T.C.

ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

BAZI KORELASYON KATSAYILARININ 1.TİP HATA VE TESTİN GÜCÜ BAKIMINDAN KARŞILAŞTIRILMASI

Elif TUĞRAN

Zootekni Anabilim Dalı

Tezin Sunulduğu Tarih:**19/06/2015**

Tez Danışmanı:

Prof. Dr. Mehmet MENDEŞ

ÇANAKKALE

Elif Tuğran tarafından Prof.Dr. Mehmet MENDEŞ yönetiminde hazırlanan ve **19/06/2015** tarihinde aşağıdaki jüri karşısında sunulan "**Bazı Korelasyon Katsayılarının 1.Tip Hata ve Testin Gücü Bakımından Karşılaştırılması**" başlıklı çalışma, Çanakkale Onsekiz Mart Üniversitesi Fen Bilimleri Enstitüsü **Zootekni Anabilim Dalı**'nda **YÜKSEK LİSANS TEZİ** olarak oybirliği/oyçokluğu ile kabul edilmiştir.

JÜRİ

Prof.Dr. Mehmet MENDEŞ ……

Başkan

Prof.Dr. Türker SAVAŞ ……

Üye

Doç.Dr. Erkut AKKARTAL ……

Üye

Prof.Dr.Levent GENÇ
Müdür
Fen Bilimleri Enstitüsü

Sıra No:……

İNTİHAL (AŞIRMA) BEYAN SAYFASI

Bu tezde görsel, işitsel ve yazılı biçimde sunulan tüm bilgi ve sonuçların akademik ve etik kurallara uyularak tarafımdan elde edildiğini, tez içinde yer alan ancak bu çalışmaya özgü olmayan tüm sonuç ve bilgileri tezde kaynak göstererek belirttiğimi beyan ederim.

Elif TUĞRAN

TEŞEKKÜR

Bu tezin gerçekleştirilmesinde, çalışmam boyunca benden bir an olsun yardımlarını esirgemeyen saygı değer danışman hocam, Prof.Dr. Mehmet MENDEŞ'e çalışmanın şekillenmesinde ve her aşamasında sarf ettiği düşünce ve emeğinden dolayı kendisine sonsuz teşekkürlerimi sunuyorum.

Çalışma süresince tüm zorlukları benimle göğüsleyen eşime, anneme, babama ve biricik oğluma, hayatımın her evresinde bana destek olan değerli aile mensuplarına sonsuz teşekkürlerimi sunarım.

Ayrıca çalışmanın şekillenmesinde desteklerini esirgemeyen tüm bölüm hocalarıma ayrı ayrı teşekkür ederim.

Ayrıca tezimin jüri hocalarım Prof.Dr. Türker SAVAŞ ve Doç.Dr. Erkut AKKARTAL'a teşekkürlerimi sunarım.

Verilerin toplanmasında ve tezin diğer aşamalarında emekleri geçen Soner YİĞİT ve Altan ŞAN hocalarıma da teşekkürlerimi sunarım.

Elif TUĞRAN

Çanakkale, Haziran 2015

SİMGELER VE KISALTMALAR

(n)	Örnek genişliği
(ρ)	Korelasyon yapıları
$n_i \geq 10$	Örnek genişliğinin yeterli olması
LMS	En Küçük Medyan Kareler Yöntemi
R(x) ve R(y), (X_i, Y_i)	Değişken çiftlerinin ranklarını gösterir
(X_i, Y_i)	Değişken gözlem çiftleri
r_w,	Winsorize korelasyon katsayısı
γ ($g = \gamma n$)	Winsorize yüzdesini gösterir
r_{xy}	Pearson korelasyon katsayısı
n!	X ve Y özelliklerine ilişkin bütün permütasyonlar
(r_{pxy})	Pearson korelasyon katsayıları
(N)	Permütasyon sayısına
AD	Adlibitum (serbest yemleme)
R20	% 20 Kısıtlı yemleme
NF6	6 saat yemleme
(g)	Ağırlık
Z	Test istatistiği
C	Uyum çiftlerin sayısı
D	Uyumsuz çiftlerin sayısı
PM	Pearson-Moment
SR	Spearman-Rank
KT	Kendall-Tau
PT	Permütasyon-tabanlı
WK	Winsorize korelasyon katsayılarına
SD	Serbestlik dereceli
$\sum d_x d_y$	Çarpımlar toplamı
$\sum d_x^2$ ve $\sum d_y^2$	X ve Y özelliklerine ilişkin kareler toplamını
$\overline{X}_W$ ve $\overline{Y}_W$	X ve Y Winsorize ortalamalar

ÖZET

BAZI KORELASYON KATSAYILARININ 1.TİP HATA VE TESTİN GÜCÜ BAKIMINDAN KARŞILAŞTIRILMASI

Elif TUĞRAN
Çanakkale Onsekiz Mart Üniversitesi
Fen Bilimleri Enstitüsü
Ziraat Fakültesi Zootekni Anabilim Dalı Yüksek Lisans Tezi
Danışman : Prof.Dr.Mehmet MENDEŞ
19/06/2015, 35

Bu tez çalışmasında, bazı korelasyon katsayıları (Pearson-Moment, Spearman-Rank, Kendall-Tau, Winsorize ve Permütasyon tabanlı korelasyon katsayıları) farklı dağılım şekli, örnek genişliği (n) ve korelasyon yapıları (ρ) gibi değişik deneme koşulları altında 1.Tip hata olasılığı ve testin gücü bakımından karşılaştırılmıştır. Yapılan 100,000 simülasyon denemesi sonucunda Pearson-Moment korelasyon katsayısının dağılımın normal olmamasından oldukça olumsuz yönde etkilendiği ve bu olumsuz etkinin de küçük hacimli örneklerle çalışılması durumunda çok daha da belirginleştiği görülmüştür. Spearman-Rank ve Kendall-Tau korelasyon katsayıları ise dağılım şekli ne olursa olsun özellikle $n \leq 10$ olduğu koşullarda Pearson-Moment korelasyon katsayısı için çok da iyi alternatifler olmadığı görülmüştür. Diğer taraftan genel olarak dağılım şekli, korelasyon yapısı ve örnek hacmi ne olursa olsun Permütasyon ve Winsorize korelasyon katsayılarının daha güvenilir sonuçlar verdiği görülmüştür. Sonuç olarak, Pearson-Moment korelasyon katsayısının varsayımlarının yerine gelmediği durumlarda Winsorize ve Permütasyon-tabanlı korelasyon katsayılarının Pearson-Moment korelasyon korelasyon katsayısı için daha iyi alternatifler olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Korelasyon Katsayısı, I.Tip Hata, Testin Gücü, Normallik, Simülasyon

ABSTRACT

COMPARISON OF SOME CORRELATION COEFFICIENTS IN TERMS OF TYPE I ERROR RATE AND TEST POWER.

ELİF TUĞRAN
ÇanakkaleOnsekiz Mart University
Graduate School of Natural and Applied Sciences
Master of Science Thesis in Animal Science
Advisor : Assoc. Prof. Dr. MEHMET MENDEŞ
19/06/2015, 35

In this thesis, some correlation coefficients namely Pearson-Moment, Spearman Rank, Kendal Tau, Permutation-based, and Winsorized correlation coefficients were compared in terms of Type I error rate and test power under different scenarios where the underlying distributions of the variables of interest, sample sizes and correlation patterns are varied. Results of 100,000 simulation runs showed that the type I error rate and power of Pearson-Moment correlation coefficient are negatively affected by the distribution shapes especially for small sample sizes. Spearman Rank and Kendal Tau correlation coefficients are not good alternatives to the Pearson-Moment correlation especially when sample sizes are small ($n \leq 10$). In general, Permutation-based and Winsorized correlation coefficients are more robust to distribution shapes and correlation patterns, regardless of sample size. In conclusion, when assumptions of Pearson correlation coefficient are not satisfied, Permutation-based and Winsorized correlation coefficients seem to be better alternatives.

Keywords: Correlation Coefficient, Type I Error Rate, Test Power, Normality, Simulation

İÇİNDEKİLER

Sayfa No

TEZ SINAV SONUÇ FORMU ii
İNTİHAL (AŞIRMA) BEYAN SAYFASI iii
TEŞEKKÜR iv
SİMGELER VE KISALTMALAR v
ÖZET vi
ABSTRACT vii
ŞEKİLLER DİZİNİ ix
ÇİZELGELER DİZİNİ xi
BÖLÜM 1 – GİRİŞ 1
BÖLÜM 2 – ÖNCEKİ ÇALIŞMALAR 3
BÖLÜM 3 – MATERYAL VE YÖNTEM 5
3.1. Materyal 5
3.2. Yöntem 7
3.2.1. Pearson-moment korelasyon katsayısı 7
3.2.2. Spearman-Rank korelasyon katsayısı 8
3.2.3. Kendall-Tau korelasyon katsayısı 9
3.2.4. Winsorize korelasyon katsayısı 9
3.2.5. Permütasyon temelli pearson korelasyon katsayısı 10
BÖLÜM 4 – ARAŞTIRMA BULGULARI VE TARTIŞMA 12
4.1. Gerçek Veri Seti Bulguları 12
4.1.1. Pearson-moment korelasyon katsayısı 12
4.1.2. Spearman-Rank korelasyon katsayısı 13
4.1.3. Kendall-Tau korelasyon katsayısı 13
4.1.4. Winsorize korelasyon katsayısı 15
4.1.5. Permütasyon temelli pearson korelasyon katsayısı 16
4.2. Simülasyon Bulguları 18
4.3.Tartışma 25
BÖLÜM 5 – SONUÇLAR VE ÖNERİLER 29
KAYNAKLAR 32
ÖZGEÇMİŞ 36

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 4.2.l. İki değişkenli normal dağılımdan elde edilen PM; SR, WK, KT korelasyon katsayılarının ρ=0,60 elde edilen örneklem dağılımları.....22
Şekil 4.2.1.1. Pearson korelasyon katsayısının n=5 histogramı…….....22
Şekil 4.2.1.2. Pearson korelasyon katsayısının n=20 histogramı……..22
Şekil 4.2.1.3. Pearson korelasyon katsayısının n=100 histogramı……...…....22
Şekil4.2.1.4. Winsorized korelasyon katsayısının n=5 histogramı ……..…….....22
Şekil 4.2.1.5. Winsorized korelasyon katsayısının n=20 histogramı ….…..……22
Şekil 4.2.1.6. Winsorized korelasyon katsayısının n=100 histogramı ..…….....22
Şekil 4.2.1.7. Spearman–Rank korelasyon katsayısının n=5 histogramı …….....22
Şekil 4.2.1.8. Spearman–Rank korelasyon katsayısının n=20 histogramı…….....22
Şekil 4.2.1.9. Spearman–Rank korelasyon katsayısının n=100 histogramı……...22
Şekil 4.2.1.10. Kendall–Tau korelasyon katsayısının n=5 histogramı …….....22
Şekil 4.2.1.11. Kendall–Tau korelasyon katsayısının n=20 histogramı ……....22
Şekil 4.2.l.12. Kendall–Tau korelasyon katsayısının n=100 histogramı……22
Şekil 4.2.2. İki değişkenli ve 10 S.D' li t dağılımından elde edilen PM; SR, WK, KT korelasyon katsayılarından elde edilen örneklem dağılımları……...23
Şekil 4.2.2.1. Pearson korelasyon katsayısının n=5 histogramı 23
Şekil 4.2.2.2. Pearson korelasyon katsayısının n=20 histogramı …….....23
Şekil 4.2.2.3. Pearson korelasyon katsayısının n=100 histogramı …….....23
Şekil 4.2.2.4. Spearman–Rank korelasyon katsayısının n=5 histogramı…….....23
Şekil 4.2.2.5. Spearman–Rank korelasyon katsayısının n=20 histogramı.....23
Şekil 4.2.2.6. Spearman–Rank korelasyon katsayısının n=100 histogramı ….....23
Şekil 4.2.2.7. Winsorized korelasyon katsayısının n=5 histogramı……...23
Şekil 4.2.2.8. Winsorized korelasyon katsayısının n=20 histogramı ……...….23
Şekil 4.2.2.9. Winsorized korelasyon katsayısının n=100 histogramı …….....23
Şekil 4.2.2.10. Kendall–Tau korelasyon katsayısının n=5 histogramı……...23
Şekil 4.2.2.11. Kendall–Tau korelasyon katsayısının n=20 histogramı ……....23
Şekil 4.2.2.12. Kendall–Tau korelasyon katsayısının n=100 histogramı ..……....23
Şekil 4.2.3. İki değişkenli lognormal dağılımdan elde edilen PM; SR, WK, KT korelasyon katsayılarının örneklem dağılımları…….....24

