bu testlerin çoğu normalliği de değerlendirdikleri için çok katıdır.
 - Bir istisna **Levene**'nin varyansın homojenliği testidir, ki bu test normallikten sapmalara çok duyarlı değildir.
- Varyansın homojenliğinin çok değişkenli benzeri, varyans-kovaryans matrislerinin homojenliğidir.
 - IBM SPSS, **Box's M** testi kullanılabilir



Levene Testi

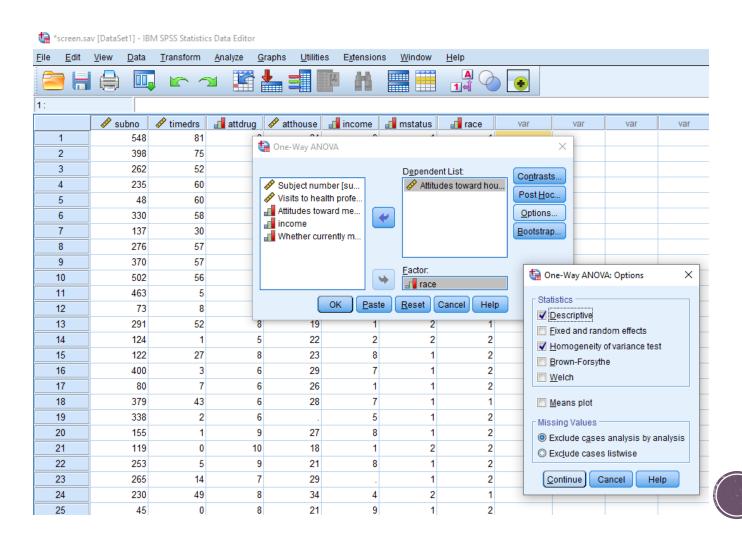
"Analyze" ana menüsünden önce "Compare Means", daha sonra da "One –Way ANOVA"ya tıklayınız.





Levene Testi

- İncelemek istediğiniz gruplama değişkenini «Factor» kutusuna; bağımlı değişkeni «Dependent List» kutusuna aktarınız.
- Options" a tıklayınız.
- "Statistics" bölümünde
 "Homogeneity of variance test"i seçiniz.
- Önce "Continue", sonra "OK"e tıklayınız.



Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Attitudes toward	Based on Mean	,607	1	462	,436
housework	Based on Median	,610	1	462	,435
	Based on Median and with adjusted df	,610	1	461,947	,435
	Based on trimmed mean	,609	1	462	,436

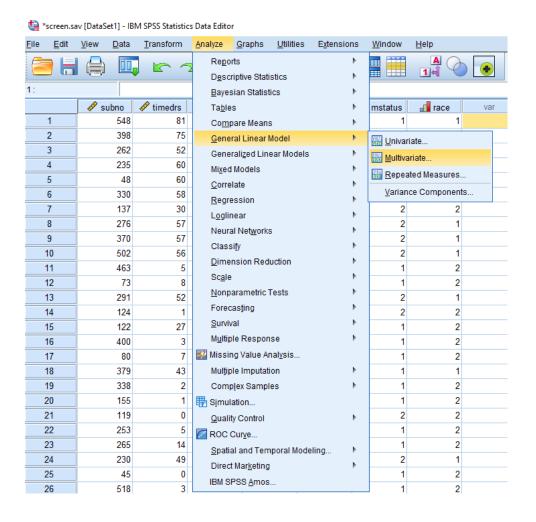


 Anlamlılık değerinin 0.05 ten büyük olması varyansların homojenliğine işaret etmektedir.



