

PAPER DETAILS

TITLE: Tekrarlanabilirlik Krizi ve Geçerlilik Krizi Kiskacındaki Psikoloji ve Sosyal Bilimlerde Krizden Çıkis İçin Öne Çıkan İki Trend: "Yeni İstatistik" ve "Bayesyen İstatistik"

AUTHORS: Ergün KARA

PAGES: 599-624

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/2910257>



DOI: 10.18039/ajesi.1240655

Two Prominent Trends In Psychology and Social Sciences In The Grip Of The Reproducibility Crisis and Validity Crisis : “New Statistics” and “Bayesian Statistics”¹

Ergün KARA²

Date Submitted: 22.01.2023

Date Accepted: 29.05.2023

Type: Review

Abstract

The reproducibility and P-value crises that started in social sciences, especially psychology, have left their mark in the last 10 years. In addition to all these, discussions of theory and validity crisis, which started from psychology but spread to social sciences, started. All these debates have imposed compulsory reform in the field of methods, applied statistical analysis, sampling and research practices in all fields of science in general and in social sciences. Within the scope of this study, the causes of these crises and the ways to get out of the crisis were discussed and suggestions were made. In this context, two historical transitions began to occur in data analysis in this short period of time. First, the use of effect size, confidence interval and meta-analysis in uncertainty estimation in classical hypothesis testing, this trend was named as “the New Statistics”. The second is the transition from classical statistics to Bayesian statistics. As a result, preferring open science, being transparent in research, utilizing the opportunities offered by new methods and statistics will contribute to the exit from this crisis and the production of healthier scientific knowledge. In social sciences, where it is difficult to create large samples, categorical and missing data are common, and difficult to meet assumption of multivariate normality. Bayesian approach, especially with the effect of obtaining posterior distributions by using a priori information and giving more accurate results in small samples, it will bring great innovation and benefit to the statistical dimension of the crisis.

Keywords: Bayesian statistics, reproducibility crisis, the validation crisis, new statistics.

Cite: Kara, E., (2023). Two prominent trends in psychology and social sciences in the grip of the reproducibility crisis and validity crisis : “New Statistics” and “Bayesian Statistics”. *Anadolu Journal of Educational Sciences International*, 13(2), 599-624. <https://doi.org/10.18039/ajesi.1240655>



¹ Eğitim Yönetimi Forumu (EYFOR-14)'nda sözlü bildiri olarak sunulmuştur.

² (Corresponding author) Dr., University of Alanya Alaaddin Keykubat, Faculty of Education, Department of Psychological Counseling and Guidance, Antalya, ergunpdr@hotmail.com <https://orcid.org/0000-0002-6129-6787>



Tekrarlanabilirlik Krizi ve Geçerlilik Krizi Kısacasındaki Psikoloji ve Sosyal Bilimlerde Krizden Çıkış İçin Öne Çıkan İki Trend: “Yeni İstatistik” ve “Bayesyen İstatistik”¹

Ergün KARA²

Gönderim Tarihi: 22.01.2023

Kabul Tarihi: 29.05.2023

Türü: Derleme

Öz

Başta psikolojinin ana aktörlerinden olduğu ve sosyal bilimlere sirayet eden tekrarlanabilirlik krizi ve p-değerinin sorgulanması son 10 yılda en çok tartışılan konulardan olmuştur. Son yapılan çalışmalar tekrar edilebilirlik krizinin sadece sosyal bilimler ile sınırlı kalmadığı temel bilimlerde etkili olduğunu göstermiştir. Bütün bunlara ek olarak yine psikoloji temelli başlayan teori ve geçerlilik krizi tartışmaları başlamıştır. Bütün bu tartışmalar genel olarak tüm bilim alanlarında özel olarak psikoloji ve sosyal bilimlerde yöntem, uygulanan istatistiksel analizler, örnekleme ve araştırma pratiklerinin değişimine yol açmıştır. Bu çalışma kapsamında kısaca tekrarlanabilirlik ve geçerlik krizlerine değinilecek ve krizden çıkış yolları kapsamında kullanılan istatistiksel yöntemler bağlamında önerilerde bulunulmuştur. Krizlerin sebeplerinden biri olarak gösterilen p-değerinin kötüye ve yanlış kullanımı öne çıkmaktadır. Bunun önüne geçebilmek için iki önemli yaklaşım önerilmektedir. Bunlardan ilki “yeni istatistik” (the New Statistics) olarak da adlandırılan ve klasik hipotez testinde etki büyüklüğü, güven aralığı ve meta-analizin belirsizlik tahminlemede kullanılmasına dayanan trenddir. İkincisi ise klasik istatistik yerine Bayesyen istatistiği öneren trenddir. Klasik istatistikten Bayesyen yöntemlere geçmek genel olarak büyük örneklem oluşturma zor olduğu, kategorik ve kayıp verinin yaygın olduğu ve çok değişkenli normalliği sağlamanın oldukça zor olduğu sosyal bilimlerde yararlı olacaktır. Bayesyen yaklaşım özellikle önsel bilgiyi kullanarak sonsal dağılımlar elde etmesi ve küçük örneklemde de daha doğru sonuçlar vermesinin etkisi ile bu krizin istatistik boyutuna büyük bir yenilik ve fayda getirecektir.

Anahtar kelimeler: Bayesyen istatistik, geçerlilik krizi, tekrarlanabilirlik krizi, geçerlilik krizi.

Atıf: Kara, E., (2023). Tekrarlanabilirlik krizi ve geçerlilik krizi kısacasındaki psikoloji ve sosyal bilimlerde krizden çıkış için öne çıkan iki trend: “Yeni İstatistik” ve “Bayesyen İstatistik”. Anadolu Journal of Educational Sciences International, 13(2), 599-624. <https://doi.org/10.18039/ajesi.1240655>

¹ Eğitim Yönetimi Forumu (EYFOR-14)'nda sözlü bildiri olarak sunulmuştur.

² (Sorumlu Yazar) Dr., Alanya Alaaddin Keykubat Üniversitesi, Eğitim Fakültesi, Rehberlik ve Psikolojik Danışmanlık Bölümü, , Antalya, ergunpdr@hotmail.com <https://orcid.org/0000-0002-6129-6787>

Giriş

Başta psikoloji olmak üzere sosyal bilimlerde 2010'da başlayan tekrarlanabilirlik krizi (The replication crisis) ve bunun sorumlularından biri olarak görülen P-değeri krizi son yıllarda en çok tartışılan konulardan biri olmaktadır (Baker, 2015). Tekrarlanabilirlik krizi, psikoloji alanında daha önce yapılan araştırmaların yeni, gelişmiş yöntemler kullanılarak ve daha büyük örneklemelerde tekrarlandığında aynı sonuçları vermemesi olarak tanımlanmaktadır (Errington ve diğerleri, 2014; Nosek ve Errington, 2017). Diğer yandan bu araştırmaların gelişmiş batılı ülkelerde (WEIRD) yapılmış olması dünyanın geri kalanını oluşturan yüzde 88 için tekrar edip edemediği konusu ise ayrı bir tartışma konusu olmaktadır. Kriz her ne kadar psikoloji temelli başlasa da geçerlilik krizinin de (The Validation Crisis) buna eklenmesi ile bütün bilim alanlarında güncel bir başlık olmaktadır (Schimmack, 2021; Muthukrishna ve Henrich 2019).

İlk olarak 2015 yılında The Reproducibility Project: Psychology adlı çalışma ile psikolojideki tekrarlanabilirlik krizi ile ilgili ilk bilimsel araştırma yayınlanmıştır. Psikolog Brian Nosek tarafından yürütülen çalışmada, psikolojideki üç önemli dergisinde (Journal of Personality and Social Psychology, Journal of Experimental Psychology: Learning, Memory, and Cognition and Psychological Science) yayımlanmış olan 100 araştırma tekrar edilmiştir. Araştırmaların sadece %36'sında anlamlı bulgulara ulaşılmıştır (p değeri 0.05'in altında). Orijinal çalışmalarda raporlanan etki büyüklüğünün ise tekrarında yaklaşık yarısı kadar olduğu görülmüştür (Open Science Collaboration 2015). 2018'de Nature Human Behavior'da yayınlanan araştırmada ise 21 sosyal ve davranışsal bilim makalesinin yaklaşık %62'sinin orijinal sonuçları başarılı bir şekilde yeniden üretebilmiştir (Rogers, 2018). Benzer şekilde, Açık Bilim Merkezi (Center for Open Science) tarafından yürütülen bir çalışmada, 60 farklı laboratuvarı (6 farklı kıtadan 36 farklı milleti temsil eden) 186 araştırmacıdan oluşan bir ekip, psikolojideki 28 klasik ve çağdaş bulgunun tekrarlarını gerçekleştirmiştir. Genel olarak, 28 araştırmacının %50'sinde, büyük örneklem gruplarına rağmen sonuçlar tekrarlanamamıştır. Bu sonuç düşünüldüğünde “örneklemede gelişmiş yöntemler kullanılmış olsa idi tekrarlanabilirlik krizi yaşanmazdı” görünüşünün de geçersiz olduğu görülmüştür (Witkowski, 2019). Son yapılan çalışmalar tekrar edilebilirlik krizinin sosyal bilimlerde olduğu gibi temel bilimlerde önemli boyutta olduğunu göstermiştir. Örneğin daha önce yapılan kanser araştırmalarının sadece yüzde 50'si tekrar etmiştir (Errington ve diğerleri 2014; Nosek ve Errington, 2017). Her ne kadar psikoloji ve kanser biyolojisi alanlarında yapılan çalışmalar öne çıksa da birçok bilim alanında aynı sorunun olduğu ortaya konulmuştur. Örneğin araştırmacılar, en çok güvenilir alanlar olarak görülen fizik ve kimyada bile tekrarlanabilirliğin ancak yüze 60-70 dolaylarında olacağını belirtmişlerdir (Baker, 2016).