Şekil 4.2.3.1. Pearson korelasyon katsayısının n=5 histogramı …….....24
Şekil 4.2.3.2. Pearson korelasyon katsayısının n=20 histogramı …….....24
Şekil 4.2.3.3. Pearson korelasyon katsayısının n=100 histogramı ……..…..24
Şekil 4.2.3.4. Spearman–Rank korelasyon katsayısının n=5 histogramı …….....24
Şekil 4.2.3.5. Spearman–Rank korelasyon katsayısının n=20 histogramı …..…..24
Şekil 4.2.3.6. Spearman–Rank korelasyon katsayısının n=100 histogramı …..…24
Şekil 4.2.3.7. Winsorized korelasyon katsayısının n=5 histogramı …….....24
Şekil 4.2.3.8. Winsorized korelasyon katsayısının n=20 histogramı ……..….24
Şekil 4.2.3.9. Winsorized korelasyon katsayısının n=100 histogramı …….....24
Şekil 4.2.3.10. Kendall–Tau korelasyon katsayısının n=5 histogramı …….....24
Şekil 4.2.3.11. Kendall–Tau korelasyon katsayısının n=20 histogramı ……..…24
Şekil 4.2.3.12. Kendall–Tau korelasyon katsayısının n=100 histogramı ….....24

ÇİZELGELER DİZİNİ

Sayfa No

Çizelge 3.1. 10 tane civcivin bel ağırlık(g) ve göğüs ağırlıklarına(g) ilişkin veriler……...5

Çizelge 3.1.1. Dikkate alınan deneme koşulları……...……7

Çizelge 4.1.2.1. Bel (gr) (X), göğüs (gr) (Y), rank (Bel) R_{xi}rank (Göğüs)D değerleri…….13

Çizelge 4.1.3.1.Bel (gr) (X), göğüs (gr) (Y), rank (Bel) R_{xi}rank (Göğüs) R_{xi}, C, D değerleri…….....13

Çizelge 4.4.1. Bel (gr) (X), göğüs (gr) (Y) değerleri ……...……15

Çizelge 4.1.5.1. Bazı permütasyon örnekleri (1-10)…….……16

Çizelge 4.1.5.2. Bazı permütasyon örnekleri (11-20)…….……...17

Çizelge 4.1.5.3. Korelasyon katsayıları, kritik tablo değerleri ve hipotez testi sonuçları……..……...17

Çizelge 4.2.1. İki değişkenli normal dağılımdan elde edilen özelliklerde 1.Tip hata ve Testin gücüne ilişkin değerler……...……19

Çizelge 4.2.2. İki değişkenli 10 serbestlik dereceli t- dağılımdan elde edilen özelliklerde 1.Tip hata ve Testin gücüne ilişkin değerler……...……20

Çizelge 4.2.3. İki değişkenli lognormal dağılımdan elde edilen özelliklerde 1.Tip hata ve Testin gücüne ilişkin değerler …….....……21

BÖLÜM 1
GİRİŞ

Uygulamada birçok araştırma, aynı deney ünitelerinden saptanan iki özellik arasındaki doğrusal ilişkinin derecesini hesaplamak amacıyla yürütülür. Bu amaçla Pearson-korelasyon katsayısından yaygın bir şekilde yararlanılır (Edgell ve Noon, 1984; Hayes, 1996; Bonett ve Wright, 2000; Zimmerman ve ark.2003). Ancak, Pearson-Moment korelasyon katsayısından beklenilen yararların sağlanabilmesi çalışılan veri gruplarında verilerin sürekli olması, normal dağılım göstermesi, veri gruplarında uç değerlerin bulunmaması, örnek genişliğinin yeterli olması ($n_i \geq 10$) gibi bir takım varsayımların yerine gelmesine bağlıdır (Wilcox, 2001; Warner, 2008; Triola, 2010; Witte ve Witte, 2010). Uygulamada bu varsayımların yerine gelmediği durumlarla oldukça sık karşılaşılmaktadır (Blairand Lawson, 1982; Fowler, 1987; Zimmerman and Zumbo, 1993: Bishara and Hittner, 2012). Buna rağmen uygulamada bu durum göz ardı edilerek doğrudan Pearson korelasyon katsayısı hesaplanmakta ve ona göre bir yoruma gidilmektedir. Bu durum özellikle hipotez testi aşamasında oldukça yanıltıcı sonuçların elde edilmesine ve gerçekte var olan önemli ilişkilerin doğru bir şekilde ortaya konulamamasına neden olmaktadır.

Pearson korelasyon katsayısının gerektirdiği varsayımların yerine gelmediği durumlarda bazen verilerin uygun bir transformasyona tabi tutularak normallikten sapmaların giderilmesine çalışılmaktadır. Ancak uygulamada böyle bir çözüm yoluna başvurulması birçok deneme koşulunda tatminkar sonuçlar vermemesinden dolayı pek önerilen bir çözüm yolu değildir (Osborne, 2002; Tabachnick ve Fidell, 2007). Böyle durumlarda Pearson-Moment korelasyon katsayısının yerine kullanılabilecek Spearman-Rank, Kendall-Tau, Winsorize, Permütasyon tabanlı korelasyon katsayısı, Ayıklanmış ortalamaya dayalı korelasyon katsayısı gibi değişik korelasyon katsayıları geliştirilmiştir (Rasmussen, 1989; Dunlap et al.1995; Wilcox, 2001). Bu korelasyon katsayıları arasında Spearman-Rank korelasyon katsayısı, Pearson korelasyon katsayısının varsayımlarının yerine gelmediği durumlarda uygulamada en yaygın olarak kullanılan korelasyon katsayısı olarak karşımıza çıkmaktadır. Halbuki bu korelasyon katsayısının özellikle sıralanmış (ordinal) ya da sıra değerleri verilmiş (rank) veri setleri için daha uygun oldukları bilinen bir gerçektir. Dolayısıyla gerek dağılım şekli ve gerek çalışılan veri gruplarında uç değerlerin bulunmasından etkilenmeyen korelasyon katsayıları olarak karşımıza çıkan

Winsorize ve Permütasayon temelli korelasyon katsayılarından etkin bir şekilde yararlanılmasında büyük yarar vardır (Calkins, 1974; Wilcox, 1993; Dunlap et al.1995; Hayes, 1996; Good, 2009; Bishara and Hittner, 2012; Mendeş, 2012). Ancak, bu korelasyon katsayılarından hiçbirisinin dikkate alınacak bütün deneme koşullarında güvenilir sonuçlar vermesi beklenen bir durum değildir. Bu durumda, söz konusu korelasyon katsayılarının hangi deneme koşulları için daha uygun olduklarının belirlenebilmesi için bunların farklı deneme koşullarındaki performanslarının (1.tip hata olasılığı ve testin gücü) ortaya konulması gerekir.

Bu tez çalışmasında Pearson-Moment, Spearman-Rank, Kendall-Tau, Winsorize ve Permütasyon temelli korelasyon olmak üzere beş korelasyon katsayısı 1.tip hata olasılığı ve testin gücü bakımından karşılaştırılmıştır. Böylece söz konusu korelasyon katsayılarının dikkate alınacak pek çok deneme koşulundaki performansları ortaya konulmuş ve bunun sonucunda da hangi korelasyon katsayısı ya da katsayılarının hangi deneme koşulları için kullanılmasının daha uygun olabileceğine yönelik bilgi üretilmiştir. Aynı zamanda bu korelasyon katsayılarının uygulanmasını göstermek üzere bu korelasyon katsayılarının nasıl hesaplandıkları gerçek bir veri seti üzerinden de gösterilmiştir.

BÖLÜM 2
ÖNCEKİ ÇALIŞMALAR

Abdullah (1990) yaptığı çalışmada Pearson, Kendall Tau, Spearman Rank ve En Küçük Medyan Kareler Yöntemine (LMS) dayalı korelasyon katsayılarını 1. tip hata ve testin gücü bakımından karşılaştırmıştır. Sonuç olarak; Pearson korelasyon katsayısının uç değerlerin varlığından oldukça fazla etkilendiğini rapor etmiştir. Ancak, Spearman Rank ve Kendall Tau korelasyon katsayılarının da uç değerlerin artması durumunda yeterince sağlam testler olmadıklarını ve bundan dolayı bu iki korelasyon katsayısının, Pearson korelasyon katsayısına sağlam birer alternatif olmadıklarını vurgulamıştır. Uç değerlerin fazla olduğu durumlarda LMS yöntemine dayalı korelasyon katsayısının tatmin edici sonuçlar verdiğini belirtmiştir.

Arnd ve Magnotta (2001) yaptıkları çalışmanın sonuçlarına göre Pearson korelasyon katsayısı için gereken ön şartlar yerine gelmediği durumlarda Kendall Tau korelasyon katsayısının 1. tip hatayı 0.05 seviyesinde koruma bakımından oldukça tatminkar sonuçlar verdiğini belirtmişlerdir.

Wilcox (2004) yaptığı çalışmada Pearson korelasyon katsayısının uç değerlerden oldukça fazla etkilendiğini, Spearman Rank ve Kendall Tau korelasyon katsayılarının ise rank değerlerini dikkate aldığı için verilerin genel yapısını yansıtmadığını belirtmiştir.

Chok (2010) yaptığı çalışmada Pearson, Spearman Rank ve Kendall Tau korelasyon katsayılarının permütasyon biçimlerini karşılaştırmıştır. Çalışmanın sonuçlarına göre normal olmayan sürekli dağılımlarda, Pearson korelasyon katsayısına dayalı permütasyon testinin diğer iki testin permütasyon karşılıkları kadar 1. tip hata olasılığını nominal seviyede (% 5.0) koruma eğiliminde olduğunu belirtmiştir. Dağılımın şekli normallikten aşırı derecede saptığında Pearson permütasyon testi bakımından gerçekleşen güç değerlerinin ($n \geq 50$ olduğunda) diğer iki teste göre daha düşük olduğunu vurgulamıştır. Ancak sürekli verilerin normal dağılım göstermemesinin Spearman Rank veya Kendall Taukorelasyon katsayılarının Pearson korelasyon katsayısının yerine kullanılabilmeleri için yeterli bir sebep olmadığını belirtmiştir.

Albayrak ve ark. (2011) yaptığı çalışmada Pearson, Spearman Rank ve Kendall Tau korelasyon katsayılarını sürekli dağılımlarda 1. tip hata olasılığı bakımından karşılaştırmışlardır. Sürekli dağılımlarda normallik ön şartının yerine gelmediği

durumlarda bile küçük hacimli örneklerle çalışılması durumunda (n<9) Pearson korelasyon katsayısının, Spearman Rank ve Kendall Tau korelasyon katsayılarına göre 1. tip hatayı daha iyi koruduğunu belirtmişlerdir. Örnek hacmi n>9 olduğunda ise Pearson ve Spearman Rank korelasyon katsayılarının Kendall Tau korelasyon katsayısına göre gerçekleşen 1. tip hata olasılığı bakımından daha iyi sonuçlar verdiklerini belirtmişlerdir.

BÖLÜM 3
MATERYAL VE YÖNTEM

3.1. Materyal

Bu çalışmanın materyalini Monte Carlo Simülasyon tekniği ile iki değişkenli normal dağılım, iki değişkenli 10 serbestlik dereceli t-dağılımı ve iki değişkenli log normal dağılımlı populasyonlardan üretilmiş tesadüf sayıları oluşturmuştur. Tesadüf sayılarının üretilmesinde Microsoft Developer Studio Fortran Power Stationun IMSL kütüphanesinden yararlanılmıştır.

Söz konusu korelasyon katsayılarının uygulamalarını göstermek amacıyla bu korelasyon katsayılarının gerçek bir veri seti üzerinde uygulaması da yapılmıştır. Bu amaçla etlik piliçlerde üç farklı yemleme programının (AD, R20, NF6) erkek civcivlerin büyümelerine etkisinin araştırıldığı bir denemeden elde edilen verilerin bir kısmı kullanılmıştır (Çizelge 3. 1.). 10 tane civcivin canlı ağırlık ve göğüs ağırlıklarına ilişkin veriler dikkate alınmış ve bu iki özellik arasındaki ilişkinin derecesi dikkate alınan bu beş korelasyon katsayısı ile belirlenmiş ve elde edilen sonuçlar karşılaştırılmıştır.

Çizelge 3.1. 10 tane civcivin bel ağırlık(g) ve göğüs ağırlıklarına(g) ilişkin veriler

Bel ağırlığı (X)	Göğüs ağırlığı (Y)
381.0	401.5
327.5	349.5
396.0	422,5
346.5	345.5
403.0	398.5
380.0	371.5
481.5	451.5
465.5	436.0
385.0	402.6
478.5	365.5

Söz konusu dağılımlar iki değişkenli standart normal dağılıma aşağıdaki transformasyonların uygulanması ile üretilmiştir.

İki değişkenli normal dağılım $\left(X^t, Y^t\right) \sim \mathrm{BVN}\left(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy}\right)$ aşağıdaki transformasyon ile üretilmiştir:

$$\begin{pmatrix} X^t \\ Y^t \end{pmatrix} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} + \left[\sum \begin{pmatrix} X \\ Y \end{pmatrix}\right]^{0.5} \tag{3.1}$$

Burada $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]$ ve $\sum = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$ **(3.1a)**

İki değişkenli 10 SD'li t-dağılımı $(X_T, Y_T) \sim t(\rho_{xy}, df)$ aşağıdaki gibi üretilmiştir:

$$\begin{pmatrix} X_T \\ X_T \end{pmatrix} = \begin{pmatrix} X^t \\ Y^t \end{pmatrix} \left[\frac{df}{\left(\frac{u}{v}\right)}\right]^{0.5} \quad \textbf{(3.2)}$$

Burada u ve v gamma dağılımından üretilen bağımsız değişkenlerdir.