Box's M Testi

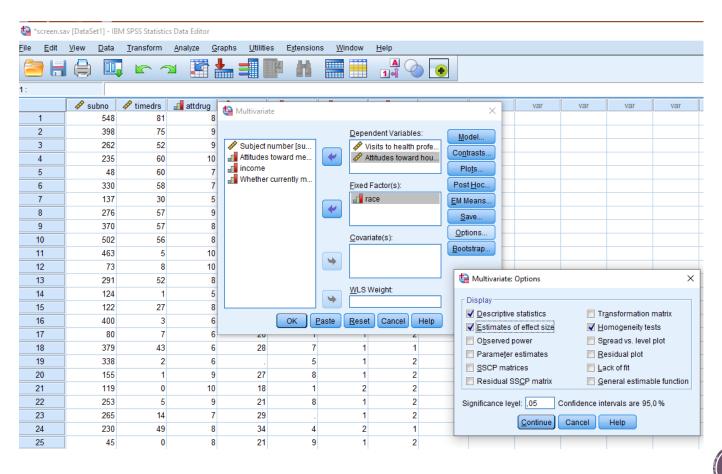
 "Analyze" ana menüsünden önce
 "General Linear Model"e, daha sonra da "Multivariate"e tıklayınız.





Box's M Testi

- İncelemek istediğiniz gruplama değişkenini «Fixed Factor(s)» kutusuna; bağımlı değişkeni «Dependent Variables» kutusuna aktarınız.
- Options" a tıklayınız.
- "Display" bölümünde "Homogeneity tests"i seçiniz.
- Önce "Continue", sonra "OK"e tıklayınız.





Box's Test of Equality of Covariance Matrices^a

Box's M	4,557
F	1,490
df1	3
df2	55177,778
Sig.	,215

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + race

Box's M Testi

• İstatiksel olarak anlamlı olmayan bir anlamlılık değeri çok değişkenli varyansların homojenliğinin sağlandığını göstermektedir.



- Veri dönüşümleri normallik, uç değerler, doğrusallık ve varyansların homojenliği varsayımlarının ihlalleri için kullanılabilmesine rağmen, genellikle önerilmez. Bunun nedeni, bir analizin içinde dönüştürülen değişkenlerin yorumlanmasının zorluğudur.
- Veri dönüştürmeye karar verildiğinde, değişkenin dönüşümden sonra normalliğini kontrol etmek önemlidir.
 - Çoğu zaman ilk dönüşümden sonra, sıfıra daha yakın çarpıklık ve basıklık değerleri, daha az uç değerler, daha iyi bir dağılım elde edinceye kadar dönüştürmeye devam edilir.



- Veri dönüştürmede değişkenlerin normallikten ne kadar uzaklaştıkları önemlidir.
 - Dağılım normalden orta derecede farklıysa, önce bir Karekök dönüşümü denenir.
 - Eğer dağılım önemli ölçüde normalden uzaklaşıyorsa, bir log dönüşümü denenir.
 - Dağılım ciddi şekilde farklıysa, ters dönüşüm denenir.
 - Son olarak, normallikten sapma şiddetli ise ve hiçbir dönüşüm yardımcı görünmüyorsa, değişken ikili kategorik bir değişkene çevrilebilir.



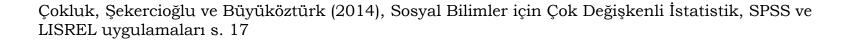
- Veri dönüştürmede değişkenlerin normallikten sapma yönü de dikkate alınır.
 Dağılımlar pozitif çarpıklığa sahip olduğunda, uzun kuyruk sağdadır. Negatif çarpıklıkları olduğunda, uzun kuyruk soldadır.
 - Negatif çarpıklık varsa, en iyi strateji değişkeni yansıtmak ve daha sonra pozitif çarpıklık için uygun dönüşümü uygulamaktır.
- Bir değişkeni yansıtmak için dağılımdaki en büyük puan bulunur. Ardından, bu değere 1 eklenerek sabit bir değer elde edilir sonra dağılımdaki her bir puan bu sabit değerden çıkarılarak yeni bir değişken oluşturulur.
- Bu şekilde, negatif çarpıklığa sahip bir değişken, dönüşümden önce pozitif çarpıklığa sahip bir değişkene dönüştürülür.
- Yansıyan bir değişkeni yorumlarken, yorumlamanın yönünü de ters çevirmek gerekecektir veya dönüşümünden sonra yeniden yansıtma düşünülebilir.



