

Bilimsel Çalışmalarda P-Değerinin Rapor Edilmesi (P<0.01?, P<0.05?, P>0.05?)

Mehmet MENDEŞ¹

Serkan SUBAŞI²

Ensar BAŞPINAR³

Geliş Tarihi: 14.06.5005

Öz: Bilindiği üzere bilimsel çalışmaların sonuçları genellikle P-değerinden yararlanılarak rapor edilmektedir. Ancak, çalışmaların istatistik analizleri sonucunda bulunan P-değerlerinin, P<0.01, P<0.05 ya da P>0.05 gibi gelenekselleşmiş bir şekilde rapor edilmesi, okuyucuların bazı yanlışlara düşmesine neden olabilmektedir. Çünkü, aynı amaçla yapılmış çalışmalara ilişkin P-değerlerinin 0.05'ten büyük olması halinde, her ne kadar verilen kararlar aynı ise de, bu çalışmalardan elde edilen sonuçların güvenilirlikleri aynı düzeyde olmamaktadır. Bu çalışmada, 15x100x200 cm boyutlarında hazırlanan beton bloğun 3 farklı bölgesinden alınan yüzey sertliği değerleri arasında istatistik olarak önemli bir farkın olup olmadığına ilişkin hipotez kontrolleri üzerinde durulmuştur. Çalışmada, aynı amaçla yapılan denemelerde kullanılan tekerrür sayılarının farklı olması durumunda P-değeri ve dolayısıyla da Ho hipotezine ilişkin varılacak karar ve söz konusu kararın doğruluk derecelerindeki değişime dikkat çekilmiştir. Sonuç olarak; istatistik analizler sonucunda bulunan P-değerlerinin, kesin olasılık değerleri olarak (P=x.xxx gibi) rapor edilmesi, çalışmadan yararlanacak üçüncü şahıslara sağlayabileceği yararlar göz önüne alınırsa daha doğru ve yararlı olacağı sonucuna varılmıştır.

Anahtar Kelimeler: P-değeri, 1.tip hata, önemlilik seviyesi, hipotez kontrolleri, istatistik

Reporting of the P-Value in Scientific Research (P<0.01?, P<0.05?, P>0.05?)

Abstract: It is known that results of scientific research are reported using P-value in general. However, the conventional use of P values P <0.01, P < 0.05 or P > 0.05 after statistical analysis may misleads the readers. When two studies reach the same conclusion P>0.05, but with different P values, reliability of these two studies cannot be the same though the decisions are identical. In this study, the hypothesis stating there is no difference among the surface hardness values from 3 different points of a concrete block with the dimensions 15x100x200 cm is checked. The study calls attention to the P-value and therefrom reached decision on Ho hypothesis, and the changes in the correctness levels of the mentioned decision, in the case of having different number of replications in the experiments carried out with the same objectives. As a result, it would be proper and more useful for the third parties if the P-values found with the statistical analysis were presented as exact probability values (P=x.xxx).

Key Words: P-value, type I error, significance level, hypothesis testing, statistics

Giriş

Bilimsel araştırmaların sonuçları genellikle P-değeri kullanılarak rapor edilir (Little 1981). P-değeri ya da önemlilik seviyesi, doğru olan kontrol hipotezinin (H_0) reddedilmesi durumunda içine düşülecek yanlış olma olasılığının bir göstergesidir (Schervish 1996; Huang ve ark. 1997). Mesela belirli bir özellik bakımından A ve B gibi iki muamele grubunun ortalamasını karşılaştırmak amacıyla bir deneme kuran araştırmacı, denemesini kurmakla birlikte $H_0 : \mu_A = \mu_B$ şeklinde ifade edilen kontrol hipotezini de kurmuş olur. Bu şekilde kurulan kontrol hipotezi ile "grup ortalamaları arasında fark yoktur" veya "grup ortalamaları arasındaki farklar istatistiksel olarak önemli olmayan ve tamamen tesadüften ileri gelen farklardır" fikri ileri sürülmüş olur.