İki değişkenli log-normal dağılım ise iki değişkenli normal dağılıma üssel transformasyonun uygulanması ile üretilmiştir:

$e^{(X^t, Y^t)} = (X_{log-n}, Y_{log-n}) \sim \text{LogNormal}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy})$ (Kotz et al., 2000; Ruscioand Kaczetow, 2008; Balakrishnan and Lai, 2009; Chok, 2010).

Bu çalışmada dikkate alınan deneme koşulları Çizelge 3.1'de topluca verilmiştir. Çizelge 3.1'deki deneme koşulları altında üretilen tesadüf sayıları üzerinden Pearson-Moment, Spearman-Rank, Kendall-Tau, Permütasyon-temelli ve Winsorize korelasyon katsayıları hesaplanmıştır. Daha sonra

$H_0 : \rho = 0$ hipotezi

$H_1 : \rho \neq 0$ hipotezine karşı test edilmiştir. Bu işlem 100,000 defa tekrar edilmiştir. Yapılan 100,000 simülasyon denemesi sonucunda ret edilen H_0 hipotezi sayısı belirlenmiş ve bu sayı toplam simülasyon sayısı olan 100,000'ne bölünerek her bir korelasyon katsayısı için 1.Tip hata olasılıkları elde edilmiştir.

Testin güç değerlerinin elde edilebilmesi için ise gerçekte $\rho \neq 0$ iken (0.25, 0.50, 0.75, 1.00) yapılan 100,000 simülasyon denemesi sonucunda da ρ'nun sıfırdan farklı olduğu hipotez sayısı belirlenmiş ve daha sonra bu sayı toplam simülasyon sayısı olan 100,000'ne bölünerek elde edilmiştir. Permütasyon-tabanlı korelasyon katsayısı için P-değerlerinin elde edilmesinde gerekli olan permütasyon sayısı 10,000 olarak alınmıştır.

Çizelge 3. 1. 1. Dikkate alınan deneme koşulları

Dağılım Şekli	Örnek Genişliği (n)	Populasyon Korelasyon Katsayısı (ρ)
İki-Değişkenli Normal İki-Değişkenli Log-normal İki-Değişkenli t(10)	5, 10, 20, 30, 40, 50, 70, 90, 100	0.00, 0.25, 0.50, 0.75, 0.90

3.2. Yöntem

Bu tez çalışmasında, iki değişken arasındaki doğrusal ilişkilerin derecesinin hesaplanması amacıyla kullanılan korelasyon katsayılarından Pearson-Moment korelasyon katsayısı, Spearman-Rank, Kendall-Tau, Winsorize ve Permütasyona temelli Pearson korelasyon katsayısı olmak üzere beş korelasyon katsayısı üzerinde durulmuştur.

3.2.1. Pearson-Moment korelasyon katsayısı

n gözlem bulunan iki değişkenli bir veri seti için, $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, Pearson- Moment korelasyon katsayısı r_{xy} :

$$r_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

$$= \frac{\sum d_x d_y}{\sqrt{(\sum d_x^2)(\sum d_y^2)}} \qquad \textbf{(3.3)}$$

şeklinde hesaplanır. Buradaki $\sum d_x d_y$ çarpımlar toplamını $\sum d_x^2$ ve $\sum d_y^2$ ise sırasıyla X ve Y özelliklerine ilişkin kareler toplamını göstermektedir.

Bu şekilde hesaplanan korelasyon katsayısı ya da doğrusal ilişkinin derecesinin istatistiksel olarak önemli olup olmadığının belirlenmesi amacıyla gerekli olan hipotez takımı:

$H_0 : \rho_{xy} = 0$

$H_1 : \rho_{xy} \neq 0$ şeklinde kurulur. H_0 hipotezinin test edilmesi için gerekli olan test istatistiği ise:

$$t = \frac{r_{xy}}{\sqrt{\frac{1 - r_{xy}^2}{n-2}}} \tag{3.4}$$

şeklinde hesaplanır. Bu şekilde hesaplanan t-test istatistiği (n-2) serbestlik dereceli t-dağılımı gösterir. Eğer bu şekilde hesaplanan t-değeri, (n-2) serbestlik dereceli t-tablo değerine eşit ya da büyük ise H_0 hipotezi ret edilir. Aksi durumda ise kabul edilir.

3.2.2. Spearman-Rank korelasyon katsayısı

Spearman-Rank korelasyon katsayısı hesaplanırken ranklar, yani sıra değerleri üzerinden işlemler yürütülür. R(x) ve R(y), (X_i, Y_i) değişken çiftlerinin ranklarını göstermek üzere Spearman-Rank korelasyon katsayısı aşağıdaki gibi hesaplanır.

$$r_s = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\left(\sum_{i=1}^{n} R(x_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)\left(\sum_{i=1}^{n} R(y_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)}} \tag{3.5}$$

Gözlemler arasında aynı ranklı değerler yoksa X ve Y gibi iki değişkene ilişkin Spearman-Rank korelasyon katsayısı gözlemlerin sıra numaraları üzerinden Pearson korelasyon katsayısında olduğu gibi hesaplanır.

Yani

$$r_s = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{3.6}$$

şeklinde hesaplanır. Eğer çalışılan veri setlerinde aynı ranklı değerler varsa bu durumda Spearman-Rank korelasyon katsayısı $d_i = X_i - Y_i i = 1,2,\dots,n$ göstermek üzere

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{3.7}$$

eşitliği ile hesaplanır.

$H_0 : \rho_{xy} = 0$ hipotezini test etmek amacıyla hesaplanan Spearman-Rank korelasyon katsayısı, bu amaçla özel hazırlanmış olan tablo değerinden büyük ya da eşit ise H_0 hipotezi ret edilir.

3.2.3. Kendall-Tau korelasyon katsayısı

(X_i, Y_i) değişken çiftleri arasındaki Kendall-Tau korelasyon katsayısı r_τ (4.1) no'lu eşitlikteki gibi hesaplanır (Knight, 1966).

$$\frac{\textbf{uyumlu gözlem çifti sayısı} - \textbf{uyumsuz gözlem çifti sayısı}}{\frac{1}{2}n(n-1)} \tag{3.8}$$

Eğer (X_i, Y_i) ve (X_j, Y_j) gözlemler çiftlerinin $X_i > X_j$ ve $Y_i > Y_j$ ya da $X_i < X_j$ ve $Y_i < Y_j$ şartlarını sağlıyorlarsa uyumlu gözlemler çifti, sağlamıyorlarsa yani eğer $X_i > X_j$ ve $Y_i < Y_j$ ya da $X_i < X_j$ ve $Y_i > Y_j$ ise uyumsuz gözlemler çifti olarak adlandırılır..

$H_0 : \rho_{xy} = 0$ hipotezini test etmek amacıyla gerekli olan Z test istatistiği

$$Z = \frac{3r_\tau\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} \tag{3.9}$$

şeklinde hesaplanır. Eğer $Z \geq 1.96$ ise H_0 hipotezi ret edilir. Aksi durumda ise H_0 hipotezi kabul edilir.

3.2.4. Winsorize korelasyon katsayısı

Pearson-Moment korelasyon katsayısı uç değerlerden oldukça olumsuz yönde etkilenmektedir. Wilcox (2001) bu gibi durumlarda Winsorize korelasyon katsayısının hesaplanmasını önermiştir. Dolayısıyla Winsorize korelasyon katsayısı özellikle veri gruplarında uç değerlerin bulunması durumunda kullanılması önerilen bir korelasyon katsayısıdır.

İki özellik arasındaki Winsorize korelasyon katsayısının hesaplanması için öncelikle bu özellikler bakımından en büyük ve en küçük gözlem değerleri bulunur. Daha sonra her

iki taraftaki en büyük ve en küçük gözlem değerlerinin yerlerine bunlara en yakın gözlem değerleri konulur.Bu şekilde elde edilen yeni değişken çiftleri arasında hesaplanan Pearson korelasyon katsayısına winsorize korelasyon katsayısı denir. Yani $(X_1, Y_1), (X_2, Y_2),...,(X_n, Y_n)$ iki değişkenli bir dağılımdan rastgele alınan bir örnek olmak üzere X_i değerleri W_i değerlerine, Y_i değerleri de V_i değerlerine dönüştürülmek üzere Winsorize korelasyon katsayısı r_w, winsorize edilmiş veri seti üzerinden Pearson-Moment korelasyon katsayısının hesaplanması ile elde edilir. Yani:

$$r_w = \frac{\sum (W_i - \overline{X}_W)(V_i - \overline{Y}_W)}{\sqrt{\sum (W_i - \overline{X}_W)^2 \sum (V_i - \overline{Y}_W)^2}} = \frac{\sum d_W d_V}{\sqrt{\sum d_W^2 \sum d_V^2}} \qquad \textbf{(3.10)}$$

Buradaki $\overline{X}_W$ ve $\overline{Y}_W$ X ve Y değişkenlerinin winsorize ortalamalarını göstermektedir. γ Winsorize yüzdesini göstermek üzere ($g = \gamma n$), r_w istatistiği (n-2g-2) serbestlik dereceli t-dağılımı gösterir (Wilcox, 1993).

3.2.5. Permütasyon temelli pearson korelasyon katsayısı

Özellikle çalışan örnek genişliğinin küçük olduğu durumlar için daha pratik olan Permütasyon-tabanlı korelasyon katsayısının hesaplanma aşamaları aşağıdaki gibidir:

a) İlk olarak orijinal veri seti üzerinden X ve Y özelliklerine ilişkin Pearson korelasyon katsayısı (r_{xy}) hesaplanır.

b) X ve Y özelliklerine ilişkin gözlemlerin kendi içindeki tüm permütasyon örnekleri (n!) oluşturulur. Örnek genişliği arttıkça oluşturulacak permütasyon örneği sayısı artık 100,000'lerle ifade edilecek kadar artar.

c) Oluşturulan her permütasyon örneği için Pearson korelasyon katsayıları (r_{pxy}) yeniden hesaplanır.

d) Permütasyon örneklerinden hesaplanan korelasyon katsayılarından, en başta orijinal veriler üzerinden hesaplanılan Pearson korelasyon katsayılarına eşit ya da büyük olanların sayısı belirlenir. Daha sonra bu sayı, permütasyon sayısına (N) bölünerek P-değeri elde edilir. Bulunan P-değeri 0.05 ile karşılaştırılarak $H_0 : \rho_{xy} = 0$ hipotezi test edilir.

Yani:

$$P \cong \frac{\left|r_{pxy}\right| \geq \left|r_{xy}\right|}{N} \quad \text{(Tanizaki, 2004)} \qquad \textbf{(3.11)}$$

BÖLÜM 4
ARAŞTIRMA BULGULARI VE TARTIŞMA

4.1.Gerçek Veri Seti Bulguları

Söz konusu korelasyon katsayılarının uygulamalarını göstermek amacıyla bu korelasyon katsayılarının gerçek bir veri seti üzerinde uygulaması da yapılmıştır.

4.1.1.Pearson-Moment korelasyon katsayısı

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}$$

$$\sum d_x d_y = (381.0)(327.5) + \ldots + (478.5)(365.5) - \frac{(4044.5)(3944.6)}{10} = 11078.5$$

$$\sum d_x^2 = 381.0^2 + 327.5^2 + \ldots + 478.5^2 - \frac{4044.5^2}{10} = 26026{,}2$$

$$\sum d_Y^2 = 401.5^2 + 349.5^2 + \ldots + 365.5^2 - \frac{3944.6^2}{10} = 11681{,}9$$

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}} = 0.635 = \%63.5$$

Bu şekilde hesaplanan korelasyon katsayısının önemlilik testi (n-2=10-2=8) serbestlik dereceli t-dağılımı gösterir. t-değeri ise:

$$t = \frac{r_{xy}}{S_r} \qquad \textbf{(4.1)}$$

$$S_r = \sqrt{\frac{1 - r_{xy}^2}{n-2}} = \sqrt{\frac{1 - 0.635^2}{8}} = 0.273 \text{ ve}$$

$$t = \frac{0.635}{0.273} = 2.325$$

2.325 olarak bulunan t-değeri n-2=10-2=8 serbestlik dereceli t-tablo değeri olan 2.306'dan daha büyük olduğu için Ho hipotezi ret edilir.