Diğer yandan tekrar etmeyen çalışmalara alan yazında tekrar edenler kadar atıf almaya devam etmektedir. Bu dinamik alan yazını zayıflatmakta, araştırma giderlerini artırmakta ve

aslında tekrarlanabilirlik için farklı bir yaklaşımın geliştirilmesi gerektiğini ortaya koymaktadır (Yang, Youyou, ve Uzzi, 2020). Bu anlamda hem kavramsal hem de doğrudan tekrar edilebilir olmanın bilimsel ilerleme açısından hayati öneme sahip olduğu görülmektedir.

Geçerlilik krizi de güncel konulardan biri olmaya devam etmektedir (Schimmack, 2021; Muthukrishna ve Henrich 2019). Günümüzde geçerlilik ile ilgili birçok temel ilke Cronbach ve Meehl (1955) tarafından yazılan ve bir klasik olan Yapı geçerliliği (Construct Validity) makalesine dayanmaktadır. Cronbach ve Meehl APA tarafından desteklenen çalışmalarında yapı geçerliliğini kendi bakış açıları doğrultusunda tanımlamışlardır. Yapı geçerliliği bir testte bazı nitelik ve özelliklerin (operasyonel olarak tanımlanmamış) ölçülmesi gerektiğinde dahil edilip yorumlanmıştır (Cronbach ve Meehl, 1955). Günümüzde ise yapı geçerliliği, gözlemlenen test puanlarındaki değişim ile teorik bir yapıdaki buna karşılık gelen varyasyonu yansıtan gizil bir değişken arasındaki ilişki olarak tanımlanmaktadır (Schimmack, 2010). Tıpkı tekrarlanabilirlik krizinde olduğu gibi şimdi de geçerlilik krizinin daha başlarında bulunduğumuz belirtilmektedir (Schimmack, 2021). Psikolojide sıklıkla kullanılan birçok ölçme aracı büyük olasılıkla yeterli yapı geçerliliği sağlamamaktadır. Her ne kadar terimin tanımında günümüzde artık değişim yaşanmış olsa da araştırma pratiklerinde eski yöntemler kullanılmaya devam etmektedir. Yine yapı geçerliliğinin kesin kriterlerini oluşturmada eksikler olduğu görülmektedir (Schimmack, 2021). Araştırmacılar 2010'ların tekrarlanabilirlik krizi ile geçtiğini 2020'lerin ise muhtemelen geçerlilik krizi ile geçeceğini iddia etmektedirler (Schimmack, 2021). Araştırmacılar geçerlik ve güvenirlik çalışmalarının hız kazanıp alanda yaygınlaşması ve bir nihai sonuca ulaşması durumunda 2030'larda ilk defa geçerli olan ölçme araçları ile tekrarlanabilir çalışmalar yapılabileceğini öngörmektedir. Cronbach ve Meehl'in (1955) yapı geçerliliği ile ilgili yaptıkları çalışmaların üzerinde yıllar geçmiş olmasına rağmen geçerlilik çalışmalarında yeterince ilerleme sağlanmamış olması bu olasılığın gerçekleşmesinin de zor olduğunu göstermektedir (Schimmack, 2021).

Ortaya çıkan tekrarlanabilirlik ve geçerlilik krizlerini birçok açıdan ele almak mümkündür; bilimin teşvik-ödül yapısının pozitif sonuçların raporlanmasını teşvik etmesi ve istatistik metodolojisindeki sorunlu yaklaşımlar öne çıkan iki başlık olmuştur. Genel olarak sorunlu araştırma pratikleri (John, Loewenstein ve Prelec, 2012; Simmons, Nelson ve Simonsohn, 2016), istatistiksel güç yönünden yetersiz (underpowered) çalışmalar (Bertamini ve Munafò, 2012), yayın yanlılığı (Bakker, van Dijk, ve Wicherts, 2012), yayın yapmadaki acelecilik, şeffaflıktaki eksiklik, hakem değerlendirme sistemiyle ilgili sorunlar, kesin olmayan teoriler ve yetersiz istatistiksel prosedürler, küçük örneklemeler, uygun olmayan istatistiksel modellerle iyi tasarlanmamış deneyler yapma, yaygın metodolojik bilgi eksikliği gibi krizin sayısız nedeni tespit edilmiştir (Gelman ve Carlin, 2014; Maxwell, Lau ve Howard, 2015; Shrout ve Rodgers, 2018; Bird, 2021; Bryan, Tipton ve Yeager, 2021; Colling ve Szűcs, 2021).

Burada sayılan ya da yer almayan birçok nedeni tek tek ele almak bu makale bağlamında mümkün değildir.

Bu çalışmada istatistik metodolojisindeki sorunlu yaklaşımlar ve öne çıkan iki trend olan Yeni İstatistik ve Bayesyen İstatistik ele alınacaktır. Bu yöntemlere geçmeden önce krizin etrafında dönen tartışmalarda p değerinin oynadığı değinmek gerekmektedir. Özellikle P-hacking öne çıkan bir kavram olmuştur. Ayrıca araştırmacının sonuçları elde ettikten sonra hipotez oluşturduğu bir yöntem olan P-Hacking de tartışmalı araştırma uygulamalarına sebep olarak gösterilmektedir (Kerr, 1998). P-hacking, kısaca veri analizi sırasında araştırmacının istedik sonuçlar elde etmek için farklı yöntemler kullanarak anlamlı p değerini ($p < 0.05$) tutturmaya çalışmaktır. Uç değerleri seçerek çıkarmak, şişirilmiş I. Tip hata oranını kontrol etmeden yeni veri toplamak ya da kontrol değişkenlerine karar verirken bilinçli olarak seçici davranmak sık kullanılan örnekler olarak gösterilebilir (John ve diğerleri, 2012; Simmons ve diğerleri, 2011). Her ne kadar birçok p-hacking yöntemi olmasına rağmen hepsinin temel amacı anlamlı olmayan p değerini anlamlı hale getirmektir. P-hacking genellikle kötü niyetli bir amacın sonucu olsa da bazen veride gizli olan gerçek sonuçları ortaya çıkarmak amacı gibi iyi niyetten de doğabilir (Nelson ve diğerleri, 2018). Ayrıca birçok araştırmacı uyguladıkları veri analizi yöntemlerinin yanlış pozitif oranını (false-positive rates) artırdığının farkında olmamaktadır (Simmons ve diğerleri, 2011). Kimi araştırmacılar p değerini .05'ten küçük bulmak için çaba harcamaktadır. Bunun sonucu olarak P-hacking yöntemleri araştırma sonuçlarını beklendiği yönde manipüle etmek için kullanılmaktadır. Burada temel motivasyonun hızlı bir şekilde yayın yapmak olduğunu vurgulamak yerinde olacaktır (Siegfried, 2010; Pashler ve Wagenmakers, 2012; Wasserstein, Schirm ve Lazar 2019).

P değeri ilgili tartışmalar birçok farklı çözümün öne sürülmesini birlikte getirmiştir. Benjamin ve diğerleri, (2018) araştırmaların tekrarlanabilir olmamasının istatistiksel olarak anlamlı (statistically significant) bulunan bilimsel çalışmalara olan güveni sarstığını belirtmişlerdir. Bilimsel olarak anlamlı dediğimiz çalışmalar için anlamlılık değerinin çok düşük olduğunu vurgulamışlardır. İstatistiksel olarak anlamlı bulguları $p < 0,05$ ile ilişkilendirmek, yüksek oranda yanlış pozitif sonuç verdiğini bu sebeple istatistiksel anlamlılık eşliğinin $P < 0.05$ yerine $P < 0,005$ olarak bir değişmesini önermişlerdir. Bu değişiklik isteyenlerden işi biraz daha ileriye götürenler de olmuştur. Örneğin Basic and Applied Social Psychology (BASP) dergisi, p değerinin kullanımını ve sıfır hipotez anlamlılık testinin kullanımını dergi kapsamında tamamen yasaklamıştır (Trafimow ve Marks 2015). Kimi bilim insanları "istatistiksel olarak anlamlı" (statistically significant) terimini kullanmayı tamamen bırakmayı tavsiye etmişlerdir (Wasserstein, Schirm ve Lazar, 2019). Benjamin ve diğerleri, (2018) istatistiksel olarak anlamlı teriminin yeniden tanımlanmasını istemişlerdir. Lakens ve diğerleri, (2018) araştırmacıların anlamlılık düzeylerini doğrulamaları için bir dizi öneride bulunmuştur. Benjamin ve Berger

(2019) artık yeni bir döneme girdiğimizi bu dönemin adının ise “ $p < 0,05$ sonrası dönem” (post $p < 0,05$ era) olduğunu vurgulayıp p -değerinin büyük oranda kullanılmaya devam edeceğini düşündüğümüzde yanlış yorumlamayı önleyecek geçici öneriler sıralamışlardır. İlk olarak istatistiksel anlamlılık terimi kullanılıyorsa, 0,05 eşliğini 0,005 ile değiştirip “anlamlılık” (significant) yerine “önemli veya fikir verici” (suggestive) teriminin kullanılmasını tavsiye etmişlerdir.