- Dönüşümlerinizi işlemlerinden varsayımlar tekrar kontrol edilmelidir.
- Örneğin, bir değişken sadece orta derecede pozitif çarpıksa, bir kare kök dönüşümü, değişkeni orta derecede negatif çarpık yapabilir ve bu durumda dönüştürmenin bir avantajı yoktur.
- Genellikle en yararlı olanı bulana kadar birkaç dönüşüm denemek gerekebilir.



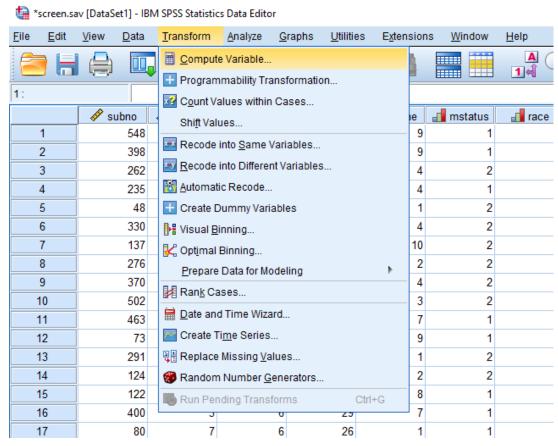
Orijinal Dağılımın Şekli	Dönüştürme Türü	SPSS Komutları			
Orta düzeyde pozitif çarpık	Karekök	NEWX=SQRT(X)			
Yüksek düzeyde pozitif çarpık	Logaritma	NEWX=LG10(X)			
Değer <0	Logaritma	NEWX=LG10(X+a)^a			
Aşırı pozitif çarpık	Ters Çevirme	NEWX=1/X			
Değer<0	Ters Çevirme	NEWX=1/(X+a)			
Orta düzeyde negatif çarpık	Yansıtma & Karekök	NEWX=SQRT(b-X)^b			
Yüksek düzeyde negatif çarpık	Yansıtma & Logaritma	NEWX=LG10(b-X)^b			
Aşırı negatif çarpık	Yansıtma & Ters Çevirme	$NEWX=1/(b-X)^b$			
a= En küçük değeri en az bir yapmak için her puana bir sabit eklenir b= En küçük değerin en az 1 olması için her puandan bir sabit çıkarılır					





Veri Dönüştürme SPSS Uygulama

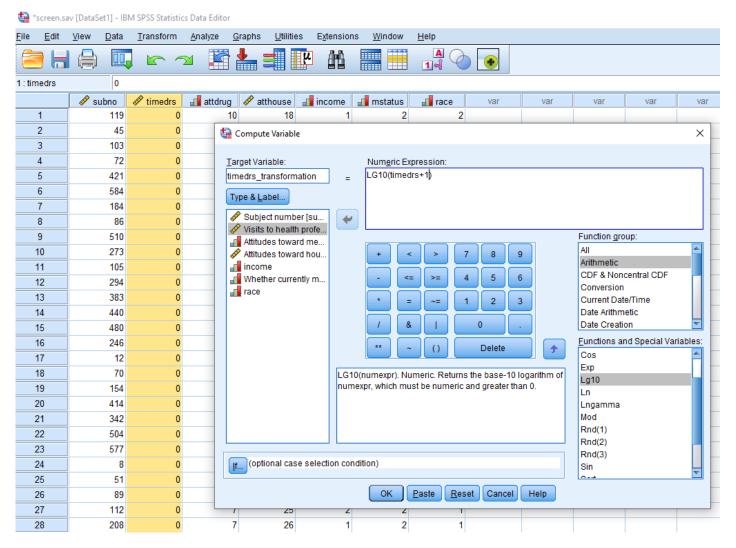
 "Transform" ana menüsünden "Compute Variable"a tıklayınız.