4. 1. 2. Spearman Rank korelasyon katsayısı

Çizelge 4. 1. 2. 1. Bel (gr) (X), göğüs (gr) (Y), rank (Bel), rank (Göğüs), D değerleri

Bel(gr) (X)	Göğüs(gr) (Y)	Rank (Bel) X	Rank (Göğüs) Y	D (X_i-Y_i)	D^2
381.0	401.5	4	6	-2	4
327.5	349.5	1	2	-1	1
396.0	422.5	6	8	-2	4
346.5	345.5	2	1	1	1
403.0	398.5	7	5	2	4
380.0	371.5	3	4	-1	1
481.5	451.5	10	10	0	0
465.5	436.0	8	9	-1	1
385.0	402.6	5	7	-2	4
478.5	365.5	9	3	6	36
					$\sum d_i^2 = 56$

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(56)}{10(99)} = 0.661 = \%66.1 \qquad \textbf{(4.2)}$$

0.661 olarak hesaplanan Spearman-Rank korelasyon katsayısı kritik tablo değeri olan 0.648'den daha büyük olduğu için Ho hipotezi ret edilir. Dolayısıyla bu iki özellik arasında pozitif yönde %66.1'lik önemli bir doğrusal ilişki bulunmaktadır.

4. 1. 3. Kendall Tau korelasyon katsayısı

Bir veri setinde Kendall-Tau korelasyon katsayısının hesaplanabilmesi için öncelikle uyumlu (C) ve uyumsuz (D) çiftlerin sayılarının bulunması gerekir. Bu veri seti için uyumlu çiftlerin sayısı C= olarak, uyumsuz çiftlerin sayısı da D= olarak bulunur. Bu durumda Kendall Tau korelasyon katsayısı:

C:Uyum çiftlerin sayısı

D:Uyumsuz çiftlerin sayısını göstermek üzere aşağıdaki gibi hesaplanır.

$$r_{tau} = \frac{(C-D)}{(C+D)} \qquad \textbf{(4.3)}$$

Hesaplanan korelasyon katsayısının önemlilik testinin yapılmasında ise

$$Z = \frac{3 * r_{tau}\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} \qquad \textbf{(4.4)}$$

şeklinde hesaplanan Z-değerinden yararlanılır. Eğer hesaplan Z-değeri 1.96 değerinden daha büyük ya da bu değere eşit ise Ho hipotezi ret edilir.

Çizelge.4. 1. 3. 1. Bel(gr) (X),göğüs(gr) (Y),rank (Bel)R_{xi},rank (Göğüs)$R_{xi,}$ C,D değerleri

Bel(gr) (X)	**Göğüs(gr) (Y)**	**Rank (Bel) (R_{xi})**	**Rank (Göğüs) (R_{yi})**	**C**	**D**
327.5	349.5	1	2	8	1
346.5	345.5	2	1	8	0
380.0	371.5	3	4	6	1
381.0	401.5	4	6	4	2
385.0	402.6	5	7	3	2
396.0	422.5	6	8	2	2
403.0	398.5	7	5	2	1
465.5	436.0	8	9	1	1
478.5	365.5	9	3	1	0
481.5	451.5	10	10	-	-
				$\sum C = 35$	$\sum D = 10$

$$r_{tau} = \frac{(C-D)}{(C+D)} = \frac{35-10}{35+10} = 0.56 = \%56$$

$$Z = \frac{3 * r_{tau}\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} = \frac{3(0.56)\sqrt{90}}{\sqrt{50}} = 2.254$$

olarak bulunur. Z=2.254 değeri kritik değer olan 1.96'dan daha büyük olduğu için Ho hipotezi ret edilir. Dolayısıyla bu etlik piliçlerin bel ağırlıkları ile göğüs ağırlıkları arasında önemli düzeyde doğrusal bir bir ilişkinin bulunduğu sonucuna varılır.

4. 1. 4.Winsorize korelasyon katsayısı

Bu veri setinde Winsorize korelasyon katsayısının hesaplanması için gerekli olan winsorize yüzdesi olan γ -değeri %10 olarak belirlenmiştir. Buna göre yeni veri seti aşağıdaki gibi olur.

Çizelge 4.4.1. Bel (gr) (X), göğüs (gr) (Y) değerleri

Bel(gr) (X),	Göğüs(gr) (Y),
381.0	401.5
346.5	349.5
396.0	422.5
346.5	349.5
403.0	398.5
380.0	371.5
478.5	436.0
465.5	436.0
385.0	402.6
478.5	365.5

$$\sum d_x d_y = (381.0).(401.5) + (346.5).(349.5) + \ldots + (478.5).(365.5) - \frac{(4060.5).(3933.1)}{10} = 8692.0$$

$$\sum d_x^2 = 381.0^2 + 346.5^2 + \ldots + 478.5^2 - \frac{(4060.5)^2}{10} = 22984.2$$

$$\sum d_y^2 = 401.5^2 + 349.5^2 + \ldots + 365.5^2 - \frac{(3933.1)^2}{10} = 9765.0$$

$$r_w = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}} = \frac{8692.0}{\sqrt{(22984.2)(9765.0)}} = 0.580 = \%58.0 \qquad \textbf{(4.5)}$$

$$S_{rw} = \sqrt{\frac{(1 - rw_{xy}^2)}{(n-2)}} = \sqrt{\frac{(1 - 0.580^2)}{8}} = 0.288 \qquad \textbf{(4.6)}$$

$$T_W = \frac{r_w}{S_{rw}} = \frac{0.580}{0.288} = 2.014 \qquad \textbf{(4.7)}$$

γ Winsorize yüzdesini göstermek üzere($g = \gamma n$), r_w istatistiği (n-2g-2) serbestlik dereceli t-dağılımı gösterir. $g = \gamma n = 0.10*10 = 1$ olduğundan r_w istatistiği (n-2g-2)=10-2-2=6 serbestlik dereceli t-dağılımı gösterir. Buna göre 2.014 olarak hesaplanan Tw değeri, 6

serbestlik dereceli t-tablo değeri olan 2.447'den daha küçük olduğundan H_0 hipotezi ret edilir. Dolayısıyla bu iki özellik arasında %58.0 olarak bulunan korelasyon katsayısının istatistiksel olarak önemli olmadığı sonucuna varılır.

4. 1. 5. Permütasyon-Tabanlı korelasyon katsayısı

Buradaki veri setimiz için n=10 olduğundan bu veri setinden $10! = 10.9.8.7.6.5.4.3.2.1 = 3628800$ tane permütasyon örneğinin oluşturulması mümkündür. Uygulamada bu kadar permütasyon örneği üzerinden exact P-değerinin hesaplanması her zaman mümkün olamadığı için genel olarak 5,000 ya da 10,000 permütasyon örneği baz alınarak yaklaşık P-değeri hesaplanmaktadır. Ancak bu veri seti için P-değeri 3628800 tane permütasyon örneği üzerinden hesaplanmıştır. Dolayısıyla burada hesaplanan exact P-değeridir. Diğer taraftan bu tez çalışmasında oluşturulması mümkün olan 3628800 tane permütasyon örneğinin rapor edilmesi mümkün olamadığından bunlar arasından rastgele seçilen 20 tanesi rapor edilmiştir (Çizelge 4.1. 5. 1 ve Çizelge 4.1. 5. 2).

Bu veri seti için Permütasyon temelli korelasyon katsayısı için exact P-değeri:

$$P \cong \frac{\left|r_{pxy}\right| \geq \left|r_{xy}\right|}{N} = 0.026 \qquad \textbf{(4.8)}$$

olarak bulunmuştur. Bu değer kritik değer olan 0.05'ten daha küçük olduğu için Ho hipotezi ret edilir.

Çizelge 4. 1. 5. 1. Bazı permütasyon örnekleri (1-10)

B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)
466	350	403	381	452	399	436	403	479	380	396	350	372	366	346	366	366	436	366	385
372	385	466	399	479	381	385	482	372	402	452	346	482	381	328	385	482	466	466	402
479	328	402	347	372	396	350	403	385	381	479	482	346	350	436	466	402	346	452	403
452	347	403	350	380	350	381	479	396	423	372	402	402	380	482	396	372	347	479	372
396	399	396	385	403	347	328	396	346	466	381	385	385	423	402	452	381	423	396	381
366	482	380	366	482	385	399	402	350	403	436	347	452	328	381	380	399	385	346	350
403	436	328	372	366	466	423	380	366	399	399	423	479	466	399	403	396	403	482	347
380	423	479	423	403	346	347	372	328	347	403	466	403	403	403	350	350	328	436	403
402	381	482	466	436	423	366	403	452	436	366	403	396	347	347	479	479	380	399	380
346	403	436	452	402	328	466	452	403	482	380	328	399	436	423	372	452	403	328	423
r_{xy}=0,762		r_{xy}=0,700		r_{xy}=0,025		r_{xy}=0,257		r_{xy}=0,134		r_{xy}=0,231		r_{xy}=0,270		r_{xy}=0,039		r_{xy}=0,435		r_{xy}=0,181	

Çizelge 4. 1. 5. 2. Bazı permütasyon örnekleri (11-20)

B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)	B(X)	G(Y)
346	403	436	346	402	452	381	403	366	372	452	366	396	399	366	423	436	372	399	403
366	328	396	466	372	482	403	380	350	380	479	372	479	328	328	436	350	346	479	366
381	372	402	328	350	366	423	479	452	328	423	380	346	436	403	380	466	381	380	482
466	479	385	372	396	328	466	346	466	396	328	385	466	347	482	402	403	403	423	436
452	403	403	482	423	399	385	328	399	385	346	381	381	423	479	347	328	380	372	385
402	423	403	399	380	466	482	399	346	403	403	350	372	402	350	399	452	402	396	328
399	396	350	347	436	347	402	347	402	482	403	396	403	452	403	385	347	479	402	346
350	482	366	452	479	385	366	372	436	381	466	482	380	366	452	346	423	482	347	452
436	347	423	479	403	403	436	452	423	347	347	399	385	350	466	381	385	399	381	350
380	385	381	380	381	346	350	396	403	479	402	436	482	403	372	396	396	366	403	466
r_{xy}=0,135		r_{xy}=0,120		r_{xy}=0,201		r_{xy}=0,163		r_{xy}=0,189		r_{xy}=0,179		r_{xy}=0,497		r_{xy}=0,709		r_{xy}=0,009		r_{xy}=0,260	

Çizelge 4. 1 5. 3. Korelasyon katsayıları kritik tablo değerleri ve hipotez testi sonuçları

Korelasyon katsayısı	**Değer**	**Kritik tablo değeri**	**Karar**
Pearson	0.635	2.306	H_0:ret
Spearman	0.661	0.648	H_0:ret
Kendall	0.560	1.96	H_0:ret
Winsorize	0.580	2.447	H_0:kabul
Permütasyon	0.026	0.05	H_0:ret

Çizelge 4. 1. 5. 3'den de görüleceği üzere Winsorize korelasyon katsayısı hariç, diğer korelasyon katsayılarına ilişkin H_0 hipotezleri ret edilmiştir. Dolayısıyla Pearson, Spearman-Rank, Kendall-Tau ve Permütasyon tabanlı korelasyon katsayılarından yararlanılarak bu 10 etlik pilicin göğüs ve sırt ağırlıkları arasında hesaplanan korelasyon katsayıları istatistiksel olarak önemli bulunurken, Winsorize korelasyon katsayısından yararlanılması durumunda söz konusu iki özellik arasındaki korelasyon katsayısının istatistiksel olarak önemli olmadığı sonucuna varılmıştır.

4. 2. Simülasyon Bulguları

Dikkate alınan deneme koşulları altında Pearson-Moment (PM), Spearman-Rank (SR), Kendall-Tau (KT), Permütasyon-tabanlı (PT) ve Winsorize (WK) korelasyon katsayılarına ilişkin gerçekleşen 1.Tip hata olasılıkları ile testin güç değerleri Çizelge 4. 2. 1 –Çizelge 4. 2. . 3'te topluca verilmiştir. Örneklerin iki değişkenli normal dağılım gösteren populasyonlardan alınmaları durumunda (ki bu durum Pearson-Moment korelasyon katsayısı için ideal durumdur), örnek genişliğinin çok küçük olduğu durumlar hariç (n=5) genel olarak bütün korelasyon katsayıları bakımından gerçekleşen 1.Tip hata olasılıklarının %5 civarında olduğu görülmüştür (Çizelge 4. 2. .1). Bu deneme koşullarında PM korelasyon katsayısı bakımından gerçekleşen 1.Tip hata olasılıkları genel olarak %4.94-5.11 arasında değişirken, SR, KT, WK ve PT korelasyon katsayıları bakımından gerçekleşen 1.Tip hata olasılıklarının sırasıyla %1.74-5.04, %1.62-5.07, %4.90-5.03 ve %5.03-5.18 arasında değiştiği görülmüştür. Bu korelasyon katsayıları testin güç değerleri bakımından karşılaştırıldığında ise; beklenildiği üzere bütün korelasyon katsayılarının güç değerlerinin etki büyüklüğü ve örnek genişliğinin artmasına paralel olarak arttığı ve giderek birbirine yaklaştığı görülmüştür. Özellikle $n \geq 30$ ve $\rho \geq 0.75$ iken bütün korelasyon katsayıları bakımından gerçekleşen güç değerleri genel olarak ya tam olarak %100 değerine ulaşmış ya da giderek %100'e yaklaşmıştır.