Kimi araştırmacılar krizin abartıldığını belirtse de (Gibson, 2021) çoğu bilim insanı bir krizin var olduğunu kabul edip zayıf metodolojik ve istatistiksel uygulamaları bundan sorumlu tutmuştur. Tüm bu tartışmalar, genel olarak bilimin tüm alanlarında, özelde ise sosyal bilimler ve psikolojide yöntem, uygulamalı istatistiksel analiz, örnekleme ve araştırma uygulamaları alanında reformu zorunlu kılmıştır. Bu bağlamda, bu kısa zaman diliminde veri analizinde iki tarihsel geçiş yaşanmaya başlamıştır (Bknz. Şekil 1). Bunlardan ilki klasik hipotez testlerinde belirsizlik tahmininde etki büyüklüğü, güven aralığı ve meta-analizin kullanılmasını önceleyen “Yeni İstatistik” (the New Statistics) (Cumming, 2014); diğeri ise klasik istatistik yerine Bayesyen istatistiği kullanmayı savunan yaklaşımdır (Kruschke ve Liddell, 2015). Bu çalışmanın amacı öne çıkan iki trend olan klasik istatistik ve Bayesyen istatistiğin avantajları ve dezavantajlarını tartışmaktır.

Şekil 1.

Veri analizinde iki yeni yaklaşımın kavramsal gösterimi (Kruschke ve Liddell, 2018, s.2).



Yeni istatistik

Yeni istatistik klasik hipotez testlerinde H_0 hipotezi (NHST), anlamlılık testi yerine belirsizlik tahmininde etki büyüklüğü, güven aralığı ve meta-analizin kullanılmasını önermektedir (Cumming, 2014). Buna göre H_0 hipotezi anlamlılık testi mevcut etki hakkında her şey ya siyah ya da beyazdır anlayışını getirdiği için yetersiz bulmaktadır. Buna göre p değeri veriyi yorumlamak ve analiz etmek için oldukça sınırlıdır. Cumming'e göre etki büyüklüğü ve güven aralıkları, H_0 hipotezi anlamlılık testinden çok daha kapsayıcı bilgiler sunmaktadır. Ayrıca meta-analiz çalışmaları birçok çalışmadan elde edilen kanıtların bir araya toplanması ve yorumlanmasına olanak sağlar. Yeni istatistik tekniklerini kullanmak araştırmacılara daha çok bilgi sağlayarak bilimin de hızlı ilerlemesine katkı sağlayacaktır (Cumming, 2013).

Cumming (2014) H_0 hipotezi anlamlılık testinin hiç kullanılmamasını önermektedir. Çünkü son yıllarda H_0 hipotezi anlamlılık testi ile güven aralıkları verilmekte ve referans olarak Cumming gösterilmektedir. Güven aralıkları p değeri ile verilmekte ancak sonuç ve yorumlamalar genellikle p değeri üzerinden yapılmaktadır. Güven aralıkları yorumlanırken güven aralığının sıfır içerip içermediği gibi basit bir açıklama kullanmak p değerinde düşülen indirgemeci basit hataya tekrar düşüldüğünü göstermektedir. Bu yaklaşım her zamanki gibi H_0 hipotezi anlamlılık testinde yaşanan eski sorunların tekrar etmesine sebep olacak ve araştırma bütünlüğüne ve kümülatif nicel bir disipline yönelik ekstra bir etkiye sahip olmayacaktır. Cumming mümkün olduğunda araştırmalarda p değerinin raporlanmamasını bunun yerine güven aralıklarının raporlanmasını ve kapsayıcı bir şekilde yorumlanmasını önermektedir. Nitekim kitabında bunun nasıl yapılacağını 8 adımda kapsamlı bir şekilde anlatmıştır (bknz. Cumming, 2013).

Yeni istatistiğin bir diğer önemli bileşeni olan etki büyüklüğü araştırmacının ilgilendiği özne ile ilgili olan her şeyin miktarıdır. Aslında bu bir çalışmada bağımlı değişkenimizin ölçtüğü şey ile ilgilidir. Terimin kendisi hem Türkçede “etki” hem de İngilizcede (effect) yarattığı tuhaflık yüzünden sanki birebir bir etki olması gerekiyor hissiyatı yaratsa da bunun gerekli olmadığı vurgulanmaktadır. Evrenin etki büyüklüğü basitçe evrendeki bir etkinin gerçek değeridir (Cumming, 2014). Araştırmalarda şeffaf iletişim tercih etmek bilim camiasında anlaşmayı teşvik etmenin anahtarıdır. Diğer yandan raporlanan popüler etki büyüklüklerinin sorunlu yapısına değinmek gerekmektedir. Örneğin sıkça raporlanan Cohen d en az bilgilendirici olan değerlerden birisi olagelmıştır. Raporlanan diğer değerlerle birlikte Cohen U3 değerini raporlamak bulguları daha erişilebilir hale getirebilir (Hanel ve Mehler, 2019).

Belirli bir alandaki araştırma literatürünü nicel olarak özetlemek için kullanılabilecek en önemli metodolojik araç meta-analizdir (Gurevitch ve diğerleri, 2018). Meta-analizler, genel bir anlayışa ulaşmak için aynı araştırma sorusunu ele alan çok sayıda çalışmanın sonuçlarını özetler. Bu nedenle, analiz birimi bireysel düzeyden toplu düzeye, ideal olarak belirli bir araştırma sorusuna göre toplanan tüm kanıtlara doğru değişir. Meta-analizlerin birkaç güçlü yönü vardır. Meta-analiz çalışmaları sayesinde araştırmacılar çalışılan bilimsel kavramı bir bütün olarak görebilmektedir. Meta-analizde etki büyüklüğü kullanışlı katkı sağlayarak, H0 hipotezinde test edilmesinde yaşanan problemleri ortadan kaldırmaya yardım etmektedir. Meta-analiz çalışmanın istatistiksel anlamlılığı göz önünde tutması, verileri etkili kullanması, araştırma sorularına odaklanması ve düzenleyicilik etkisini işin içine dahil etmesi ile birçok avantaj sunmaktadır (Friesse ve Frankenbach, 2020).

Son kırk yılda meta-analiz, araştırmaları özetlemek için oldukça popüler bir yöntem olarak ortaya çıkmıştır. Tekrarlanabilirlik krizinde öne çıkan yöntemlerden olan meta-analizi bu kadar çekici kılan kimi unsurlar bulunmaktadır. İlk olarak, yürütülen çok sayıda meta-analiz araştırmacıların önceki araştırmaların genel olarak özetlenmesiyle ilgilenmelerinin yanında tekrarlanabilirlikleri ile de sıklıkla meşgul olduklarını göstermektedir. İkinci olarak psikolojide yayınlanan meta-analiz çalışmaları azımsanmayacak düzeyde sıfır olmayan etki büyüklükleri üretmekte bu da psikolojide yapılan çalışmaların tekrar ettiği izlenimini vermektedir (Cafri, Kromrey ve Brannick, 2010). Üçüncü olarak meta-analizlerde bulunan çalışmaların birleşimi herhangi bir tek çalışmadan daha güçlü (robust) olarak görülmektedir (Williams, Polanin ve Pigott, 2017). Son olarak meta-analizler, psikoloji araştırmalarının hangi alanlarının tekrarlandığını ve hangilerinin tekrarlanmadığını gösteren açıklayıcı bir işlev görmektedirler (Siddaway, Wood ve Hedges, 2019). Araştırmacılar meta-analiz çalışmalarının tekrarlanabilirlik krizinde etkili olabilmesi için kimi önerilerde bulunmuşlardır (Sharpe ve Poets, 2020). Buna göre; meta-analitik uygulamalar, yayımlanmış meta-analizlerin sistematik incelemelerini içermelidir. Yayınlanan meta-analizlere hangi bilgilerin dahil edileceğine ilişkin yönergeler izlemez. Örnek ve iyi tasarlanmış meta-analitik uygulamalarla ilgili çalıştaylar düzenlemeli, el kitapları ve yönergeler hazırlanması önerilmektedir. Henüz incelenmemiş araştırma konularının meta-analizlerinin yapılması için teşvik sistemi oluşturabilir (Sharpe ve Poets, 2020).

Yeni istatistiğin getirdiği öneriler kuşkusuz çok büyük katkılar sunmaktadır. Araştırmacıları p değerine bağımlı olmaktan kurtarması ve bir alternatif sunması oldukça değerlidir. Ancak yeni istatistiğin de kimi sınırlılıkları bulunmaktadır. Yeni istatistiği tercih etmek Bayes yaklaşımıyla gelen faydalardan vazgeçmek anlamına gelmektedir. Kimi araştırmacıların vurguladığı gibi güven aralıkları yanılgılara ve yanlış yorumlara meyilli olmaları sebebi ile p değeri aynı kaderi paylaşma olasılığı taşımaktadırlar (Morey ve diğerleri, 2016). Bir diğer

sınırlılık ise, güven aralıklarının incelenen etkinin var olduğunu varsaymasıdır; başka bir deyişle, her çıkarım sorununun, hipotez testinden ziyade bir parametre tahmini sorunu olduğu anlamına gelmesidir. Etki büyüklüğü tahmininin önemli olduğu ve dikkat edilmesi gerekse de büyüklük sorusu (“etki ne kadar büyük?”) ancak mevcudiyet sorusu (“bir etki var mı?”) ikna edici bir şekilde ele alındıktan sonra devreye girmesi gerekmektedir (Morey ve diğerleri, 2014). Bu ve benzeri birçok eleştiriden dolayı Bayesyen istatistik yeni alternatif olarak öne çıkmaktadır.