Bu fikrin geçerli olup olmadığı ancak $H_0 : \mu_A = \mu_B$ hipotezinin, $H_1 : \mu_A \neq \mu_B$ şeklinde kurulan ve "grup

ortalamaları arasında fark vardır" veya "iki grup ortalaması arasındaki fark istatistiksel olarak önemlidir" şeklinde ifade edilen karşıt ya da alternatif hipoteze karşı test edilmesi ile mümkün olur (Zar 1999; Hassan ve Khurshid 2001).

İstatistik analizler sonucunda H_0 hipotezinin ret edilmesi durumunda, içine düşülecek yanlış olma olasılığının ne kadar olduğu hesaplanır. Tıp, ziraat, biyoloji, orman gibi birçok bilim dalında yapılan çalışmalar için genel olarak kabul görmüş yanlış olma olasılığı ise $\alpha=0.05$ (%5) ve $\alpha=0.01$ (%1)'dir (Cohen 1994).

Yapılan istatistik analizler sonucunda içine düşülecek yanlış olma olasılığının 0.05'ten daha küçük olması durumunda, H_0 hipotezi ile ileri sürülen görüşler ret edilir. Diğer taraftan, yanlış olma olasılığının 0.05'ten daha büyük olması durumunda ise söz konusu görüşler kabul edilir (Winer ve ark. 1991, Motulsky 1995, Schervish 1996).

¹ Çanakkale Onsekiz Mart Üniv. Ziraat Fak.–Çanakkale

² Gazi Üniv. Teknik Eğitim Fak. Yapı Eğitimi Bölümü-Ankara

³ Van Yüzüncü Yıl Üniv. Tıp Fak. Biostatistik ABD-Van

Günümüzde hipotez kontrolleri Minitab, Statistica, SPSS, SAS, Systat, NCSS gibi oldukça gelişmiş istatistik paket programlarından yararlanılarak yapılmakta ve yapılan istatistik analizler sonucunda bulunan P-değerinden yararlanılarak H_0 hipotezine ilişkin bir hükme varılmaktadır. Yani, H_0 hipotezinin doğru ya da geçerli olup olmadığı istatistik analiz sonuçlarında bulunan P-değerinden anlaşılmaktadır. Eğer $P > \alpha$ ise H_0 hipotezi ile ileri sürülen görüşlerin geçerli olduğu (H_0 hipotezi ile ileri sürülen görüşlerin doğruluğunun ret edilemeyeceği) anlaşılırken, $P < \alpha$ olması durumunda ise H_0 hipotezi ile ileri sürülen görüşlerin geçerli olmadığı (H_0 hipotezi ile ileri sürülen görüşlerin doğruluğunun ret edilebileceği) anlaşılır (Gibbons 1985, Frick 1996, Hassan ve Khurshid 2001). İstatistik analizler sonucunda bulunan P-değeri 0.01'den büyük ve 0.05'ten küçük ya da eşit ise ($0.01 < P \leq 0.05$), ortalamalar arasındaki farklılığın 0.05 (%5) düzeyinde önemli olduğu, P-değeri 0.01'e eşit ya da daha küçük ise ($P \leq 0.01$) söz konusu farkın 0.01 (%1) düzeyinde önemli olduğu anlaşılır (Little 1981, Huang ve ark. 1997). Diğer taraftan, P-değerinin 0.05'ten büyük bulunması halinde, H_0 hipotezinin geçerli olduğu yönünde bir hükme varılır. Diğer bir ifade ile P-değerinin 0.05'ten büyük olarak bulunması, H_0 hipotezinin reddedilemeyeceği anlamına gelir. Örneğin, yukarıda verilen çalışma için $P=0.06$ olarak bulunmuşsa, "iki muamele grubunun ortalaması arasında fark yoktur" şeklinde ileri sürülen görüşler kabul edilir. Ancak, P-değerinin 0.06 olarak bulunması, varılan hükmün tamamen doğru ya da geçerli olduğu anlamına gelmez. Yani, H_0 hipotezinin kabul edilmesiyle söz konusu muamele gruplarının ortalamaları arasında hiç bir farkın olmadığı anlamı çıkartılmamalıdır. Çünkü $P=0.06$ olarak bulunması, üzerinde durulan özellik bakımından bu muamele gruplarının ortalamaları arasında hiç bir farkın olmadığı değil, fark olduğunun kabul edilemeyeceğidir (Lehmann 1959, Gibbons 1985, Barnard 1990, Chia 1997, Goodman 1999). Özetle, "P-değerinin 0.05'ten büyük çıkması $H_0 : \mu_A = \mu_B$ şeklinde kurulan kontrol hipotezinin doğruluğunun ret edilemeyeceği anlamına gelmektedir (fail to reject)". Çünkü, bahsedilen çalışma için P-değeri 0.11 ya da 0.56 ya da 0.94 olarak bulunmuş olsaydı yine aynı hükme varılacaktı. Ancak, her durum için elde edilen sonuçların güvenilirlikleri, yani doğruluk dereceleri aynı düzeyde değildir. Örneğin, $P=0.06$ olarak bulunduğu bir denemede "muamele gruplarının ortalamaları arasında fark vardır" şeklinde bir hükme varılması, yani H_0 hipotezinin ret edilmesi durumunda içine düşülecek yanlış olma olasılığı 0.06'dır (halbuki 0.05'i geçmesi istenmez). Diğer bir ifade ile doğru karar verme olasılığımız %6 iken, yanlış karar verme olasılığımız %94'tür. P-değerinin 0.11 olarak bulunduğu bir denemede de aynı hüküm verilmekte ancak, bu hipotezin reddedilmesi durumunda içine düşülecek yanlış olma olasılığı 0.11 olmaktadır. P-değerinin 0.56 ve 0.94 bulunduğu denemeler için içine düşülecek yanlış olma