Örneklerin çekildikleri populasyonların dağılımlarının iki değişkenli 10 SD'li t-dağılımı olması durumunda ampirik olarak gerçekleşen 1.Tip hata olasılıkları ve testin güç değerleri Çizelge 4. 2. 2'de verilmiştir. Çizelge 4. 2. 2'den de görüleceği üzere bu koşullarda elde edilen 1.Tip hata olasılıkları, dağılımların iki değişkenli normal olması durumunda (Çizelge 4. 2. 1) gerçekleşen 1.Tip hata olasılıklarına oldukça benzerdir. Bu beklenen bir durumdur. Çünkü 10 serbestlik dereceli iki değişkenli t-dağılımı, iki değişkenli normal dağılıma oldukça benzer bir dağılımdır.

Örneklerin alınmış oldukları populasyonların dağılımlarının iki değişkenli log normal olması durumunda ise dikkate alınan korelasyon katsayıları bakımından gerçekleşen 1.Tip hata olasılıkları ve testin güç değerleri Çizelge 4. 2. 3'te topluca verilmiştir. Çizelge 4. 2. 3'ten de görüleceği üzere özellikle Pearson-Moment korelasyon katsayısı dağılımın normal olmamasından ya da normallikten sapmaların belirgin düzeylerde olmasından oldukça olumsuz yönde etkilenmiştir.

Bu deneme koşullarında Pearson-Moment korelasyon katsayısı bakımından gerçekleşen 1.Tip hata olasılıklarının çalışılan örnek hacminin 100 olması durumunda bile %6.34 olarak tahmin edilmiştir. Bu durumun testin gücüne yansıması ise özellikle n<10 olduğu durumlarda belirgindir.

Çizelge 4. 2. 1. İki-değişkenli normal dağılımdan elde edilen özelliklerde 1.Tip hata ve Testin gücüne ilişkin değerler

Örneklem hacmi	**Korelasyon katsayısı**	**Rho (ρ)**				
		1. Tip hata	**Testin gücü**			
		0	**0.25**	**0.50**	**0.75**	**0.90**
5	Pearson	4.96	9.10	16.15	30.02	58.55
	Spearman-Rank	1.74	2.14	3.13	5.62	13.62
	Kendal-Tau	1.62	2.09	3.24	5.75	13.55
	Permütasyon	5.94	10.81	17.21	30.14	55.66
	Winsorize	4.90	9.42	16.44	30.11	58.74
10	Pearson	5.11	13.22	31.97	62.88	92.93
	Speraman-Rank	4.92	7.61	17.36	38.81	75.85
	Kendal-Tau	4.21	7.58	16.91	38.05	74.91
	Permütasyon	5.63	16.57	36.82	68.76	95.18
	Winsorize	5.03	13.66	31.96	61.49	91.57
20	Pearson	4.94	22.35	57.77	91.58	99.77
	Spearman-Rank	4.90	11.96	37.15	76.70	98.59
	Kendal-Tau	4.87	11.55	36.45	75.82	98.31
	Permütasyon	5.55	24.34	57.84	91.66	99.79
	Winsorize	5.02	20.21	51.62	86.07	99.56
30	Pearson	4.98	28.35	72.92	97.82	100.00
	Spearman-Rank	4.97	17.13	55.14	92.51	99.95
	Kendal-Tau	4.90	16.60	54.88	92.36	99.92
	Permütasyon	5.03	30.35	74.58	98.06	100.00
	Winsorize	5.01	26.15	67.69	96.20	100.00
50	Pearson	4.98	40.28	90.50	100.00	100.00
	Spearman-Rank	5.04	37.36	86.22	99.76	100.00
	Kendal-Tau	5.05	26.11	78.81	99.39	100.00
	Permütasyon	5.04	42.53	90.54	100.00	100.00
	Winsorize	4.95	59.52	98.90	100.00	100.00
100	Pearson	4.99	64.70	99.86	100.00	100.00
	Spearman-Rank	4.96	60.00	99.05	100.00	100.00
	Kendal-Tau	5.07	47. 70	98.83	100.00	100.00
	Permütasyon	5.18	34.50	82.08	99.51	100.00
	Winsorize	4.98	59.52	99.15	100.00	100.00

Çizelge 4. 2. 2. İki-değişkenli 10 serbestlik dereceli t- dağılımdan elde edilen özelliklerde 1.Tip hata ve Testin gücüne ilişkin değerler

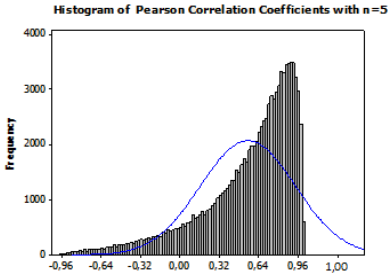
Örneklem hacmi	Korelasyon katsayısı	Rho (ρ)				
		1. Tip hata	Testin gücü			
		0	0.2	0.4	0.6	0.8
5	Pearson	5.08	9.06	16.17	28.41	53.05
	Spearman-Rank	1.69	2.09	3.21	5.46	12.01
	Kendal-Tau	1.61	2.04	3.17	5.92	13.47
	Permütasyon	5.88	10.14	18.04	29.17	51.56
	Winsorize	4.98	9.15	16.01	28.80	53.70
10	Pearson	5.11	13.22	30.32	58.34	89.65
	Speraman-Rank	4.93	7.36	17.13	37.12	72.21
	Kendal-Tau	4.65	7.30	17.00	38.29	75.05
	Permütasyon	5.07	13.95	34.22	64.42	92.07
	Winsorize	5.11	13.71	30.74	58.74	89.73
20	Pearson	4.98	20.07	52.44	87.21	99.72
	Spearman-Rank	4.91	11.62	35.61	74.88	98.36
	Kendal-Tau	4.78	11.54	36.33	75.47	98.69
	Permütasyon	4.97	21.03	51.75	88.69	99.72
	Winsorize	5.04	19.55	49.50	84.88	99.52
30	Pearson	4.98	26.50	68.26	96.72	100.00
	Spearman-Rank	5.06	16.41	53.36	91.18	99.92
	Kendal-Tau	4.92	16.52	54.71	92.55	100.00
	Permütasyon	4.81	26.65	68.71	97.43	100.00
	Winsorize	5.16	25.53	66.40	95.51	100.00
50	Pearson	4.99	37.57	87.12	100.00	100.00
	Spearman-Rank	5.03	36.49	86.54	99.67	100.00
	Kendal-Tau	4.93	26.09	78.95	99.41	100.00
	Permütasyon	5.33	38.54	87.52	100.00	100.00
	Winsorize	5.00	36.71	85.51	100.00	100.00
100	Pearson	4.99	60.02	99.33	100.00	100.00
	Spearman-Rank	4.95	58.54	98.51	100.00	100.00
	Kendal-Tau	5.06	47.72	98.47	100.00	100.00
	Permütasyon	5.25	60.11	98.98	100.00	100.00
	Winsorize	5.04	58.04	98.71	100.00	100.00

Bu tez çalışmasında aynı zamanda dikkate alınan deneme koşullarına göre hesaplanan korelasyon katsayılarının örnekleme dağılımları hakkında da bilgi edinilmiştir. Bunun için söz konusu deneme koşulları altında hesaplanan korelasyon katsayılarından pozitif ve negatif olanların frekansları belirlenmiştir. Daha sonra hesaplanan bu korelasyon katsayılarının histogramları çizilmiştir (Şekil 4.2.1- 4.2.3). Bu çalışma sonucunda dikkate alınan bütün deneme koşulları için oluşturulacak örnekleme dağılımı sayısı çok fazla (135

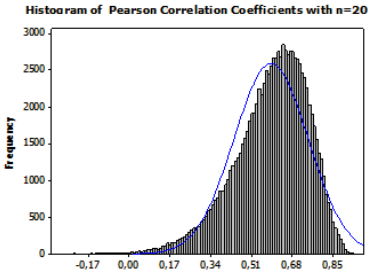
tane) olduğu için çalışmada sadece ρ≥0.50 ve n=5, 20 ve 100 olduğu durumlardaki örnekleme dağılımları rapor edilmiştir. Her bir örnekleme dağılımı 50,000 simülasyon sonucunda elde edilmiştir. Bu bulgular ışığında söz konusu korelasyon katsayılarının güç değerleri bakımından dolaylıda olsa karşılaştırılmasına imkan verecektir. Normalde pozitif frekans sayısı fazla olan korelasyon katsayısının daha güçlü olması beklenir. Beklenildiği gibi örnek genişliğinin artmasına paralel olarak pozitif korelasyon katsayısı sayısı da artmıştır (Şekil 4.2.1- 4.2.3).

Çizelge 4. 2. 3. İki-değişkenli lognormal dağılımdan elde edilen özelliklerde 1.Tip hata ve Testin gücüne ilişkin değerler

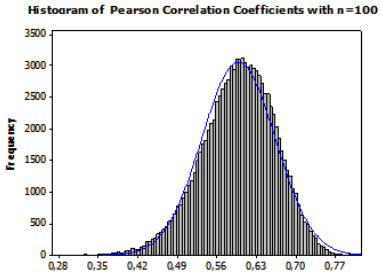
Örneklem hacmi	Korelasyon Katsayısı	Rho (ρ)				
		1. Tip hata	Testin gücü			
		0	0.2	0.4	0.6	0.8
5	Pearson	7.65	12.93	21.41	33.58	33.82
	Spearman-Rank	1.59	2.23	3.45	6.03	13.63
	Kendal-Tau	1.65	2.11	3.35	5.92	13.46
	Permütasyon	5.79	10.45	17.36	27.56	49.81
	Winsorize	7.51	13.77	21.52	33.64	55.09
10	Pearson	7.22	16.33	30.09	51.06	51.50
	Speraman-Rank	4.93	7.80	17.21	39.22	76.03
	Kendal-Tau	4.62	7.47	16.69	38.55	75.89
	Permütasyon	5.30	12.82	25.11	46.87	80.16
	Winsorize	5.36	13.61	28.84	54.71	87.28
20	Pearson	7.01	19.55	42.03	73.17	73.21
	Spearman-Rank	4.88	12.21	37.49	77.06	98.62
	Kendal-Tau	4.77	11.86	36.52	75.74	98.40
	Permütasyon	5.21	15.05	33.81	63.51	94.17
	Winsorize	5.25	18.55	46.16	80.19	99.07
30	Pearson	6.89	22.08	50.13	83.57	99.64
	Spearman-Rank	5.09	17.05	54.57	92.61	100.00
	Kendal-Tau	4.91	16.52	54.24	92.32	99.94
	Permütasyon	5.20	18.36	43.06	77.64	98.99
	Winsorize	5.21	24.11	60.54	93.18	100.00
50	Pearson	6.57	27.42	64.51	95.19	98.22
	Spearman-Rank	5.05	37.53	87.18	99.89	100.00
	Kendal-Tau	4.94	26.04	78.62	99.67	100.00
	Permütasyon	5.23	23.19	57.79	91.72	100.00
	Winsorize	5.17	33.28	81.01	99.54	100.00
100	Pearson	6.34	37.54	85.16	99.73	99.73
	Spearman-Rank	5.03	60.25	99.29	100.00	100.00
	Kendal-Tau	5.06	47.31	98.17	100.00	100.00
	Permütasyon	5.17	34.20	81.28	99.41	100.00
	Winsorize	5.13	52.50	98.26	100.00	100.00