Bayesyen İstatistik

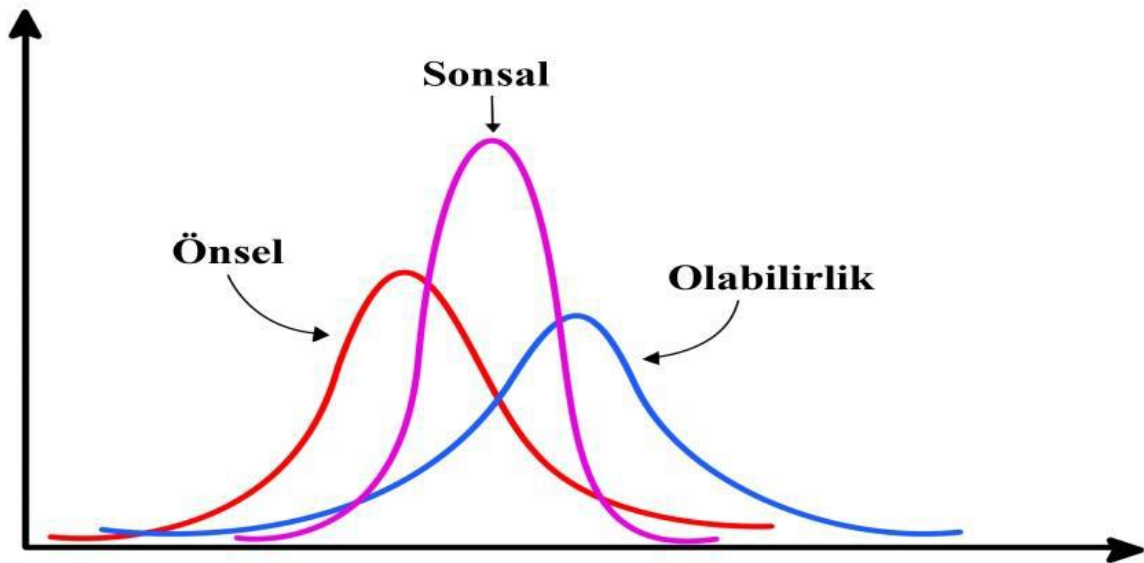
Artan teknolojik ve metodolojik gelişmeler ile günümüzde istatistiksel problemlerin çözümü için daha karmaşık modeller oluşturulmakta ve verinin modeli en iyi yansıtan sonucuna ulaşılmaya çalışılmaktadır. Modeller ile ilgili karar verilirken klasik yaklaşımda, P-değeri temel karar verici olarak ele alınmaktadır. P-değerine, yanlış kullanımı ve yorumlamada getirdiği zorluklar sebebiyle son yıllarda analizlerde daha az yer verilir olmuştur (Benjamin ve diğerleri, 2018). İstatistiksel parametre tahmininde klasik istatistiğe alternatif olarak model seçimlerinde ve hipotez testlerinde Bayes istatistik yaygın olarak kullanılmaktadır. Bayesyen yaklaşım, bilimsel karar verme sürecinde olasılık gibi oldukça önemli bir kavrama dayanmaktadır. 1783’te Thomas Bayes tarafından ortaya atılan öznel olasılık kavramından doğan yaklaşım, Bayes teoremine dayandığı için Bayesyen yaklaşım ismini almıştır (Bayes, 1763). Bayesyen istatistik, istatistiksel bir modeldeki parametreler hakkındaki bilginin geçmiş çalışmalardaki verilerdeki bilgilerle güncellendiği Bayes teoremine dayalı bir veri analizi yaklaşımıdır (Bayes, 1763).

Davranış bilimlerindeki birçok problem için, araştırmacılar, benzer veya geçmiş verilere ait analizlerden veya diğer kaynaklardan iyi bir ön bilgiye sahip olabilirler. Bilinmeyen parametreler hakkında daha önceki çalışmalardan elde edilen bu bilgiler önsel (prior) olarak adlandırılır (Erkan, 2019). Yani parametreler hakkında daha önceki deneyimlerden elde edilen bilgiler, önsel (prior) bilgidir. Önsel bilgi, Bayesyen çıkarımlarda önemli bir rol oynar (Gill, 2002). Önsel bilgi ile gözlem verilerine ilişkin olabilirlik fonksiyonu birlikte kullanılarak elde edilen dağılımlara, sonsal (posterior) dağılımlar denir. Sonsal dağılımların ardından, parametreler için sonsal tahminler elde edilir. Sonsal dağılımlar, bilinmeyen bütün gözlenemeyen (gizil) parametreler hakkında bilginin güncel durumunu ortaya koyar (Congdon, 2001). Bayesyen yaklaşımda sonsal dağılımın (posterior distribution) elde edilebilmesi için önsel dağılımın belirtilmesi gerekir. Model parametreleri için önsel dağılımın belirtilmesi ve olasılıkları hakkında doğrudan açıklama yapılmasına izin vermesi, klasik yaklaşıma göre ayırt edici bir özellik olarak ortaya çıkmaktadır. Ancak bunu yapabilmek için araştırmacının analiz öncesinde model parametreleri için hangi önsel dağılımı kullanacağını belirlemesi gerekmektedir (Erkan, 2019).

Sonsal dağılımlar önsel dağılımların bir fonksiyonudur. Örneklem büyüklüğünün artması, önsel bilginin sonsal bilgi üzerindeki etkisinin azalması anlamına gelir. Önsel bilginin miktarı, sonsal dağılımın üzerinde belirleyici önemli bir role sahiptir. Bu yüzden önsel dağılımın seçimi araştırmacının ne kadar bilgi ve tecrübeye sahip olduğuna, bilginin ne kadarının doğru olduğuna veya ne kadarını kullanması gerektiğine inanmasına göre değişebileceği, Bayes analizinin temel unsurudur (Şehribanoğlu, 2012). Sonuç olarak Bayesyen yaklaşımda sonsal dağılımın elde edilme süreci, temeli Bayes teoremine dayanan Bayesyen yaklaşımlarla elde edilir. Bayesyen yaklaşımda bilinmeyen parametre θ sabit bir parametre değil, bir raslantı değişkeni olarak kabul edilir. Bu sayede klasik yaklaşımın aksine parametreler hakkında olasılıklar dikkate alınarak yorumlanabilir. Bayesyen yaklaşımın temel amacı, veri elde edildikten sonra bilinmeyen parametre dağılımlarını elde etmek için önceden elde edilmiş bilgilerden yararlanarak (önsel) yeni bilgilerle (sonsal) parametrelere ait tahminlerde bulunmaktır (Ekici, 2009). Bayesyen yaklaşımda parametre için önsel, olabilirlik ve sonsal dağılım Şekil 2’de verilmiştir

Şekil 2.

Parametre İçin Önsel, Olabilirlik ve Sonsal Dağılım (Erkan, 2019, s.27).



Bayesyen yaklaşımın uygulamasında önsel dağılım kullanılarak elde edilen sonsal dağılımlar kullanıldığından uzun yıllar uygulama aşamasında zorluklar yaşanmıştır. Sonsal dağılımların elde edilmesi uzun zaman almakta, yüksek boyutlu integrallerin kullanılması sebebiyle genel kullanımı yaygınlaşamamıştır. Ancak Markov Zinciri Monte Carlo yöntemleri (MCMC) Metropolis-Hasting algoritması, Gibbs örnekleme algoritması ve gelişen hesaplama teknolojileri sayesinde bu yaklaşım öncelikle fizik alanında ve uzay bilimlerinde kullanılmış ve daha sonra diğer alanlarda kullanılmaya başlanmıştır (König ve van de Schoot, 2017).

“Bayesyen yaklaşım, özü Bayes Teoremine dayandırılarak yapılandırılmış bir yaklaşım sistemidir. Bu disiplinin başlangıç aksiyomlarının yorumlanmasında, pek çok konu ve kavramın ele alınışında bu yaklaşımlardan biri diğerine alternatif olmuştur. Dolayısıyla pek çok istatistiksel kavram bu yaklaşımda farklı yorumlanmakta ve ele alınmaktadır (Ekici, 2009).“

Genel olarak klasik istatistik sıfır (yokluk) hipotezi ile psikoloji ve temel bilimlerde yaşanan tekrarlanabilirlik krizinin bir parçası olan p değerinin yanlış kullanılması ilişkilendirilmektedir (Chambers, ve diğerleri, 2014; Szucs ve Ioannidis, 2016). Literatürde Bayesyen yaklaşımın kullanımının yaygınlaşması ile bu sorunların büyük oranda ortadan kalkacağı yönünde fikir birliği oluşmuştur (Benjamin ve diğerleri, 2018; Etz ve Vandekerckhove, 2016).

“Bayes istatistiğine ilişkin yukarıdaki bilgiler göz önünde bulundurulduğunda, klasik yaklaşım ile Bayes yaklaşımı arasındaki temel farkın, parametrelerin doğasından kaynaklandığını söylemek mümkündür. Klasik yaklaşımda popülasyonlar için parametrelerin sabit karakterli olduğu ve örneklemden sağlanan bilgiler ile tahmin edilebileceği kabul edilir. Buna karşın Bayes yaklaşımında, parametrelerin rassal olduğu varsayılır ve bir dağılım formu ile karakterize edileceği kabul edilir. Başka bir ifade ile klasik yaklaşımın aksine Bayes yaklaşımında bütün bilinmeyenlerin bir olasılık dağılımına sahip oldukları dikkate alınır. Her iki yöntem arasındaki bu temel farklılık, tahminleme ve yorumlama sürecinde de göz önünde tutulur (Şehribanoğlu, 2012). “

Bayes istatistiksel analizi, daha sonraki inançları (posterior beliefs) elde etmek için veriler (data) ışığında parametreler ve hipotezler hakkındaki önceki inançları (prior beliefs) nasıl güncelleyeceğimizi söylemektedir. Bayes Teoreminin kendisi doğrudan koşullu olasılığın geleneksel tanımından hareket etmektedir. Bunu şu şekilde ifade edebiliriz: önceki inançlar (prior beliefs) → veri (data) → sonraki inançlar (posterior beliefs) veya matematiksel bir ifade ile $p(\theta) \rightarrow y \rightarrow p(\theta|y)$ şeklinde gösterilmektedir.