olasılıkları ile yanlış karar verme olasılıkları sırasıyla; 0.56, 0.44 ve 0.94, 0.06 olacaktır.

Dikkat edileceği üzere bu dört durum için de aynı yönde hükümlere varılmasına karşılık, varılan hükümlerin güvenilirlik düzeyleri aynı değildir. Bu gibi durumlarda alışılmış bir biçimde "muamele grupları arasında önemli bir fark bulunmamıştır" demek yerine, bu ifadeyi kullanıp yanına $P=0.06$, $P=0.11$, $P=0.56$ ve $P=0.94$ gibi olasılık değerlerinin verilmesi daha doğrudur. Diğer taraftan, "gerçekte gruplar arasında fark yokken", "fark vardır" şeklinde bir sonuca ulaşma olasılığı örnek hacmi ile de oldukça yakından ilişkilidir (Barnard 1990, Ramp ve Yancey 1991, Goodman 1999). Aynı amaçla yapılan fakat farklı örnek hacimleriyle yürütülen iki denemede, örnek hacimlerinin farklı olmasından dolayı farklı P-değerlerinin elde edilmesi ve dolayısıyla deneme sonuçlarının farklı olması beklenen bir durumdur. İstatistik analizler sonucunda bulunan P-değeri, ölçüm değerlerinin varyasyonu da oldukça yakından ilişkilidir. Aynı muameleye tabi tutulan deney ünitelerinin ölçüm değerleri arasındaki farklılıkların (varyasyonun) fazla olması, istatistik analizler sonucunda bulunan P-değerinin 0.05'ten büyük olmasına yol açmaktadır. Bunun sonucunda da gerçekte grup ortalamaları arasında istatistik olarak önemli bir farkın bulunmasına karşılık, gruplar arasında önemli bir fark olmadığı şeklinde oldukça yanıltıcı bir sonuca varılabilmektedir. Çünkü, aynı muameleye tabi tutulan deney üniteleri arasındaki büyük farklılıklar (varyansların homojenliği ön şartının sağlanamaması) etkisi araştırılmak istenen muamele grupları arasında gerçekte var olan farkların örtülmesine sebep olabilmektedir (Parkhurst 1985, Frick 1996, Mendeş 2002). Böylece, gerçekte gruplar arasında var olan farklar ortaya konulamamaktadır. Bu durum, uygulamada çok sık karşılaşılan problemlerden birisi olarak karşımıza çıkmaktadır. Bunun için, özellikle t-testi ve F-testinin (ANOVA F testi) uygulanacağı durumlarda istatistik analize başlanmadan önce, verilerde özellikle varyansların homojenliği ön şartının sağlanıp sağlanmadığının test edilmesinde büyük yarar vardır. Eğer bu ön şart sağlanıyorsa söz konusu verilerin değerlendirilmesinde bu testler kullanılabilir ve elde edilen analiz sonuçları da güvenilir sonuçlar olur. Diğer taraftan, varyansların homojenliği ön şartının sağlanmadığı durumlarda bu testler kullanılırsa önceden de belirtildiği üzere grup ortalamaları arasında var olan farkların ortaya konulamaması sonucuyla karşılaşılır ki bu da oldukça istenmeyen bir durumdur. Bu ön şartın sağlanmadığı durumlarda genel olarak iki çözüm yolu önerilmektedir. Bunlar:

a) Verilere uygun bir transformasyon uygulayarak bu ön şartın sağlanmasına çalışmak,

b) Söz konusu testlere alternatif olan Welch, Brown-Forsythe, Wilcoxon, James Second-Order, Alexander-Govern gibi testlere başvurmak (Bek ve Efe 1987, Mendeş 2002).

Bu çalışmanın amacı, bütün pozitif bilim dallarında yapılan çalışmaların istatistiksel olarak değerlendirilip elde edilen sonuçların yorumlanmasında P-değerinin nasıl rapor edileceğinin açıklanmasıdır.

Materyal ve Yöntem

Çalışmada 15x100x200 boyutlarında beton blok üretmeye olanak sağlayacak şekilde perde duvar kalıbı üretilmiştir. Perde duvar kalıbına; tamamı tek harman halinde hazırlanmış olan C20 betonu dökülmüş ve iki ayrı noktadan vibratör daldırmak suretiyle sıkıştırılmıştır. Betonun kürü sulanmak suretiyle yapılmıştır.

Test çekici ile yüzey sertliği tayini deneyi : Yüzey sertliği deneyinde; geri tepmeli çekiç (Schmidt Çekici) kullanılmıştır. Geri tepmeli çekiç ile yüzey sertliği deneyinde beton blok üç bölgeye ayrılmış ve her bölgede 50 ayrı noktaya deney uygulamakla beton blok için toplam 150 adet okuma yapılmıştır. Yüzey sertliği okuması, beton dökümünden 28 gün sonra "BS 1881; Part 202, Testing Concrete, Recommendation for Surface Hardness Testing by Rebound Hammer" standardında belirtilen metotlara uygun olarak yapılmıştır (Subaşı 2001).

İstatistik analiz: Hazırlanan beton bloğun 3 farklı bölgesinden alınan yüzey sertliği değerleri arasında istatistik olarak önemli bir farkın olup olmadığına ilişkin hipotez kontrolleri üzerinde durulmuştur. Çalışmada, aynı amaçla yapılan denemelerde kullanılan tekerrür sayılarının farklı olması durumunda P-değeri ve dolayısıyla da H_0 hipotezine ilişkin varılacak karar ve söz konusu kararın doğruluk derecelerindeki değişime dikkat çekilmiştir.

Bulgular ve Tartışma

Çalışmada, tekerrür sayıları farklı ancak aynı amaçla yürütülmüş dört denemenin istatistik analizleri (ANOVA) sonucunda bulunan P-değerleri Çizelge 1'de topluca verilmiştir.

Tekerrür sayıları farklı olan 1., 2. ve 3. çalışmalara ilişkin P-değerleri incelendiğinde, söz konusu üç P-değerinin de H_0 hipotezinin reddedilmesi için kabul görmüş olan üst sınır değerinden (0.05) daha büyük olduğu görülür (Çizelge 1). Dolayısıyla, her üç çalışma sonunda da "Üç bölge ortalaması arasında istatistiksel olarak önemli bir fark yoktur" veya "bölge ortalamaları arasındaki farklar, istatistiksel olarak önemli olmayan ve tesadüften ileri gelen farklılıklardır" şeklinde bir hükme varılmaktadır. Ancak, bu çalışmaların her biri için varılan sonuçların güvenilirlikleri aynı düzeyde değildir. Mesela, 5'er tekerrürle yürütülen 1. çalışmaya ilişkin P-değeri 0.729 olarak bulunurken, 10'ar tekerrürle yürütülen 2. çalışmaya ilişkin P-değeri 0.182 olarak bulunmuştur. 20'şer tekerrürle yürütülen 3. çalışmaya ilişkin P-değeri ise 0.066'dır. Dikkat edileceği üzere, çalışmalardaki tekerrür sayısının artması, P-değerlerinin 0.05 sınırına yaklaştırmıştır. Dolayısıyla, tekerrür sayısının artması gerçekte bölgeler arasında var olan önemli farkların daha duyarlı bir şekilde ortaya konulmasına imkan sağlamaktadır. Yukarıda bahsedilen ve aynı amaçla fakat farklı sayıda tekerrürle yürütülen bilimsel çalışmaların sonuçlarının rapor edilmesinde çok dikkatli davranılması gerekmektedir. Söz konusu bilimsel çalışmaların sonuçlarının: "Yapılan varyans analizi sonunda bölge