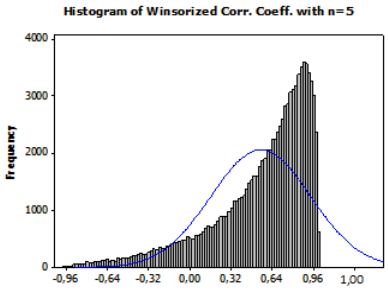
Şekil 4.2.1.1



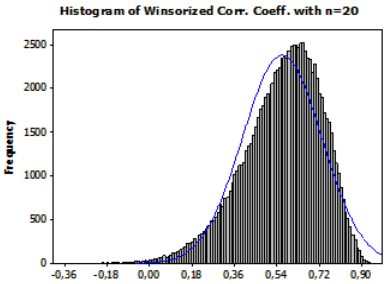
Şekil 4.2.1.2



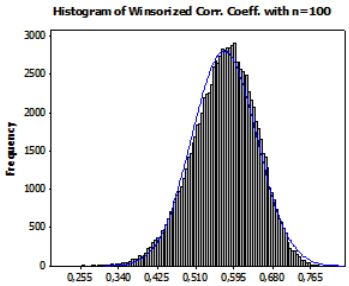
Şekil 4.2.1.3



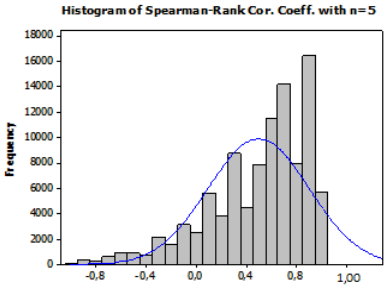
Şekil 4.2.1.4



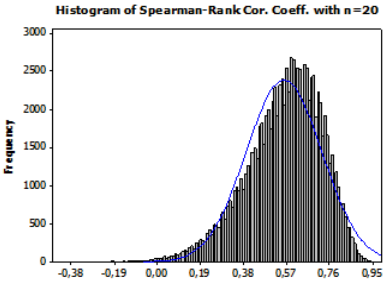
Şekil 4.2.1.5



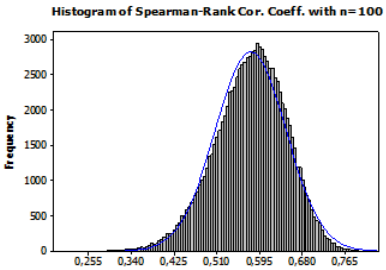
Şekil 4.2.1.6



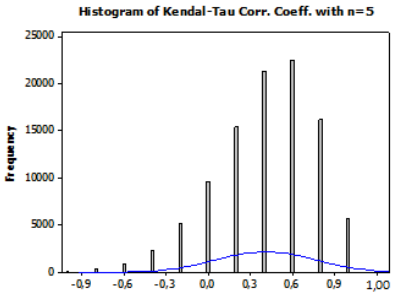
Şekil 4.2.1.7



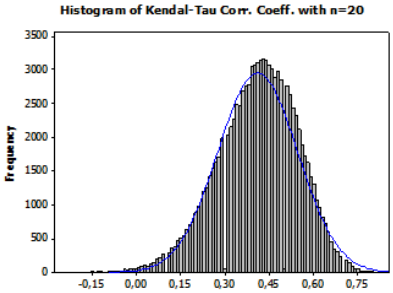
Şekil 4.2.1.8



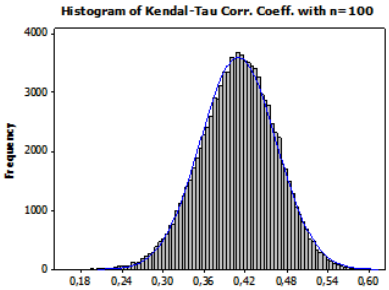
Şekil 4.2.1.9



Şekil 4.2.1.10

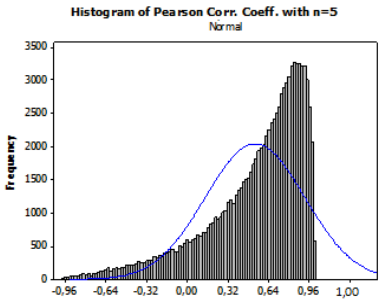


Şekil 4.2.1.11

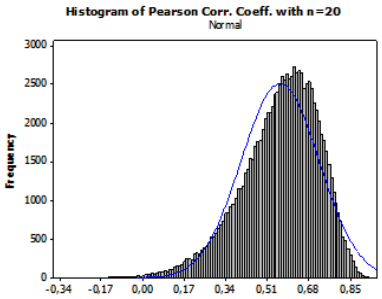


Şekil 4.2.1.12

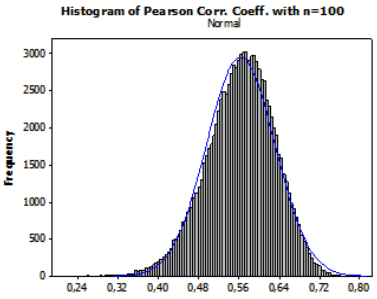
Şekil 4.2.1. İki değişkenli normal dağılımdan elde edilen PM; SR, WK, KT korelasyon katsayılarının ρ=0,60 elde edilen örneklem dağılımları



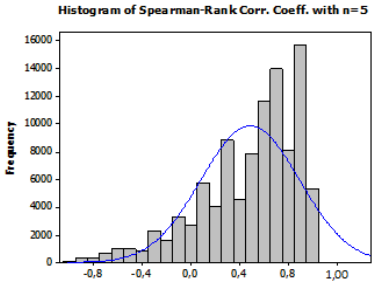
Şekil 4.2.2.1



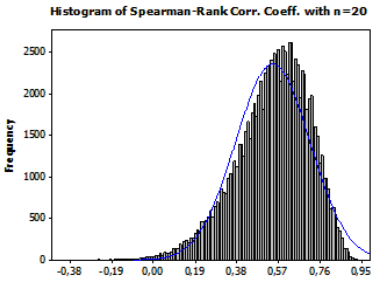
Şekil 4.2.2.2



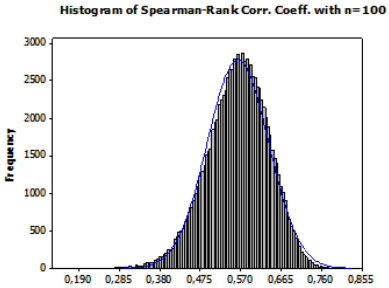
Şekil 4.2.2.3



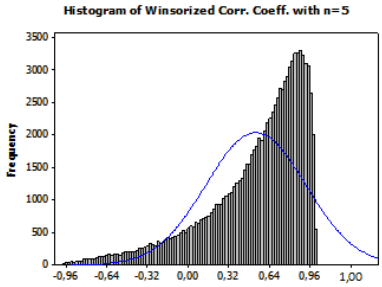
Şekil 4.2.2.4



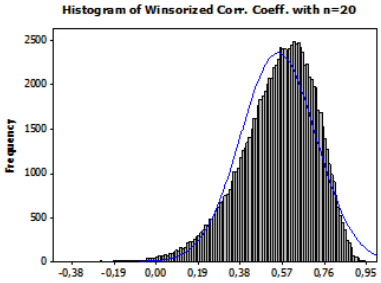
Şekil 4.2.2.5



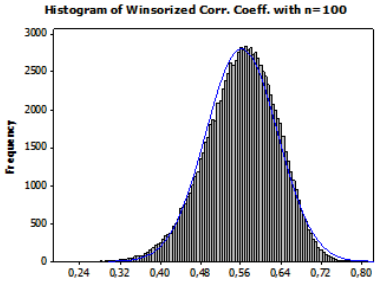
Şekil 4.2.2.6



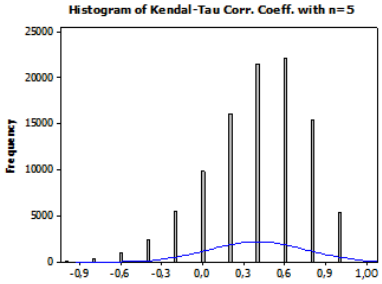
Şekil 4.2.2.7



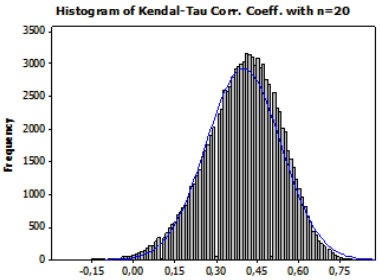
Şekil 4.2.2.8



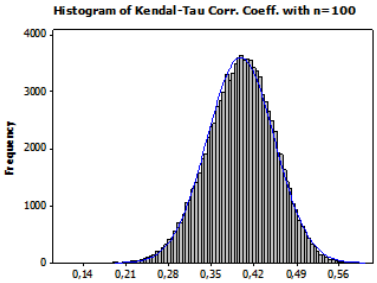
Şekil 4.2.2.9



Şekil 4.2.2.10

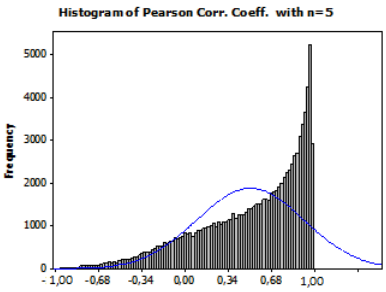


Şekil 4.2.2.11

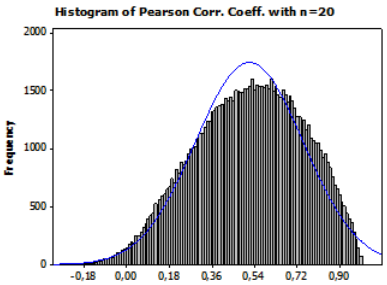


Şekil 4.2.2.12

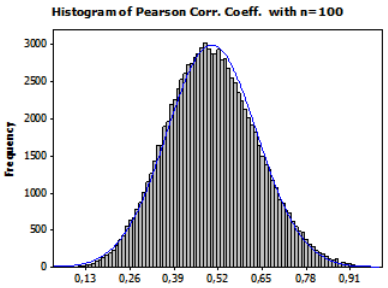
Şekil 4.2.2. İki değişkenli ve 10 S.D' li t dağılımından elde edilen PM; SR, WK, KT korelasyon katsayılarından elde edilen örneklem dağılımları



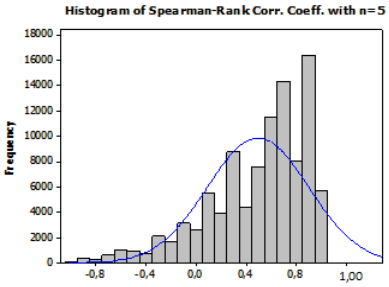
Şekil 4.2.3.1



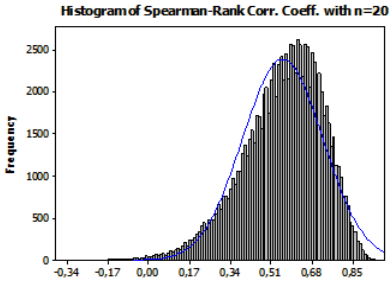
Şekil 4.2.3.2



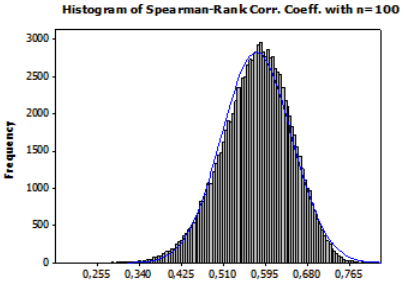
Şekil 4.2.3.3



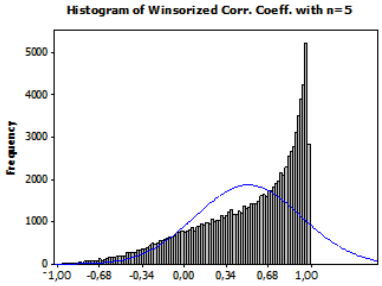
Şekil 4.2.3.4



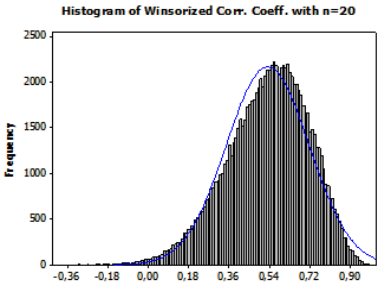
Şekil 4.2.3.5



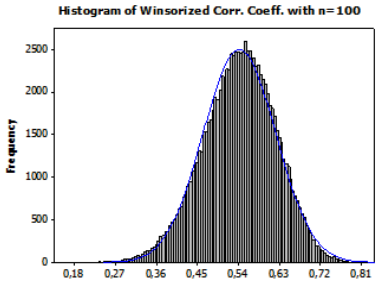
Şekil 4.2.3.6



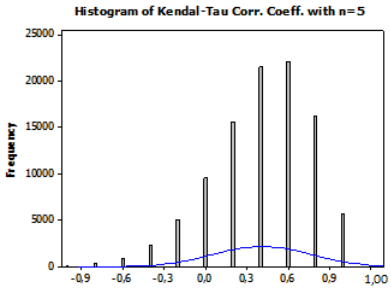
Şekil 4.2.3.7



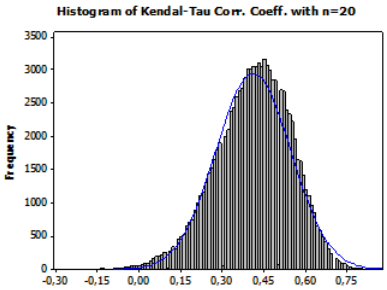
Şekil 4.2.3.8



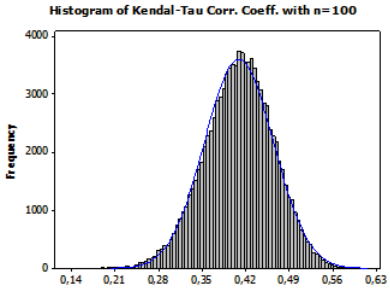
Şekil 4.2.3.9



Şekil 4.2.3.10



Şekil 4.2.3.11



Şekil 4.2.3.12

Şekil 4.2.3. İki değişkenli log normal dağılımdan elde edilen PM; SR, WK, KT korelasyon katsayılarının örneklem dağılımları