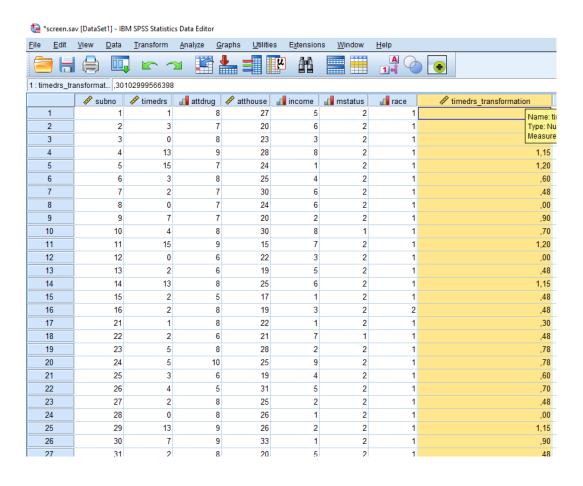
Veri Dönüştürme SPSS Uygulama

- "Target Variable" a dönüştürülecek olan değişkenin yeni adını yazınız.
- "Numeric Expression" kutusuna "lg10(timedrs+1)" yazınız.
- "OK"e tıklayınız.





Veri Dönüştürme SPSS Uygulama



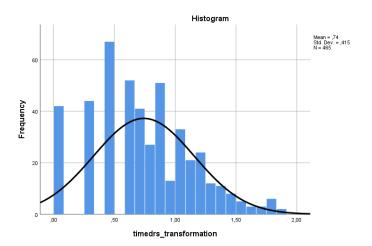
Frequencies

timedrs transformation

Statistics

unicuis_ua		
N	Valid	465
	Missing	0
Mean		,7413
Std. Deviation	on	,41525
Skewness		,228
Std. Error of	Skewness	,113
Kurtosis		-,177
Std. Error of	Kurtosis	,226
Minimum		,00
Maximum		1,91

 Log dönüşümünden sonra timedrs değişkeni için çarpıklık değeri 3,241'den 0,221'e, basıklık değeriyse 13,041'den -0,183'e düşmüştür.





- Çoklu bağlantılılık ve tekillik, değişkenler arasında çok yüksek korelasyon olduğunda ortaya çıkan bir korelasyon matrisi ile ilgili problemlerdir.
 - Çoklu bağlantılılıkta değişkenler arasında çok yüksek korelasyon (örneğin .90 ve üstü) söz konusudur;
 - Tekillikle ise, değişkenler gereksizdir; değişkenlerden biri diğer değişkenlerin iki veya daha fazlasının birleşimidir.
- Değişkenler çoklu bağlantılı veya tekil olduğunda, fazladan bilgi içerirler ve aynı analizde hepsine ihtiyaç yoktur. Bu değişkenlerin hepsinin modele yer alması modeldeki hataları artırır ve analizi zayıflatır.
- İki değişkenli veya çok değişkenli korelasyonlar çoklu bağlantı veya tekilliğe neden olabilirler.



- Örneğin, iki değişken arasındaki korelasyon katsayısının 0,90 veya **0,90'dan** daha yüksek olması çoklu bağlantı ve tekillik problemlerine işarettir.
 - **Not:** İki değişkenli korelasyon katsayılarının incelenmesi çoklu bağlantı probleminin belirlenmesi için yeterli değildir. Çünkü problem sadece iki değişken arasındaki korelasyonun yüksek olması değil, bir bağımsız değişkenin diğer bütün bağımsız değişkenlerle yüksek derecede korelasyona sahip olmasıdır.
- Çok değişkenli korelasyon çok yüksek ise, buna neden olan değişkeni bulmak için çok değişkenli istatistiklere ihtiyaç duyulduğundan tanı biraz daha zordur.



- Çoklu bağlantılılık ve tekillik hem mantıksal hem de istatistiksel sorunlara neden olur. Mantıksal problem, bir yapının analizini (faktör analizi, temel bileşenler analizi ve yapısal eşitlik modellemesi) yapılmadığı sürece, aynı analize gereksiz değişkenlerin dahil edilmesi iyi bir fikir değildir
- Bu değişkenler gerekli değildir ve hata terimlerinin büyüklüğünü şişirdikleri için aslında analizi zayıflatırlar.
- Bir yapı analiz edilmiyorsa veya analizde aynı değişkenin tekrarlanan ölçümleri ele alınmıyorsa aynı analize iki değişkenli korelasyonu .70 ve üstü olan iki değişken dahil etmeden önce dikkatlice düşünmek gerekir.
- Bu durumda değişkenlerden biri çıkarılabilir veya gereksiz değişkenlerden bileşik bir puan oluşturulabilir.