Veri elde edilmeden önce parametreye ilişkin kesin olmayan bilgilerimizi kapsayan olasılık dağılımına önsel dağılım denilmektedir. Sonsal dağılımı elde etmek için ise önsel dağılım ve örnekleme ilişkin bilgiler kullanılır. Bir parametre ile ilgili bütün çıkarımlar elde edilen sonsal dağılım aracılığıyla yapılır. Bayes tahminleme bulunan sonsal dağılımın beklenen değeridir (Jackman, 2009). Bayes teoremi gözlenen olayların olasılıkları açısından ifade edilir. Buna A ve B gibi iki olayın sırası ile gerçekleşme olasılıkları dersek bu iki olayın koşullu olasılığı aşağıdaki eşitlikteki gibidir.

$$P(A; B) = \frac{P(B; A) \times P(A)}{P(B)}$$

Bayes teoremi koşullu olasılıklara dayanır. Bir olayın gerçekleşme olasılığının ek bilgilerle değişebileceğini gösterir. Yani bir örneklem uzayındaki A ve B gibi iki olayın marjinal

ve koşullu olasılıkları arasında ilişki kurar. $P(A)$ ve $P(B)$ sırasıyla A ve B olayının gerçekleşme olasılıkları olmak üzere bu iki olaya ait koşullu olasılık yukarıda verilmiştir. Eşitlikte ifade edilen Bayes teoremi $P(B)>0$ olma koşulu ile özetle B'nin gerçekleştiği durumda A'nın gerçekleşme olasılığını açıklamaktadır. Burada; $P(B|A)$, A'nın gerçekleştiği durumda B'nin gerçekleşme olasılığını, $P(A)$ A'nın marjinal olasılığını (A'nın önsel olasılığı olarak da adlandırılır.), $P(B)$, B'nin marjinal olasılığını ve $P(A|B)$, B'nin gerçekleştiği durumda A'nın gerçekleşme olasılığını ifade eder. Genel olarak $P(A|B)$, B'nin olasılığını dahil ederek hesaplanan A'nın sonsal olasılığı olarak adlandırılır. Burada $P(B)$, $P(A|B)$ olasılık değerinin 0 ile 1 arasında olması için normalleştirme sabiti olarak işlev görür (Erkan, 2019).

Eşitliği Koptur'un (2020) verdiği bir örnek ile açıklamak gerekirse;

- Bir hastalığın görülmesi olasılığı %1,
- Bir bireyde hastalık mevcut ise testin pozitif çıkma olasılığı %80 olsun,
- Bir bireyde hastalık mevcut değilken testin pozitif çıkma olasılığı %9.6 olsun. Bu durumda, test pozitif çıktığında bireyin gerçekten hasta olma olasılığı

$$p(\text{BireyHasta}|\text{TestPozitif}) = \frac{p(\text{TestPozitif}|\text{BireyHasta})p(\text{BireyHasta})}{p(\text{TestPozitif})}$$

olduğundan,

$$p(\text{BireyHasta}|\text{TestPozitif}) = \frac{0.8 * 0.01}{0.01 * 0.8 + 0.99 * 0.096} = 0.077$$

bulunur. Yani, test pozitif çıktığında bireyin gerçekten hasta olma olasılığı yaklaşık olarak %7.8dir. Yukarıdaki örnekte, bireyin aynı testten bir kere daha yaptırdığını düşünelim. Bu durumda bireyin hasta olma olasılığını 0.077 olarak güncellersek

$$p(\text{Birey Hasta} | \text{Test Pozitif}) = \frac{0.8 * 0.077}{0.077 * 0.8 + 0.923 * 0.096} \approx 0.41$$

olur. Yani iki test de pozitif sonuç verdiğinde bireyin hasta olma olasılığı yaklaşık %41 olmaktadır. Bayes teoremi, yeni bilgiler edindikçe olasılıkları güncellemeye doğal yoldan izin vermektedir.

Bayesyen tahminleme yöntemleri kullanımı birçok bilim alanında yükselişe geçmiştir (Kruschke, Aguinis ve Joo, 2012; Rietbergen, ve diğerleri, 2017; van de Schoot, ve diğerleri, 2017). Son on yılda ise Bayes için popüler yazılımların artması ve klasik yaklaşımın zorlandığı çok karmaşık ve fazla hesaplama gerektiren modellerde Bayes yaklaşımın getirdiği esneklik ve kolaylık sayesinde özellikle psikoloji alanında dikey bir yükseliş olmuştur (van de Schoot ve diğerleri, 2017; Wagenmakers ve diğerleri, 2008). Yapısal eşitlik modellemelerinde küçük örneklem ve modelin karmaşıklığının klasik istatistikte kabul edilemez parametre çözümleri ve yanlış tahminlerle sonuçlandığı görülmüştür (Muthén ve Asparouhov, 2012; Wagenmakers ve

diğerleri, 2008). Bu sosyal bilimlerde doğal olarak küçük olan popülasyonlarda (Veen ve Egberts, 2020), erişilmesi zor hedef gruplarda (Coleman ve diğerleri, 2002) ya da finansal yetersizlikler nedeniyle veri toplamanın zor olduğu gruplardaki çalışmalarda Bayesyen veri analizinin kullanımı büyük bir konfor sunmaktadır.

Bayesyen yaklaşımın klasik istatistiğe göre birçok avantajı bulunmaktadır. Küçük örneklemelerde klasik istatistik yerine Bayesyen kullanılması literatürde yaygın olarak önerilmektedir (Wagenmakers, ve diğerleri, 2008). Klasik istatistikte parametre tahmininde Maksimum Olabilirlik tahminleme yöntemi kullanılmakta ve asimptotik olarak normal dağılım varsayımının sağlanması gerekmektedir. Ancak tıp ve psikoloji gibi büyük örneklem oluşturmanın zor olduğu ya da sosyal bilimlerdeki kategorik ve kayıp verinin olduğu alanlarda çok değişkenli normalliği sağlamak oldukça zordur. Bu da parametre tahminlerinde yanlış sonuçların ortaya çıkmasına sebep olabilmektedir (Rupp ve diğerleri, 2004). Bayesyen yaklaşım özellikle önsel bilgiyi kullanarak sonsal dağılımlar elde etmesinin de etkisi ile tekrarlanabilirlik krizinin istatistik boyutuna büyük bir yenilik ve fayda getirmektedir (Rupp ve diğerleri, 2004).

Çalışmanın başında da vurgulanan The Reproducibility Project: Psychology adlı çalışma ile psikolojideki tekrarlanabilirlik krizi ile ilgili ilk bilimsel araştırma yayınlandığında analizler klasik istatistik ile yapılmıştır. Araştırmaların sadece %36'sında anlamlı bulgulara ulaşılmıştır (p değeri 0.05'in altında). Orijinal çalışmalarda raporlanan etki büyüklüğünün ise tekrarında yaklaşık yarısı kadar olduğu görülmüştür (Open Science Collaboration 2015).

Etz ve Vandekerckhove (2016) The Reproducibility Project: Psychology adlı çalışma yayınlandıktan bir yıl sonra açık olan veriyi kullanarak analizleri Bayesyen istatistik ile tekrarlamışlardır. Klasik istatistikten farklı olarak hipotezler hem H_0 hem de alternatif hipotez için test edilmiştir. Çalışmada yayımlanmaya bağlı yanlışlıkta ele alınmıştır. Genel olarak çalışmaların %75'i elde kanıtlar bağlamında benzer sonuçlar üretmiştir. Elde kanıtlar çoğunlukla Bayes dilinde konuşmak gerekirse oldukça zayıf bulunmuştur (yani Bayes faktörü < 10). Çalışmaların çoğu (%64), hem orijinalde hem de tekrar edilende çalışmada H_0 hipotezi veya alternatif hipotez için güçlü kanıt sağlamamıştır. Araştırmacılar Reproducibility Project'te birçok çalışmanın tekrar etmemesinin sebebi olarak bu çalışmalarda psikoloji literatüründeki küçük örneklem ve yayın yanlışlığı kaynaklı olan etki büyüklüklerinin abartılmasını göstermişlerdir. Bunun sonucu olarak geleneksel istatistiğin ihtiyacı olan büyük örneklem ihtiyacı Bayesyen yaklaşımın önemini ortaya çıkarmıştır.

Psikoloji biliminde küçük örneklem ve güç bakımından yetersiz çalışmalar yaygın olmakta, yayın yanlışlığı ise hâkim olan bir durum olmuştur. Bu iki zayıflık yanlış sonuçlar üreten bir literatür yaratmıştır (Ioannidis, 2005). Reproducibility Project: Psychology projesinin Bayesyen yeniden analizi sonucunda birçok çalışmanın tekrar etmemesinin sebebi olarak varlıklarına dair kanıtların başlangıçta kabul edilemez derece zayıf olması gösterilmiştir.