4.3. Tartışma

Uygulamada hemen hemen her alanda aynı deney ünitelerinden tespit edilen değişken çiftleri arasındaki doğrusal ilişkilerin derecesinin araştırıldığı pek çok çalışma yapılmaktadır. Mesela bir araştırıcı vücut ağırlığı ile sistolik kan basıncı arasındaki doğrusal ilişkinin derecesini araştırabilir. Benzer şekilde başka bir araştırıcı bitki boyu ile elde edilen ürün miktarı arasındaki doğrusal ilişkinin derecesi ile ilgilenebilir. Başka bir araştırıcı öğrencilerde haftalık olarak televizyon seyretme süresi ile bu öğrencilerin başarı not ortalamaları arasındaki ilişki ile ilgilenebilir. Bu amaçla kullanılmak üzere değişik ilişki katsayıları ya da korelasyon katsayıları geliştirilmiştir. Bu durumda araştırıcı merak ettiği ilişkinin derecesini güvenilir bir şekilde hesaplamasını sağlayacak korelasyon katsayısını belirlemek durumundadır. Zira her korelasyon katsayısının dikkate alınan bütün deneme koşullarında güvenilir sonuçlar vermesi beklenen durum değildir. Her ne kadar uygulamada bu amaçla en yaygın kullanılan korelasyon katsayının Pearson-Moment korelasyon katsayısı olduğu dikkati çekiyor olsa da yapılan pek çok çalışma bu korelasyon katsayısının varsayımlarından sapmaların olması ya da yerine gelmemesinden oldukça olumsuz yönde etkilendiğini ortaya koymuştur. Dolayısıyla bu gibi durumlarda Pearson korelasyon katsayısı yerine Spearman-Rank, Kendal-Tau, doğrusal olmayan transformasyon temelli Pearson korelasyon katsayısı, Permütasyon tabanlı korelasyon katsayısı ve Winsorize korelasyon katsayısı gibi alternatif korelasyon katsayılarının kullanılması elde edilecek sonuçların güvenilirliği bakımından büyük yarar vardır. Literatürler incelendiğinde bu alternatif korelasyon katsayılarından özellikle Spearman-Rank ve Kendal-Tau korelasyon katsayılarının kullanıldığı (özellikle Spearman-Rank) çalışmaların yoğunluğu dikkati çekmektedir. Halbuki her ne kadar bu iki korelasyon katsayısı ordinal ve rank verilmiş veri grupları için avantajlı katsayılar olsalar da sürekli veri setleri için performansları net olarak ortaya konulmamıştır. Aynı zamanda bu iki korelasyon katsayısının uygulamada yaygın olarak karşılaşılan deneme koşullarındaki performanslarının da yeteri kadar ortaya konulmamış olması başka bir eksikliktir. Bu gibi durumlarda uygulamada başvurulan diğer pratik bir çözüm yolu ise veri setlerinin uygun bir transformasyona tabii tutulduktan sonra elde edilecek transforme edilmiş veriler üzerinden Pearson korelasyon katsayısının hesaplanmasıdır (Kowalski and Tarter, 1969; Rasmussen, 1989; Rasmussen and Dunlap, 1991; Dunlap et al., 1995). Ancak, örneklerin alındıkları populasyonların dağılımlarının çok eğri ve örnek genişliğinin de küçük olduğu durumlarda transformasyonların kullanılması genel olarak tatmin edici sonuçlar verememektedir. Aynı zamanda transformasyonların kullanılması sonuçların yorumlama

aşamasında da bazı güçlüklülerle karşılaşılmasına neden olabilmektedir (Osborne, 2002; Tabachnick and Fidell, 2007). Dolayısıyla Pearson-Moment korelasyon katsayısının varsayımlarının yerine gelmediği durumlarda transformasyonlara başvurmak yerine, Winsorize korelasyon katsayısı ve Permütasyon tabanlı korelasyon katsayısı gibi bazı güçlü korelasyon katsayılarının kullanılması elde edilecek sonuçların güvenilirliği bakımından oldukça önemlidir. Bununla birlikte, bu korelasyon katsayılarının özellikle uygulamada yaygın olarak karşılaşılan deneme koşullarındaki performanslarının ortaya konulması için kapsamlı çalışmalara ihtiyaç vardır. Bu noktadan hareketle bu tez çalışmasında Pearson korelasyon katsayısının varsayımlarının yerine gelmediği durumlarda bu korelasyon katsayısı ile birlikte buna alternatif olarak kullanılabilen Spearman-Rank, Kendall-Tau, Winsorize ve Permütasyon-tabanlı korelasyon katsayılarının farklı deneme koşullarındaki performanslarının ortaya konulması amacıyla kapsamlı bir simülasyon çalışması yapılmıştır. Yapılan simülasyon çalışması sonucunda normallikten aşırı sapmaların olmaması halinde Pearson-Moment korelasyon katsayısının güvenilir sonuçlar verdiği, ancak normallikten belirgin sapmaların olmasından Pearson korelasyon katsayısının olumsuz yönde etkilendiği ve olumsuz etkinin de özellikle küçük hacimli örneklerle çalışılması durumlarında çok daha belirginleştiği görülmüştür.

Pearson-Moment korelasyon katsayısının varsayımlarının yerine gelmediği durumlarda her ne kadar uygulamada genel olarak Spearman-Rank korelasyon katsayısı en çok önerilen ve dolayısıyla da kullanılan korelasyon katsayısı olarak karşımıza çıkıyorsa da Spearman-Rank ve Kendall-Tau korelasyon katsayılarının bu çalışmada dikkate alınan deneme koşullarındaki performanslarının pek de iyi olmadığı görülmüştür. Fowler (1987) yaptığı çalışmada normal olmayan veri gruplarında Spearman-Rank korelasyon katsayısının Pearson korelasyon katsayısına göre daha güçlü olduğunu bildirmiştir. Benzer şekilde Zimmerman ve Zumbo (1993) yaptığı çalışma sonucunda Spearman-Rank korelasyon katsayısının karışık-normal ve normal olmayan dağılımlar için güçlü bir korelasyon katsayısı olduğunu rapor etmiştir.

Aynı zamanda bu korelasyon katsayısının üssel dağılımlı populasyonlarla çalışılması durumunda deneme başında kararlaştırılan 1.Tip hata olasılığını aynı düzeyde koruma bakımından Pearson-Moment korelasyon katsayısına göre daha güvenilir sonuçlar verdiğini bildirmişlerdir. Aslında pek çok çalışmada da belirtildiği üzere Spearman-Rank ve Kendal Tau korelasyon katsayıları özellikle çalışılan verilerin ordinal ya da rank

şeklinde elde edildiği veya ifade edildiği durumlar için uygun korelasyon katsayılarıdır (Siegel, 1957; Lieberson, 1964; Sheskin, 2007; Khamis, 2008).

Bu çalışmanın bulguları ile yukarıda tartışılan çalışma bulguları arasında gözlenen belirgin farkların çalışmalarda dikkate alınan deneme koşullarındaki farklılıklardan (mesela dağılım şekli, örnek genişliği, dikkate alınan simülasyon sayısı gibi) kaynaklandığı düşünülmektedir.

Dikkate alınan korelasyon katsayıları arasında Permütasyon tabanlı korelasyon katsayısı ile Winsorize korelasyon katsayısının dağılım şekli, örnek hacmi ve uç değerlere karşı en güçlü katsayılar oldukları görülmüştür.

Dolayısıyla Pearson-Moment korelasyon katsayısının varsayımlarının yerine gelmediği durumlarda bu iki korelasyon katsayının, Spearman-Rank ve Kendall-Tau korelasyon katsayılarına tercih edilmesinin daha uygun olduğu sonucuna varılabilir. Hayes (1996) yaptığı çalışmada özellikle normallik varsayımının yerine gelmediği durumlarda Permütasyon tabanlı testlerin kullanılmasının deneme başında kararlaştırılan 1.Tip hata olasılıklarının deneme sonunda da aynı düzeyde koruması bakımından oldukça güvenilir sonuçlar verdiğini bildirmiştir. Benzer şekilde Good (2005), Good (2009), Mielke ve Berry (2007) yaptıkları çalışmalar sonucunda ilgilenilen değişkenlerin dağılımlarının normal olmaması ve çalışılan örnek genişliklerinin de küçük olması durumlarında permütasyon tabanlı yöntemlerin kullanılmasının elde edilecek sonuçların güvenilirlikleri bakımından oldukça önemli olduğunu bildirmişler ve bu gibi durumlarda permütasyon tabanlı yöntemlerin kullanılmasını önermişlerdir.

Bununla birlikte Keller-McNulty ve Higgins (1987), Rasmussen (1989), Hayes (1996; 1998; 2000), Akkartal ve ark. (2010), bildirdikleri gibi permütasyon tabanlı yöntemlerin özellikle normallik varsayımının bozulduğu ve küçük hacimli örneklerle çalışılması durumlarında kararlaştırılan 1.Tip hata olasılıklarını deneme sonunda da aynı düzeylerde koruyabildikleri, ancak bu yöntemlerin birçok varsayımın bozulduğu deneme koşullarında her zaman güvenilir sonuçlar vermelerinin beklenilen bir durum olmadığının göz ardı edilmemesi gerektiğini ifade etmişlerdir.

Yapılan simülasyon çalışmaları sonucunda Pearson korelasyon katsayısının varsayımlarının yerine gelmediği durumlarda sürekli iki değişken arasındaki doğrusal ilişkinin derecesinin hesaplanmasında genel olarak Winsorize ya da Permütasyon-tabanlı

korelasyon katsayılarından yararlanılmasının daha uygun olduğu sonucuna varılabilir. Her ne kadar örnek hacminin artmasına paralel olarak oluşturulacak permütasyon örneklerindeki olağanüstü artış, Permütasyon-tabanlı korelasyon katsayısının hesaplanmasında bazı zorluklarla karşılaşmasına neden olabilirse de artık günümüz bilgisayar teknolojisinde bu zorluklarla rahatlıkla baş edilebilmektedir.

BÖLÜM 5
SONUÇLAR VE ÖNERİLER

Örneklerin çekildikleri populasyonların dağılımlarının, iki değişkenli normal dağılım gösteren populasyonlardan alınmaları durumunda örnek genişliğinin çok küçük olduğu durumlar hariç (n=5) genel olarak bütün korelasyon katsayıları bakımından gerçekleşen 1.Tip hata olasılıklarının %5 civarında olduğu görülmüştür (Çizelge 4. 2. 1).

n=5 olduğu durumlarda hesaba katıldığında S.R ve K.T haricindeki katsayıların yine %5 civarında olduğu görülmüştür. Bu açıdan iki değişkenli normal dağılımda her durumda P.K W.K ve P.T nin daha güçlü olduğu söylenebilir.

Bu korelasyon katsayıları testin güç değerleri bakımından karşılaştırıldığında ise; beklenildiği üzere bütün korelasyon katsayılarının güç değerlerinin etki büyüklüğü ve örnek genişliğinin artmasına paralel olarak arttığı ve giderek birbirine yaklaştığı görülmüştür. Özellikle $n \geq 30$ ve $\rho \geq 0.75$ iken bütün korelasyon katsayıları bakımından gerçekleşen güç değerleri genel olarak ya tam olarak %100 değerine ulaşmış ya da giderek %100'e yaklaşmıştır.

1.Tip hata olasılıkları ile testin güç değerleri birlikte değerlendirildiğinde genel olarak PM, WK ve PT korelasyon katsayılarının SR ve KT korelasyon katsayılarına tercih edilebileceği rahatlıkla ileri sürülebilir. PM, WK ve PT korelasyon katsayılarının, SR ve KT korelasyon katsayılarına olan üstünlükleri özellikle $n \leq 10$ olduğu koşullarda çok daha belirgindir.

Örneklerin çekildikleri populasyonların dağılımlarının iki değişkenli 10 SD'li t-dağılımı olması durumunda ampirik olarak gerçekleşen 1.Tip hata olasılıkları ve testin güç değerleri Çizelge 4. 2. 2'de verilmiştir. Çizelge 4. 2. 2'den de görüleceği üzere bu koşullarda elde edilen 1.Tip hata olasılıkları, dağılımların iki değişkenli normal olması durumunda (Çizelge 4. 2. 1) gerçekleşen 1.Tip hata olasılıklarına oldukça benzerdir. Bu beklenen bir durumdur. Çünkü 10 serbestlik dereceli iki değişkenli t-dağılımı, iki değişkenli normal dağılıma oldukça benzer bir dağılımdır.

Örneklerin alınmış oldukları populasyonların dağılımlarının iki değişkenli log normal olması durumunda dikkate alınan korelasyon katsayıları bakımından gerçekleşen 1.Tip hata olasılıkları ve testin güç değerleri Çizelge 4. 2. 3'te topluca verilmiştir.

Çizelge 4. 2. 3' ten de görüleceği üzere özellikle Pearson-Moment korelasyon katsayısı dağılımın normal olmamasından ya da normallikten sapmaların belirgin düzeylerde olmasından oldukça olumsuz yönde etkilenmiştir.