- Çoklu bağlantılılık ve tekillik, değişkenler için SMC (squared multiple correlation, R^2) değeri hesaplanarak incelenebilir.
- SMC yüksekse, değişken diğer değişkenlerle yüksek düzeyde ilişkilidir ve çoklu bağlantılılık vardır. SMC değeri 1 ise, değişken diğer değişkenlerle mükemmel bir şekilde ilişkilidir ve tekilliğiniz vardır.
- Birçok program her değişken için SMC değerlerini toleransa (1 SMC) dönüştürür ve SMC yerine toleransla ilgilenir.
- Tekilliğin incelenmesinde genellikle
 - Bir değişkenin diğer değişkenlerden türetilip türetilmediği ya da
 - Bir değişkenin iki değişkenin bileşimi olup olmadığı kontrol edilir. Böyle bir durum varsa değişkenlerden biri silinir.



- Çoklu bağlantının belirlenmesinde her bir değişken için tolerance (1-SMC) değeri incelenebilir.
 - Tolerance değeri, değişkendeki gözlenen varyansın modeldeki diğer bağımsız değişkenler tarafından açıklanmayan kısmıdır.
 - Tolerans (1 SMC) çok düşükse, değişken analize dahil edilmez.
 - Varsayılan tolerans düzeyleri .01 ile .0001 arasında; SMC'ler ise .99 ile .9999 arasında değişmektedir.
 - Tolerance değerinin 0,10'dan küçük olması çoklu bağlantı problemini gösterir.



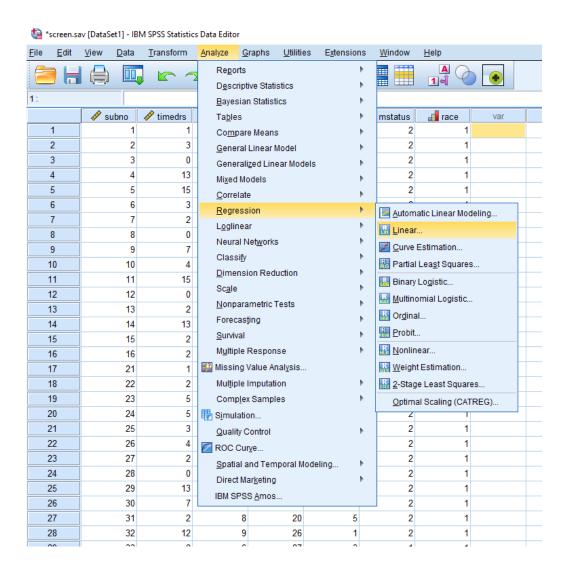
- Çoklu bağlantının belirlenmesinde her bir değişken için VIF =1/tolerance değeri incelenebilir.
 - Eğer bağımlı ve bağımsız değişkenler arasında hiç korelasyon yoksa, SMC=0 ve tolerans değeri 1'e eşit olur. Bu durumda VIF (1/tolerance) değeri de 1'e eşit olacaktır. VIF ≥ 10 olması çoklu bağlantı problemine işarettir.
- Çoklu bağlantının belirlenmesi için koşul indeksi (condition index, CI) incelenebilir
 - Koşul indeksi, bir değişkenin diğerlerine karşı bağlılığının veya bağımlılığının bir ölçüsüdür.
 - SMC, tolerans ve VIF her bir değişken için hesaplanırken; CI, her bir boyut/faktör için hesaplanır.
 - Yüksek CI'ye sahip bir boyut, iki veya daha fazla değişkenin varyansına güçlü bir şekilde katkıda bulunursa (yüksek bir varyans oranına sahip olduğunda) bir bağlantılılık sorunu oluşur.
- Belsely ve ark. (1980) tarafından önerilen çoklu bağlantılılık kriteri, belirli bir boyut için koşul indeksinin **30'**dan daha büyük olması ve en az iki değişken için varyans oranlarının **0,50**'den büyük olması çoklu bağlantı probleme işaret eder.