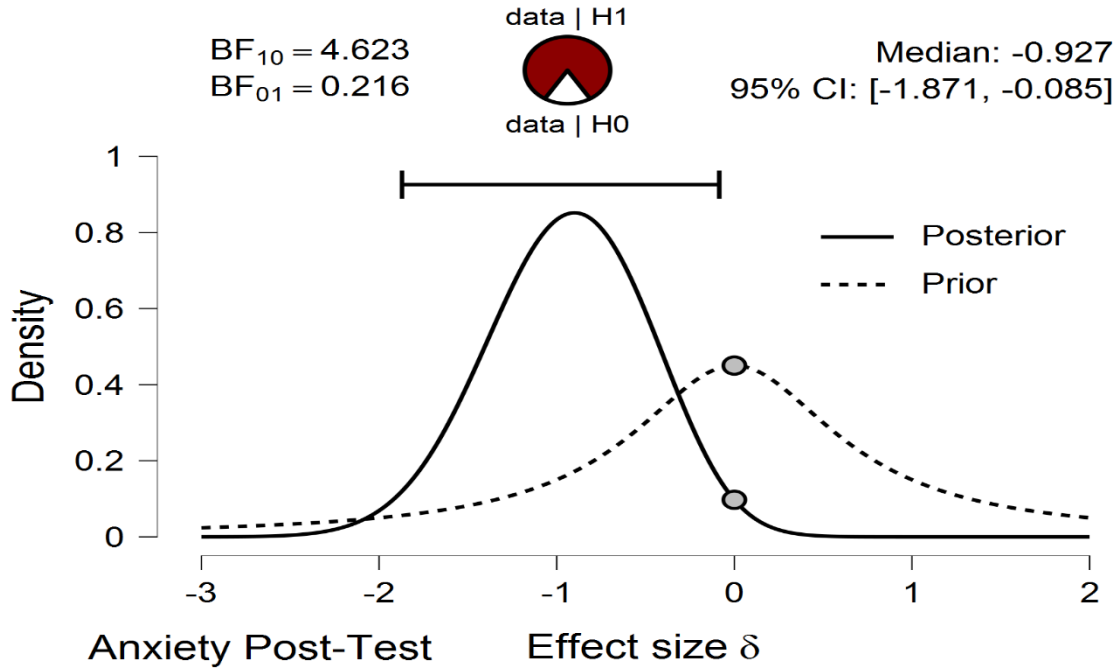
Bayesyen analizde orijinal ve replikasyon sonuçları arasında belirgin bir tutarsızlık bulunmadığı sonucuna varılmıştır. Aslında, çalışmaların %75'inde replikasyon çalışması, yayın yanlılığı olasılığını dikkate aldıktan sonra orijinal çalışma ile benzer sonuçlara ulaşmıştır. Anlamlı bir fark bulunan çalışmalarda da bu fark örneklem büyüklüğüne bağlanmıştır (Etz ve Vandekerckhove, 2016).

Reproducibility Project: Psychology, alanı dönüştüren ve sonuçlarının etkisinin hemen görüldüğü önemli bir çalışma olmuştur. Araştırmacılar psikolojide deneysel araştırmaların, mümkün olduğunda, etki büyüklüklerine dayalı güç analizlerinden elde edilen örneklem büyüklüklerini aşması gerektiğini vurgulamışlardır (Etz ve Vandekerckhove, 2016). Yıllarca kabul gören ortodoks istatistiğe dayanmak genel olarak psikoloji çalışmalarını (Berger ve Sellke, 1987) ve özel olarak da replikasyon çalışmalarını küçük örneklem sorununun gölgesinde bırakmıştır.

Bayesyen istatistiksel yöntemlerin en büyük bir avantajı, örneklem büyüklüklerinin önceden belirlenmesinin gerekli olmamasıdır. Bu durum, veri analizinde araştırmacıların verileri incelemesine ve yeterli bilgi toplandığına kanaat getirildiklerinde veri toplamayı bırakma gibi olanak tanımaktadır (Rouder, 2014). Bu tür bir örneklem planının nasıl uygulaması gerektiğini ayrıntılı açıklaması için Matzke ve diğerleri (2015) incelenebilir. Ayrıca bu tasarımın detaylı adım adım kılavuzunu ve tartışmasını için Schönbrodt ve diğerleri (2017) okunabilir. Sonraki aşamada planlanan örneklem büyüklüğüne ulaşırsa ve sonuçlar bilgi vermezse, daha fazla veri toplanabilir veya araştırmacılar burada durup sonuçlarındaki belirsizliği kabul edip bunu yayınlayabilir (Rouder, 2014).

Bayes faktörleri, teoriyi veriyle ilişkilendirmek için bir modelin diğerine (örneğin, H1'e karşı H0) kanıtların simetrik bir ölçüsünü sağlar. Bu özellikler, psikolojideki tekrarlanabilirlik krizinin bazı sorunlarının çözülmesine yardımcı olabilir. Kanıtın simetrisi, H1 için olduğu kadar H0 için de kanıt olabileceği anlamına gelir; ya da Bayes faktörü her iki yönde de yetersiz kanıt gösterebilir (Dienes, 2016). Makalenin yazarının Bayes istatistik kullandığı araştırmasındaki (bkznz: Kara, Türküm ve Turner, 2023) örneği inceleyecek olursak; yapılan çalışmada kaygı son testinin deney ve kontrol grubu açısından farklılaşıp farklılaşmadığı T testi ile sınanmıştır (Şekil 3). Pizza grafiğinde görüldüğü gibi Bayes T testi sonucu hem H1 hem de H0 için sonuç vermektedir.

Bayes faktörleri verilerin bir teoriye karşı olup olmadığı (ve hiçbir şey için sayılmadığını) gösterir böylece, tekrarlanabilirlik çalışmalarında aslında H0 mi veya H1'i mi desteklediği sonuçlara kolayca ulaşılabilir. Kanıtların her iki yönde de aynı derecede güçlü olduğu ölçümler yapılabilir. H0 ve H1 hipotezlerini destekleyen kanıtların (sonuçların) elde edilmesi sayesinde akademik yayın sırasında sonuçlar anlamlı mı değil tartışmasını anlamsız kılabilir. Bilimsel yayınlara denge gelmesine yardımcı olacağı düşünülmektedir (Dienes, 2016).

Şekil.3.*Bayes T testi grafiği*

Bayes faktörleri de haclenebilir ancak Bayes istatistikte bu sorunu çözebilecek bir birkaç önemli özellik bulunmaktadır. Öncelikle Bayes her iki yönde kanıtlara izin vermektedir. Böylece araştırmacıların sadece bir yönde hile yapma eğilimleri (klasik istatistik dili ile; pozitif yönde sonuç bulma) daha az olabilmektedir. Bayes faktörlerdeki durma süresi prensibi (The Stopping Rule Principle) bir kanıt ölçütü olarak durma süresine duyarsız olmaktadır. Test ailelerinin keyfi bir şekilde tanımlanmasına olanak vermez. Çelişkili bir durum planlandığında post hoc testleri önemsiz hale gelmektedir (Dienes, 2016).

Bayesyen yaklaşımın yukarıda sayılan avantajlarına ek olarak klasik istatistiğe göre birçok avantajı bulunmaktadır:

- Bayesyen yaklaşım klasik istatistikte olduğu gibi hangi modelin doğru olduğunu söylemek yerine eldeki veri ve diğer bilgiler ışığında model seçmek için tercih sunmaktadır. Bu şekilde geleceğe yapılacak tahminlerde en iyi modellerin seçilmesine olanak tanımaktadır (Kruschke, Aguinis ve Joo, 2012).
- Bayesyen hipotez testleri klasikteki gibi sıfır (yokluk) hipotezine dayanmamaktadır. Test edilebilir hipotez aralığı genişler ve alternatif hipotez hakkında da bilgi vererek önemlilik testine dayanmayan sezgisel yorumlar yapılabilinmektedir (Kruschke, Aguinis ve Joo, 2012).
- Mevcut verilere ek olarak gerçek ön bilgi kullanımına olanak verdiği için veri için sonsal dağılımın ortalama ve yüzdelik gibi yararlı istatistiklerini temin ederek daha

güvenilir sonuçlar vermektedir. Mevcut tüm bilgi doğru şekilde kullanıldığında daha etkili sonuçlar vermektedir (Lee ve Song, 2004).

- Bayesyen yaklaşım daha az hesaplama gerektirdiği için daha fazla ve yeni modeller kolaylıkla analiz edilebilmektedir. Bu anlamda zaman ve maliyet tasarrufu da sağlamaktadır. (Lee ve Song, 2004).
- Bayesyen tahminlemenin ardışık yapısından dolayı çıkarsama ile parametreler üzerindeki belirsizlikler azaltılmaktadır (Makowski, ve diğerleri, 2019).
- Bayesyen yaklaşımda klasikte çözümü olmayan birçok problemin çözümü kolayca bulunmaktadır (örn: Behrens-Fisher problemi) (Dienes ve Mclatchie, 2018).
- Bayesyen yaklaşım daha güvenilir sonuçlar vermektedir, örneğin karışık ve kirli veri grubunda daha doğru sonuçlar vermekte, tip 1 hata hataya yatkınlığı oldukça düşürmekte ve önsel bilginin eklenmesi ile daha doğru ve güvenilir sonuçlar vermektedir (Andrews ve Baguley, 2013; Etz ve Vandekerckhove, 2016).
- Bayesyen yaklaşımla sadece bilinmeyen parametreler için sonsal dağılım değil, gözlem değerleri için de sonsal öngörü dağılımları elde edilmektedir. Bu nedenle öngörü amacıyla da kullanılmaktadır. Bayesyen yaklaşım geçmişi görerek, sorgulama yapmaktadır (Kruschke, Aguinis ve Joo, 2012; Wagenmakers ve diğerleri, 2018).
- Pek çok uygulamalı çalışmalarda, son zamanlara kadar karmaşık problemlerin Bayes analizi oldukça güç ve probleme özel metod geliştirmeyi gerektirmekteydi. Ancak Markov zinciri kullanarak Monte Carlo simülasyon yöntemi sayesinde, gerekli nümerik integrallerin alınması sağlanabilmektedir. Bu yöntemle, sonlu sayıda gözlem değeri kullanarak, sonsuz sayıda veri elde etmek mümkündür. Böylece çözümü analitik olarak zor olan bazı problemlerin, benzetim teknikleri ve bilgisayar yazılımları sayesinde hızlı biçimde çözülmesi sağlanabilmektedir (Yardımcı ve Erar, 2005).