Bu tez çalışmasında Pearson-Moment korelasyon katsayısı bakımından gerçekleşen 1.Tip hata olasılıklarının çalışılan örnek hacminin 100 olması durumunda bile %6.34 olarak tahmin edilmiştir. Bu durumun testin gücüne olumsuz yönde yansıması ise özellikle $n<10$ olduğu durumlarda belirgindir. Spearman-Rank ve Kendall-Tau korelasyon katsayıları ise beklenildiği üzere dağılım şekillerinden pek fazla etkilenmemiştir. Bu iki korelasyon katsayısı özellikle örnek genişliğinden etkilenmektedir. Ancak, örnek genişliğinin artması; söz konusu etkinin göz ardı edilebilecek düzeylere düşürmüştür.

Bu tez çalışmasında genel olarak Pearson-Moment, Winsorize ve Permütasyon-tabanlı korelasyon katsayılarının Spearman-Rank ve Kendall-Tau korelasyon katsayılarına göre daha güçlü oldukları sonucuna varmak mümkündür.

PM, WK ve PT korelasyon katsayılarının, SR ve KT korelasyon katsayılarına olan üstünlükleri özellikle $n\leq10$ olduğu koşullarda çok daha belirgindir. Dolayısıyla 1.Tip hata ve testin güç değerleri birlikte değerlendirildiğinde, bu deneme koşullarında Winsorize ve Permütasyon-tabanlı korelasyon katsayılarının kullanılmasının diğer üç korelasyona göre daha avantajlı oldukları rahatlıkla söylenebilir. Diğer taraftan 1.Tip hata olasılıkları ile testin güç değerleri birlikte değerlendirildiğinde genel olarak PM, WK ve PT korelasyon katsayılarının SR ve KT korelasyon katsayılarına tercih edilebileceği rahatlıkla ileri sürülebilir. Bu deneme koşullarında özellikle küçük hacimli örneklerle çalışılması durumunda ($n\leq10$) Permütasyon-tabanlı korelasyon katsayısı ile Winsorize korelasyon katsayısının hem 1.Tip hata olasılığını %5 seviyesinde koruması bakımından hem de gerçekleşen güç değerleri bakımından diğer korelasyon katsayılarından daha üstün oldukları görülmüştür. Dolayısıyla küçük hacimli örneklerle çalışılması durumlarında Permütasyon-tabanlı korelasyon katsayısı ile Winsorize korelasyon katsayısının diğer korelasyon katsayılarına göre tercih edilebileceği rahatlıkla söylenebilir. Diğer taraftan beklenildiği gibi örnek genişliğinin artmasına paralel olarak bütün korelasyon katsayılarının giderek birbirlerine benzer sonuçlar verme eğilimine girdikleri görülür. Dolayısıyla $n\geq20$ olduğu durumlarda herhangi bir korelasyon katsayısından yararlanılabileceği rahatlıkla söylenebilir.

Söz konusu korelasyon katsayıları arasındaki farklar özellikle örnek genişliğinin küçük ($n \leq 10$) olduğu durumlarda belirgindir. Dolayısıyla bu korelasyon katsayılarının güç değerleri arasındaki farklar çalışılan örnek genişliğinin küçük olması durumlarında daha net görülür.

Genel olarak Winsorize, Pearson ve Permütasyon-tabanlı korelasyon katsayıları bakımından pozitif olarak hesaplanan korelasyon katsayıları, Spearman-Rank ve Kendall-Tau korelasyon katsayılarına göre daha fazla olduğu için bu üç korelasyon katsayısının Spearman-Rank ve Kendall-Tau korelasyon katsayılarından daha güçlü oldukları sonucuna varılır.

Bu beş korelasyon katsayısı arasında gücü en düşük olan korelasyon katsayısı ise Kendall-Tau'dur.

KAYNAKLAR

Abdullah M.B., 1990. On a Robust Correlation Coefficient. *Journal of the Royal Statistical Society. Series D (TheStatistician).*39(4): 455-460.

Akkartal E., Mendes M., Mendes E., 2010. Determination of SuitablePermutation Numbers in Comparing Independent Group Means: A Monte Carlo Simulation Study. *Journal of Scientific & Industria lResearch.* 69: 422-425.

Albayrak R., Özgümüş E., Özkan M. M. ve Kocabaş Z., 2011. Normallikten Sapma ve Farklı Örnek Genişliklerinde Parametrik ve Parametrik Olmayan İlişki Ölçütlerinin 1. Tip Hatası. *7. Ulusal Zootekni Bilim Kongresi,* Adana. 139.

Arndt S., Magnotta V., 2001. Generating Random Series with Known Values of Kendall's Tau. *Computer Methods and Programs in Biomedicine* 65: 17–23.

Balakrishnan N., Lai C.D., 2009. *Continuous Bivariate Distributions*, 2nd Ed.Springer, New York.

Bishara A.J., Hittner J.B., 2012. Testing the Significance of a Correlation with Non-normal Data: Comparison of Pearson, Spearman, Transformation, and Resampling Approaches. *Psychological Methods* May 7 (PMID:22563845).

Blair R., Lawson S., 1982. Another Look at the Robustness of the Product-moment Correlation Coefficientto Population Non-normality. *Florida Journal of Educational Research.*24: 11-15.

Bonett D. G., Wright T. A., 2000. Sample Size Requirements for Estimating Pearson, Kendall and Spearman Correlations. *Psychometr.* 65(1): 23–28.

Calkins D. S., 1974. Some Effects of Non-normal Distribution Shape on the Magnitude of the Pearson Product Moment Correlation Coefficient. *Interamerican Journal of Psychology* 8: 261-288.

Chok N.S., 2010. Pearson's vs Spearman's and Kendall's Correlation Coefficientsfor Continuous Data (Yüksek Lisans Tezi). University of Pittsburgh.

Dunlap W., Burke M., Greer T., 1995. The Effect of Skew on the Magnitude of Product-Moment Correlations. *Journal of General Psychology* 122(4): 365-377.

Edgell S., Noon S., 1984. Effect of Violation of Normality on the TTest of the Correlation Coefficient. *Psychological Bulletin* 95(3): 576-583.

Ernst M.D., 2004. Permutation Methods: A Basis for Exact Inference. *Statistical Science* 19: 676-685.

Fowler R., 1987. Power and Robustness in Product-Moment Correlation. *Applied Psychological Measurement* 11(4): 419-428.

Good P., 2005. *Permutation, Parametric and Bootstrap Tests of Hypotheses, (3rd Ed.*)

Springer-Verlag, New York.

Good P., 2009. Robustness of Pearson Correlation.*Interstat. 15*(5): 1-6.

Hayes A. F., 1996. PERMUSTAT: Randomization Tests for the MacIntosh. *Behavior Research Methods, Instruments, &Computers.*28: 473-475.

Hayes A. F., 1998. SPSS Procedures for Approximate Randomization Tests. *Behavior Research Methods, Instruments & Computers* 30: 536-543.

Hayes A. F., 2000. Randomization Tests and the Equality of Variance Assumption When Comparing Group Means. *Animal Behaviour* 59: 653-656.

Khamis H., 2008. Measures of Association: How to Choose? *Journal of Diagnostic Medical Sonography* 24: 155-162.

Knight W., 1966. A Computer Method for Calculating Kendall's Tau with Ungrouped Data. *Journal of the American Statistical Association* 61 (314): 436–439.

Keller-McNulty S., Higgins J. J., 1987. Effect of Tail Weight and Outliers on Power and Type-I Error of Robust Permutation Tests for Location. *Communications in Statistics B-Simulation and Computation* 16 (1): 17-35.

Kotz S., Balakrishnan N., Johnson N.L., 2000. *Continuous Multivariate Distributions: Volume 1: Modelsand Applications, 2nd Ed*. John Wiley & Sons, New York.

Kowalski C. J., Tarter M. E., 1969. Co-ordinate Transformations to Normality and the Power of Normal Tests for Independence *Biometrika* 56: 139-148.

Lieberson S., 1964. Limitations in the Application of Non-Parametric Coefficients of Correlation. *American Sociological Review* 29: 744-746.

Mielke P. W., Berry K. J., 2007. *Permutation Methods. A Distance Function Approach, (2nd) Ed.*Springer, New York .

Osborne J., 2002. Notes on the use of data transformations. *Practical Assessment, Research & Evaluation, 8(6).* Retrieved June 20, 2015 from http://PAREonline.net/getvn.asp?v=8&n=6

Rasmussen J., 1989. Data Transformation, Type I Error Rate and Power. *British Journal of Mathematical and Statistical Psychology* 42(2): 203-213.

Rasmussen J., Dunlap W., 1991. Dealing with Non-normal Data: Parametric Analysis of Transformed Data vs. Nonparametric Analysis. *Educational and Psychological Measurement* 51(4): 809-820.

Ruscio J., Kaczetow W., 2008. Simulating Multivariate Non-normal Data Using an Iterative Algorithm. *Multivariate Behavioral Research* 43: 335-381.

Sheskin D.J., 2007. *Handbook of Parametric and Nonparametric Statistical Procedures, (4th) Ed*. Chapman & Hall/CRC, Boca Raton, FL.

Siegel S., 1957.Nonparametric Statistics. *The American Statistician*. 11: 13-19.

Tabachnick B., Fidell L., 2007. *Using Multivariate Statistics, 5th Ed.* Allyn & Bacon/Pearson Education, Boston, MA.

Tanizaki H., 2004. On Small Sample Properties of Permutation Tests: Independence Between Two Samples. *International Journal of Pure and Applied Mathematics*. 13 (2): 235-243.

Triola M., 2010.*Elementary Statistics, 11th Ed.* Addison-Wesley/Pearson Education, Boston, MA.

Warner R., 2008. *Applied statistics: From Bivariate through Multivariate Techniques*. Sage Publications, Thousand Oaks: CA.

Wilcox R.R., 1993. Some Results on a Winsorized Correlation Coefficient. *British Journal of Mathematical and Statistical Psychology* 46: 339-349.

Wilcox R. R., 2001.*Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy* Springer, New York.

Wilcox R.R., 2004. Inferences Based on a Skipped Correlation Coefficient. *Journal of Applied Statistics* 31: 131–143.

Witte R., Witte J., 2010. *Statistics, (9th) ed*. Wiley, New York.

Yiğit S., Akkartal E., Mollaoğulları A. ve Mendeş M., 2012. Bazı Korelasyon Katsayılarının 1. Tip Hata ve Testin Gücü Bakımından Karşılaştırılması, *13. Uluslararası Ekonometri ve İstatistik Sempozyumu* Gazimağusa, KKTC.

Zimmerman D., Zumbo B., 1993. Significance Testing of Correlation Using Scores, Ranks, and Modified Ranks. *Educational and Psychological Measurement*53(4): 897-904.

Zimmerman D.W., Williams R.H., 1997. Properties of the Spearman Correction for Attenuation for Normal and Realistic Non-normal Distributions. *Applied Psychological Measurement* 21 (3): 253-270.

Zimmerman D. W., Zumbo B. D., Williams R. H., 2003. Bias in Estimation and Hypothesis Testing of Correlation. *Psicologica* 24: 1335-158.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Elif TUĞRAN

Doğum Yeri : BOLU

Doğum Tarihi : 23.10.1981

EĞİTİM DURUMU

Ön Lisans Öğrenimi: Abant İzzet Baysal Üniversitesi Mudurnu Meslek Yüksekokulu Tavukçuluk Bölümü (2001 - 2002)

Lisans Öğrenimi : Çanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi, Çanakkale (2002 – 2005)

Yüksek Lisans Öğrenimi : Çanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi Zootekni Ana Bilim Dalı, Çanakkale

Bildiği Yabancı Diller : İngilizce

BİLİMSEL FAALİYETLERİ

a) Yayınlar -SCI –Diğer:

1.Gebeliğin Son Döneminde Farklı Rasyonlarla Beslenen Süt Keçilerinde Bazı Davranış ve Doğum Özellikleri Arası İlişkiler. 5. Ulusal Zootekni Bilim Kongresi. 5-8 Eylül 2007, Van.

2. Profile Analysis and Growth Curve for Body Mass Index of Broiler Chickens Reared Under Different FeedRestrictions in Early Age"Archiv für Tierzucht-Archives of Anımal Breeding, 50 (4) 403-411, (2007)

b) Bildiriler -Uluslararası -Ulusal . Farklı Keçi Genotiplerinde Plasenta Özellikleri. 6. Ulusal Zootekni Kongresi, 24-26 Haziran 2009, Erzurum

İŞ DENEYİMİ

Çalıştığı Kurumlar ve Yıl :

(2001 -2002) :Abant İzzet Baysal Üniversitesi Mudurnu Meslek Yüksekokulu Tavukçuluk Bölümü Staj , Beypi AŞ. BOLU

(2004 – 2005) : Çanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi, Zootekni Bölüm Stajı: Beypi AŞ. BOLU

Ödül ve Dereceler

Haziran 2002 : Abant İzzet Baysal Üniversitesi Mudurnu Meslek Yüksekokulu Tavukçuluk Bölümü Üçüncüsü

İLETİŞİM

E-posta Adresi : elifarslanelif@yahoo.com