ortalamaları arasında istatistiksel olarak önemli bir fark bulunmamıştır (P>0.05)" ya da "yapılan varyans analizi sonucunda bölgeler arasındaki farkın istatistiksel olarak önemli olmadığı görülmüştür (P>0.05)" şeklinde rapor edilmesi, okuyucuların bazı yanlışlara düşmesine neden olabilir. Çünkü sonuçların bu şekilde rapor edilmesi, her üç bilimsel çalışma sonucunun da güvenilirliklerinin aynı düzeyde olduğu gibi yanlış bir algılamaya neden olabilir. Halbuki bu çalışmalara ilişkin P-değerleri incelendiğinde, her denemeye ilişkin varılan hükümlerin güvenilirliklerinin aynı düzeyde olmadığı görülür. 5'er tekerrürle yürütülen 1.denemeye ilişkin P-değeri incelendiğinde "bölge ortalamaları arasında fark yoktur" şeklinde kurulan H_0 hipotezinin ret edilmesi durumunda içine düşülecek yanlış olma olasılığı %72,9'dur. 10'ar tekerrürle yürütülen 2.deneme ve 20'şer tekerrürle yürütülen 3.çalışmaya ilişkin P-değerleri incelendiğinde ise H_0 hipotezinin reddedilmesi durumunda içine düşülecek yanlış olma olasılıklarının sırasıyla %18,6 ve %6,6 olduğu görülür. Bu şekilde bulunan P-değerleri daha önce de belirtildiği üzere çalışmada kullanılan tekerrür sayılarından oldukça etkilenmektedir. Nitekim burada da görüleceği üzere 5'şer tekerrürle yürütülen 1. çalışmaya ilişkin bulunan P-değeri ile 20'şer tekerrürle yürütülen 3.çalışmaya ilişkin P-değeri arasında büyük farklılıklar olduğu görülmektedir. Bu P-değerleri incelendiğinde, 3. denemenin birkaç fazla tekerrürle yürütülmesi durumunda P-değerinin sınır değeri olan 0.05'e daha çok yaklaşacağı ortadadır. Bu çalışma sonunda "bu üç bölge ortalaması arasındaki fark istatistiksel olarak önemlidir" şeklinde bir hükme ulaşma olasılığı daha yüksek olurdu. Halbuki aynı amaçla yürütülen bu üç bilimsel çalışma sonucunun da daha önce de bahsedildiği üzere "bu üç bölge ortalaması arasında istatistiksel olarak önemli olan bir farka rastlanılmamıştır" şeklinde ifade edilmesi durumunda, sanki her üç denemede varılan sonuçların güvenilirlik düzeylerinin aynı olduğu kanısına varılmasına neden olabilmirdi. Aynı, zamanda 1., 2., ve 3. çalışmalara ilişkin tekerrür sayıları dikkate alındığında, 1. çalışmaya ilişkin 0.729 olarak bulunan P-değerinin fazla güvenilir olamayacağı ortadadır. Çünkü tekerrür sayısının düşük olarak alındığı çalışmalarda elde edilecek deneysel hata, tekerrür sayısının çok tutulduğu çalışmalara göre belirgin bir şekilde daha büyük olacaktır. Deneme hatası ise tekerrür sayısı arttıkça azalmaktadır. Bunun sonucunda da gruplar arasında gerçekte var olan farkların ortaya konulabilme olasılığı da artmaktadır. Bu bağlamda düşünüldüğünde bu üç bilimsel çalışma sonuçlarının aşağıdaki şekilde rapor edilmesinin daha yararlı olabileceği söylenebilir:

a) Yapılan varyans analizi sonunda üç bölge ortalaması arasındaki farkların istatistik olarak önemli olmadığı görülmüştür (P=0.729).