Goodman (2019) Bayes faktörlerin çok çekici olduğunu ancak araştırmacıların bunu nerede ve nasıl kullanacağı konusunda bilgi sahibi olmadıklarını belirtmektedir. Bu durumu Bitcoin'e benzetmektedir. Bitcoin'inin büyük ihtimalle geleceğin parası olduğunu ama insanların mevcut durumda nasıl yaklaşacakları ve kullanacakları konusunda kafalarının karışık olduğunu belirtmiştir. Ancak kendi kendine giden arabalarının tasarımcıları, yapay zeka araştırmaları, makinelerin kalıpları tanınmasında ve karar vermesine yardımcı olmasında Bayes yazılımlar kullanıldığını belirtmektedir.

Bayes faktörlerin kullanılması konusunda kimi araştırmacılar dikkatli olunması gerektiğini belirtmişlerdir. Örneğin Gelman ve Carlin (2017), hipotez testi için Bayes faktörlerinin kullanılması, aynı amaç için kullanıldığında p-değerlerinin yarattığına benzer birçok probleme sebep olabileceğini, p-değerlerini basitçe Bayes faktörleriyle değiştirmenin bir çözüm olmadığı konusunda uyarılmışlardır.

Sonuç

Sonuç olarak yıllardır klasik istatistik dilinde yazma, çizme ve düşünme alışkanlığı geliştiği için Bayesyen istatistiğin dili ilk defa duyanlara farklı gelebilmektedir. Bayes yöntemi Sherlock Holmes çıkarım yapmak için kullandığı tekniklerle benzetilmektedir. Eldeki ip ucundan hareketle mevcut teori ve hipotezler güncellenmektedir. Yeni bilgiler geldikçe sonuçlara ilişkin yeni çıkarımlar yapılmaktadır. Geline nokta da mevcut tabloya göre, bilimsel araştırmalarda kullanılan istatistiksel tekniklerin değiştirilmesi gerektiği konusunda bir fikir birliği oluşmuş durumdadır; ancak yapılacak değişikliğin hangisi olacağı konusunda tartışmalar devam etmektedir. P değerlerinin yanlış yorumlanması ve kötüye kullanılması ile ilgili tartışma devam etmektedir. Anlamlılık ile ilgili eşik değişse bile mevcut sorunlar devam edecektir. Bu çalışmada öne çıkan iki alternatif olan Yeni İstatistik ve Bayesyen İstatistik konuları kısaca ele alınmaya çalışılmıştır. Yeni İstatistik ile yola devam etmek klasik istatistik kullanmaya devam etmek anlamına gelmektedir. Ancak Bayesyen İstatistiğin kullanılması bilimsel araştırmalara yeni bir soluk getireceği düşünülmektedir. Bayesyen yaklaşımın klasik istatistiğe göre birçok avantajı bulunmaktadır. Küçük örneklerde klasik istatistik yerine Bayesyen kullanılması literatürde yaygın olarak önerilmektedir (Rupp ve diğerleri, 2004). Klasik istatistikte parametre tahmininde Maksimum Olabilirlik kullanılmakta ve asimptotik olarak normal dağılım varsayım gerekmektedir. Ancak tıp ve psikoloji gibi büyük örneklem oluşturmanın zor olduğu ya da sosyal bilimlerde gibi kategorik ve kayıp verinin çok olduğu alanlarda çok değişkenli normalliği sağlamak oldukça zorlayıcı olmaktadır. Bu da parametre tahminlerinde yanlış sonuçların ortaya çıkmasına sebep olabilmektedir (Rupp ve diğerleri, 2004). Bayesyen yaklaşım özellikle önsel bilgiyi kullanarak sonsal dağılımlar elde etmesinin de etkisi ile bu krizin istatistik boyutuna büyük bir yenilik ve fayda getirmektedir.

Yazar katkısı: makalenin tek yazarı olarak makaledeki içerikten sorumludur.

“Bu çalışma, kapsamı gereği etik kurul onayı gerektirmemektedir.”.

Kaynakça

- Amrhein, V. ve Greenland, S. (2018). Remove, rather than redefine, statistical significance. *Nature human behaviour*, 2(1), 4-4. <https://doi.org/10.1038/s41562-017-0224-0> adresinden 28.3.2022 tarihinde erişilmiştir.
- Andrews, M., ve Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, 66(1), 1-7. <https://doi.org/10.1111/bmsp.12004> adresinden 28.3.2022 tarihinde erişilmiştir.
- Baker, M. (2015). Over half of psychology studies fail reproducibility test. *Nature News*, 27. <https://doi.org/10.1038/nature.2015.18248> adresinden 28.3.2022 tarihinde erişilmiştir.
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604). <https://doi.org/10.1038/533452a> 05.5.2023 tarihinde erişilmiştir.
- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7, 543-554. <https://doi.org/10.1177/1745691612459060> adresinden 28.3.2022 tarihinde erişilmiştir.
- Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*, (53), 370-418. <https://doi.org/10.1098/rstl.1763.0053> adresinden 28.3.2022 tarihinde erişilmiştir.
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... ve Johnson, V. E. (2018). Redefine statistical significance. *Nature human behaviour*, 2(1), 6-10. <https://doi.org/10.1038/s41562-017-0189-z> adresinden 28.3.2022 tarihinde erişilmiştir.
- Benjamin, D. J., ve Berger, J. O. (2019). Three recommendations for improving the use of p-values. *The American Statistician*, 73(sup1), 186-191. <https://doi.org/10.1080/00031305.2018.1543135> adresinden 28.3.2022 tarihinde erişilmiştir.
- Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence. *Journal of the American statistical Association*, 82(397), 112-122. <https://doi.org/10.1080/01621459.1987.10478397> 05.5.2023 tarihinde erişilmiştir.
- Bertamini, M., & Munafò, M. R. (2012). Bite-size science and its undesired side effects. *Perspectives on Psychological Science*, 7(1), 67-71. <https://doi.org/10.1177/1745691611429353> adresinden 28.3.2022 tarihinde erişilmiştir.
- Bird A. (2018). Understanding the replication crisis as a base rate fallacy. *The British Journal for the Philosophy of Science*. Advance Online Publication. New York. <https://doi.org/10.1093/bjps/axy051> adresinden 28.3.2022 tarihinde erişilmiştir.
- Bryan, C. J., Tipton, E., ve Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature human behaviour*, 5(8), 980-989. <https://doi.org/10.1038/s41562-021-01143-3> adresinden 28.3.2022 tarihinde erişilmiştir.
- Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., ve Etchells, P. (2014). Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, 1(1), 4-17. <https://doi.org/10.3934/neuroscience.2014.1.4> adresinden 28.3.2022 tarihinde erişilmiştir.

- Coleman, M., ve Briggs, A. R. (Eds.). (2002). *Research methods in educational leadership and management*. Sage. New York. <http://dx.doi.org/10.4135/9781473957695> adresinden 25.3.2022 tarihinde erişilmiştir.
- Colling, L. J., ve Szűcs, D. (2021). Statistical inference and the replication crisis. *Review of Philosophy and Psychology*, 12(1), 121-147. <http://dx.doi.org/10.1007/s13164-018-0421-4> adresinden 25.3.2022 tarihinde erişilmiştir.
- Cronbach, L. J., ve Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281. <http://dx.doi.org/10.1037/h0040957> adresinden 25.3.2022 tarihinde erişilmiştir.
- Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge. New York. <http://dx.doi.org/10.4324/9780203807002> adresinden 25.3.2017 tarihinde erişilmiştir.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29. <http://dx.doi.org/10.1177/0956797613504966> adresinden 25.3.2022 tarihinde erişilmiştir.
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78-89. <https://doi.org/10.1016/j.jmp.2015.10.003> 05.5.2023 tarihinde erişilmiştir.
- Dienes, Z., ve Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic bulletin ve review*, 25(1), 207-218. <http://dx.doi.org/10.3758/s13423-017-1266-z> adresinden 25.3.2022 tarihinde erişilmiştir.
- Earp, B. D., ve Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in psychology*, 6, 621. <http://dx.doi.org/10.3389/fpsyg.2015.00621> adresinden 25.3.2022 tarihinde erişilmiştir.
- Ekici, O. (2009). İstatistikte Bayesyen ve Klasik yaklaşımın kavramsal farklılıkları. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 12(21), 89-101. Retrieved from <https://dergipark.org.tr/en/pub/baunsobed/issue/50238/647985> adresinden 25.3.2022 tarihinde erişilmiştir.
- Erkan, G. (2017). Klasik ve Bayesci Yapısal Eşitlik Modellerinde Parametre Tahminlerinin Karşılaştırılması: Sıralı Kategorik Verilerle Bir Uygulama. [Yüksek Lisans Tezi, Hacettepe Üniversitesi]. <https://tez.yok.gov.tr/UlusalTezMerkezi/> adresinden 05.5.2023 tarihinde erişilmiştir.
- Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., ve Nosek, B. A. (2014). Science forum: An open investigation of the reproducibility of cancer biology research. *Elife*, 3, e04333. <http://dx.doi.org/10.7554/elife.04333> adresinden 25.3.2022 tarihinde erişilmiştir.
- Etz A, Vandekerckhove J (2016) A Bayesian Perspective on the Reproducibility Project: Psychology. *PLoS ONE* 11(2): e0149794. <https://doi.org/10.1371/journal.pone.0149794> 05.5.2023 tarihinde erişilmiştir.
- Etz, A., ve Vandekerckhove, J. (2016). A Bayesian perspective on the reproducibility project: Psychology *PloS one*, 11(2), e0149794. <http://dx.doi.org/10.1371/journal.pone.0149794> adresinden 25.3.2022 tarihinde erişilmiştir.