- Tespit edildiğinde bağlantılılık ile başa çıkmak için birkaç seçenek vardır.
 - İlk olarak, analizin tek amacı yordama ise, bunu göz ardı edebilirsiniz.
 - İkinci seçenek, en yüksek varyans oranına sahip değişkeni silmektir.
 - Üçüncü seçenek, bağlantılı değişkenleri toplamak veya ortalamasını almaktır
 - Dördüncü seçenek, temel bileşenleri hesaplamak analizlerde temel bileşenlerin kullanmaktır.

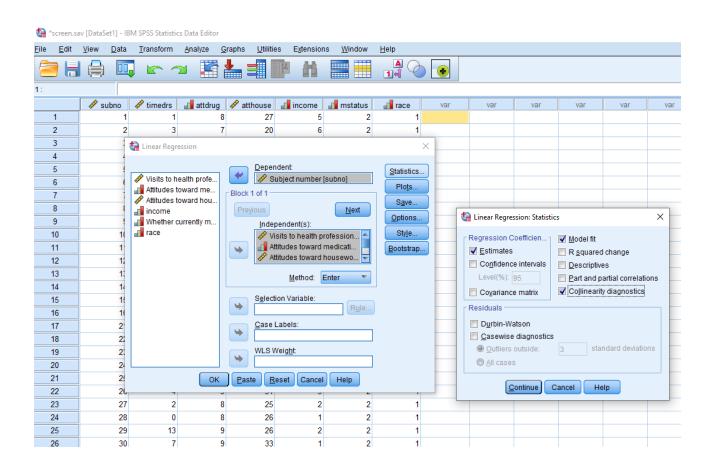


" Analyze"
 menüsünden
 "Regression"a, daha
 sonra da "Linear"a
 tıklayınız





- "Bağımlı değişkeni seçip "Dependent" kutusuna aktarınız.
- Bağımsız değişkenleri seçip (Örn, ltimedrs, attdrug, atthouse, mstatus ve race gibi). "Independent(s)" kutusuna taşıyınız.
- "Statistics"e tıklayınız.
 "Collinearity Diagnostics"i işaretleyiniz.
- "Continue" a sonra "OK" e tıklayınız





Coefficients^a

		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model		В	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	481,512	93,468		5,152	,000		
	Visits to health professionals	,443 ,835		,025	,531	,596	,969	1,032
	Attitudes toward medication	2,592	7,844	,015	,330	,741	,988	1,012
	Attitudes toward housework	-4,291 2,027		-,099	-2,117	,035	,982	1,019
	Whether currently married	-12,787	21,781	-,027	-,587	,557	,994	1,006
	race	-58,777	32,106	-,085	-1,831	,068	,996	1,004

a. Dependent Variable: Subject number

 Tolerance değerleri 0,10'dan büyük ve VIF değerleri de 10'dan küçük olduğundan çoklu bağlantı problemi olamadığı ifade edilebilir.



Collinearity Diagnostics

				Variance Proportions					
Model	Dimension	Eigenvalue	Condition Index	(Constant)	Visits to health professionals	Attitudes toward medication	Attitudes toward housework	Whether currently married	race
1	1	5,255	1,000	,00	,01	,00,	,00,	,00,	,00
	2	,608	2,940	,00	,96	,00	,00,	,00,	,00,
	3	,059	9,401	,00	,00,	,00	,02	,24	,71
	4	,044	10,943	,00	,03	,02	,31	,53	,11
	5	,026	14,090	,01	,00,	,48	,42	,07	,06
	6	,007	27,233	,99	,00,	,49	,25	,16	,11

a. Dependent Variable: Subject number

 CI değerleri 30'dan küçük olduğundan ve aynı satırda varyans oranı 0,50'den büyük iki veya daha fazla değişken bulunmadığından çoklu bağlantı probleminin olmadığı söylenebilir.