Çizelge 1. Aynı amaçla farklı tekerrürle (n) yürütülen dört çalışmaya ilişkin P-değerleri

Çalışma No	n1:n2:n3	P	Hüküm
1	5:5:5	0.729	Ho: Kabul
2	10:10:10	0.182	Ho: Kabul
3	20:20:20	0.066	Ho: Kabul
4	50:50:50	0.000	Ho: Ret

b) Yapılan varyans analizi sonunda üç bölge ortalaması arasındaki farkların istatistik olarak önemli olmadığı görülmüştür ($P=0.182$).

c) Yapılan varyans analizi sonunda üç bölge ortalaması arasındaki farkların istatistik olarak önemli olmadığı görülmüştür ($P=0.066$).

Bu üç bilimsel çalışma sonucunun bu şekilde rapor edilmesi, bu çalışmaları kendi çalışmaları için referans olarak kullanacak araştırmacılara da büyük kolaylıklar sağlar. Şöyle ki, araştırmacı her ne kadar bu üç çalışma sonucunda varılan hükümlerin aynı yönde olduğunu görse de söz konusu çalışma sonuçlarının güvenilirlik dereceleri arasında büyük farklılıkların olduğunu anlayabilir. Aynı zamanda bu çalışma sonuçlarından hareketle tekrür sayısının, elde edilecek sonuçlara etkisi hakkında bir fikir sahibi olabilir ve kendi araştırmasında dikkate alacağı tekrür sayısına ilişkin bir ipucu elde edebilir.

Diğer yandan, 50'şer tekrürlü olarak yürütülen 4. denemeye ilişkin P-değeri ise 0.000 olarak bulunmuştur. Dolayısıyla 4.deneme sonucunda varılan hüküm, 1. 2. ve 3. deneme sonunda varılan hükümlerin tam tersidir. Bu sonuçta 4. denemede ki tekrür sayısının diğer üç denemede ki tekrür sayılarından belirgin bir miktarda daha fazla olmasının da payı oldukça yüksektir. Bu çalışma sonucunun da rapor edilmesinde yukarıda bahsedildiği üzere P'nin gerçek değerinin verilerek rapor edilmesinde büyük yarar vardır. Bu durumda "yapılan varyans analizi sonucunda üç bölge ortalaması arasındaki farkların istatistiksel olarak önemli olduğu görülmüştür ($P<0.05$)" şeklinde gelenekselleşmiş bir rapor biçimi yerine: "yapılan varyans analizi sonucunda üç bölge ortalaması arasındaki farkların istatistiksel olarak önemli olduğu görülmüştür ($P=0.000$)" şeklinde rapor edilmesi daha doğru olacaktır.

Sonuç ve Öneriler

P'nin gerçek değerinin rapor edilmesi araştırmacılara, örnek hacminin H_0 hipotezinin ret veya kabul edilmesini ne kadar desteklediği ile ilgili oldukça önemli bilgiler verir. Bu değerlendirmelerden hareketle, yapılan bir çalışmanın sonuçları rapor edilirken:

1) Eğer üzerinde durulan özellik bakımından gruplar arasında istatistik olarak önemli bir fark var ise "üzerinde durulan özellik bakımından gruplar arasındaki farkların istatistik olarak önemli olduğu görülmüştür ($P<0.01$)" ya da $P<0.05$) yerine, örneğin "üzerinde durulan özellik bakımından gruplar arasındaki farkların istatistiksel olarak önemli olduğu görülmüştür ($P=0.046$)" şeklinde rapor edilmesi,

2) Eğer üzerinde durulan özellik bakımından gruplar arasında istatistik olarak önemli bir farkın olmadığı görülmüşse bu durumda da "üzerinde durulan özellik

bakımından gruplar arasındaki farkların istatistiksel olarak önemli olmadığı görülmüştür" ifadesi yerine örneğin "üzerinde durulan özellik bakımından gruplar arasındaki farkların istatistiksel olarak önemli olmadığı görülmüştür ($P=0.182$)" ifadesinin kullanılması daha doğru ve açıklayıcı olacaktır.

Sonuç olarak, deneme sonuçlarının doğru bir şekilde yorumlanabilmesi, istatistik analizler sonucunda elde edilen P değerinin doğru bir şekilde rapor edilmesiyle oldukça yakından ilişkilidir. Dolayısıyla istatistik analizler sonucunda bulunan P-değerlerinin, kesin olasılık değerleri olarak ($P=x.xxx$ gibi) rapor edilmesi, çalışmadan yararlanacak üçüncü şahıslara sağlayabileceği yararlar göz önüne alınırsa daha doğru ve yararlı olacaktır.

Kaynaklar

- Barnard, G. A. 1990. Must clinical trails be large? The interpretation of P-values and the combination of test results. *Statistics in Medicine* 9, 601-614.
- Bek, Y. ve E. Efe. 1987. Araştırma ve Deneme Metotları I. Ç.Ü.Ziraat Fakültesi Yayın No:71, Adana.
- Chia, K. S. 1997. Significant-it is-An obsession with the P-value. *Scandinavian Journal of Public Health* 23, 152-154.
- Cohen, J. 1994. The earth is round ($P<0.05$). *American Psychologist* 49:997-1003.
- Frick, R. W. 1996. The appropriate use of null hypothesis testing. *Psychological Methods* 1, 379-390.
- Gibbons, J. D. 1985. P-values. in *Encyclopedia of Statistical Sciences* eds. Kotz, S., Johnson, N. L., 6, 366-368.
- Goodman, S. N. 1999. The P-value fallacy in medical statistics. *Annals of International Medicine* 130, 995-1004.
- Hassan, M., A. M. Khurshid. 2001. Role of P-values in decision making. *Proc. Eight Stat. Semp.*, 7 (8), 123-130.
- Huang, H. M. J., R. T. O'Neil, P. Bauer and K. Kohne. 1997. The behavior of the P-value when alternative hypothesis is true? *Biometrics* 53, 11-22.
- Lehmann, E. L. 1959. *Testing Statistical Hypotheses*. New York: Wiley USA.
- Little, T. M. 1981. Interpretation and presentation of results. *Hortscience* 16:637-640.
- Mendeş, M. (2002). Normal dağılım ve varyansların homojenliği ön şartlarının gerçekleşmediği durumlarda varyans analizi tekniği yerine kullanılabilecek bazı parametrik alternatif testlerin I.tip hata ve testin gücü bakımından irdelenmesi, Ankara Üniv., Fen Bil. Enstitüsü, Doktora Tezi (Basılmamış).
- Motulsky, H. J. 1995. *Interpreting nonsignificant P-values*. Instuitive Biostatistics, Oxford University Inc. Oxford.

Parkhurst, D. F. 1985. Interpreting failure to reject a null hypothesis. Bulletin of the Ecological Society of America 66:301-302.

Ramp, W. K., J. M. Yancey. 1991. P-values and their problems. Bone and Mineral 13, 163-165.

Schervish, M. J. 1996. P values: What they are and what they are not. The American Statistician 50,3,203-206.

Subaşı, S. 2001. Kalıp Yüzey Farklılıklarının Betonun Bazı Fiziksel Özellikleri Üzerine Etkileri, Gazi Üniv., Fen Bilimleri Enst., Y.Lisans Tezi (Basılmamış).

Winer, B.J., D. R. Brown and K. M. Michels. 1991. Statistical principles in experimental design. McGraw-Hill Book Company, New York: USA.

Zar, J. H. 1999. Biostatistical analysis. Prentice-Hall Inc. Simon and Schuster/A Viacom Company, New Jersey: USA.

İletişim adresi:

Mehmet MENDEŞ

Çanakkale Onsekiz Mart Üniversitesi

Ziraat Fakültesi Zootekni Bölümü - ÇANAKKALE

Tel: 0-286-218 00 18 / 1348

GSM : 0-542-476 00 27

E-mail: mmendes@comu.edu.tr