- Friese, M., & Frankenbach, J. (2020). p-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*, 25(4), 456. DOI: [10.1037/met0000246](https://doi.org/10.1037/met0000246) 16.5.2022 tarihinde erişilmiştir.
- Gelman, A., ve Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, 9(6), 641-651. <http://dx.doi.org/10.1177/1745691614551642> 16.5.2022 tarihinde erişilmiştir.
- Gibson, E. W. (2021). The role of p-values in judging the strength of evidence and realistic replication expectations. *Statistics in Biopharmaceutical Research*, 13(1), 6-18. <http://dx.doi.org/10.1080/19466315.2020.1724560> 16.5.2022 tarihinde erişilmiştir.
- Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555(7695), 175-182. <https://doi.org/10.1038/nature25753> 16.5.2022 tarihinde erişilmiştir.
- Hanel, P. H., ve Mehler, D. M. (2019). Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication. *Public understanding of science*, 28(4), 468-485. <http://dx.doi.org/10.1177/0963662519834193> 16.5.2022 tarihinde erişilmiştir.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.1004085> 05.5.2023 tarihinde erişilmiştir.
- Jackman, S. (2009). *Bayesian analysis for the social sciences* (Vol. 846). John Wiley ve Sons. <http://dx.doi.org/10.1002/9780470686621> 15.6.2022 tarihinde erişilmiştir.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*, 23(5), 524-532. DOI: 10.1177/0956797611430953 1.11.2022 tarihinde erişilmiştir.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524-532. <https://doi.org/10.1177/0956797611430953> 22.7.2022 tarihinde erişilmiştir.
- Kara, E., Türküm, A. S., & Turner, M. J. (2023). The Effects of Rational Emotive Behaviour Therapy (REBT) Group Counselling Program on Competitive Anxiety of Student-Athletes. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, 1-18. <https://doi.org/10.1007/s10942-023-00497-z> 05.5.2023 tarihinde erişilmiştir.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and social psychology review*, 2(3), 196-217. <https://doi.org/10.1207/s15327957pspr0203> 11.5.2022 tarihinde erişilmiştir.
- Koptur, Murat (2020). Bayesci İstatistiğe Giriş. Preprint. DOI: 10.13140/RG.2.2.30359.93608/
- König, C., ve van de Schoot, R. (2018). Bayesian statistics in educational research: a look at the current state of affairs. *Educational Review*, 70(4), 486-509. <http://dx.doi.org/10.1080/00131911.2017.1350636> 18.1.2022 tarihinde erişilmiştir.
- Kruschke, J. K., Aguinis, H., ve Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, 15(4), 722-752. <http://dx.doi.org/10.1177/1094428112457829> 1.7.2022 tarihinde erişilmiştir.
- Kruschke, J. K., ve Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation,

- meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic bulletin ve review*, 25(1), 178-206. <http://dx.doi.org/10.3758/s13423-016-1221-4> 19.6.2022 tarihinde erişilmiştir.
- Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., ... ve Zwaan, R. A. (2018). Justify your alpha. *Nature human behaviour*, 2(3), 168-171. <https://doi.org/10.1038/s41562-018-0311-x> 14.7.2022 tarihinde erişilmiştir.
- Lee, S. Y., ve Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, 39(4), 653-686. http://dx.doi.org/10.1207/s15327906mbr3904_4 21.3.2022 tarihinde erişilmiştir.
- Makowski, D., Ben-Shachar, M. S., Chen, S. H., ve Lüdtke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in psychology*, 10, 2767. <http://dx.doi.org/10.3389/fpsyg.2019.02767> 15.3.2022 tarihinde erişilmiştir.
- Matzke, D., Nieuwenhuis, S., van Rijn, H., Slagter, H. A., van der Molen, M. W., & Wagenmakers, E.-J. (2015). The effect of horizontal eye movements on free recall: A preregistered adversarial collaboration. *Journal of Experimental Psychology: General*, 144(1), e1–e15. <https://doi.org/10.1037/xge0000038> 05.5.2023 tarihinde erişilmiştir.
- Maxwell, S. E., Lau, M. Y., ve Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70(6), 487. <http://dx.doi.org/10.1037/a0039400> 18.7.2022 tarihinde erişilmiştir.
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., ve Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic bulletin ve review*, 23(1), 103-123. <http://dx.doi.org/10.3758/s13423-015-0947-8> 15.2.2022 tarihinde erişilmiştir.
- Morey, R. D., Rouder, J. N., Verhagen, J., ve Wagenmakers, E. J. (2014). Why hypothesis tests are essential for psychological science: A comment on Cumming. *Psychological science*, 25(6), 1289-1290. <http://dx.doi.org/10.1037/a0039400> 13.1.2022 tarihinde erişilmiştir.
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, 69, 511-534. <https://doi.org/10.1146/annurev-psych-122216-011836> 14.1.2022 tarihinde erişilmiştir.
- Nosek, B. A., ve Errington, T. M. (2017). Reproducibility in cancer biology: Making sense of replications. *Elife*, 6, e23383. <http://dx.doi.org/10.7554/elife.23383> 17.1.2022 tarihinde erişilmiştir.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <http://dx.doi.org/10.1126/science.aac4716> 18.1.2022 tarihinde erişilmiştir.
- Pashler, H., ve Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on psychological science*, 7(6), 528-530. <https://doi.org/10.1177/1745691612465> 18.1.2022 tarihinde erişilmiştir.
- Rietbergen, C., Debray, T. P., Klugkist, I., Janssen, K. J., ve Moons, K. G. (2017). Reporting of Bayesian analysis in epidemiologic research should become more transparent. *Journal of Clinical Epidemiology*, 86, 51-58. <https://doi.org/10.1016/j.jclinepi.2017.04.008> 18.1.2022 tarihinde erişilmiştir.

- Rogers, A. (2018). "The Science Behind Social Science Gets Shaken Up—Again". *Wired*. 2022-08-25. <https://www.wired.com/story/social-science-reproducibility/> 6.5.2022 tarihinde erişilmiştir.
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic bulletin & review*, 21, 301-308. <https://doi.org/10.3758/s13423-014-0595-4> 05.5.2023 tarihinde erişilmiştir.
- Rupp, A. A., Dey, D. K., ve Zumbo, B. D. (2004). To Bayes or not to Bayes, from whether to when: Applications of Bayesian methodology to modeling. *Structural Equation Modeling*, 11(3), 424-451. https://doi.org/10.1207/s15328007sem1103_7 21.5.2022 tarihinde erişilmiştir.
- Schimmack, U. (2010). What multi-method data tell us about construct validity. *European Journal of Personality: Published for the European Association of Personality Psychology*, 24(3), 241-257. <https://doi.org/10.1002/per.771> 21.7.2022 tarihinde erişilmiştir.
- Schimmack, U. (2021). The validation crisis in psychology. *Meta-Psychology*, 5. <https://doi.org/10.15626/MP.2019.1645> 17.11.2022 tarihinde erişilmiştir.
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2), 322–339. <https://doi.org/10.1037/met0000061> 05.5.2023 tarihinde erişilmiştir.
- Sharpe, D., ve Poets, S. (2020). Meta-analysis as a response to the replication crisis. *Canadian Psychology/Psychologie canadienne*, 61(4), 377. <https://doi.org/10.1037/cap0000215> 13.4.2022 tarihinde erişilmiştir.
- Shrout, P. E., ve Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual review of psychology*, 69(1), 487-510. <https://doi.org/10.1146/annurev-psych-122216-011845> 19.7.2022 tarihinde erişilmiştir.
- Siddaway, A. P., Wood, A. M., ve Hedges, L. V. (2019). How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual review of psychology*, 70, 747-770. <https://doi.org/10.1146/annurev-psych-010418-102803> 19.6.2022 tarihinde erişilmiştir.
- Siegfried, T. (2010). Odds are, it's wrong. *Science news*, 177(7), 26. <http://www.aakkozzll.com/pdf/siegfried2.pdf> 15.5.2022 tarihinde erişilmiştir.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632> 11.7.2022 tarihinde erişilmiştir.
- Şehribanoğlu, S. (2012). Yapısal eşitlik modellerinde parametre tahmininde gibbs örnekleme. Doktora Tezi, *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü*, Van, 140. <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp> 21.9.2022 tarihinde erişilmiştir.
- Van De Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., ve Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2), 217. <https://doi.org/10.1037/met0000100> 10.10.2022 tarihinde erişilmiştir.
- Veen, D., ve Egberts, M. (2020). *The Importance of Collaboration in Bayesian Analyses with Small Samples. In Small Sample Size Solutions* (pp. 50-70). Routledge. <https://doi.org/10.4324/9780429273872-5> 1.9.2022 tarihinde erişilmiştir.

- Wagenmakers, E. J., Lee, M., Lodewyckx, T., ve Iverson, G. J. (2008). *Bayesian versus frequentist inference*. In *Bayesian evaluation of informative hypotheses* (pp. 181-207). Springer, New York, NY. <https://doi.org/10.4324/9780429273872-5> 6.5.2022 tarihinde erişilmiştir.
- Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... ve Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic bulletin ve review*, 25(1), 58-76. <https://doi.org/10.3758/s13423-017-1323-7> 4.2.2022 tarihinde erişilmiştir.
- Wasserstein, R. L., Schirm, A. L., ve Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, 73(sup1), 1-19. https://doi.org/10.1007/978-0-387-09612-4_9 2.3.2022 tarihinde erişilmiştir.
- Williams, R. T., Polanin, J. R., ve Pigott, T. D. (2017). *Meta-analysis and reproducibility*. In. M. C. Makel ve J. A. Plucker (Eds.), *Toward a more perfect psychology* (pp. 255–270). American Psychological Association. <https://doi.org/10.1037/0000033-016> 1.8.2022 tarihinde erişilmiştir.
- Witkowski T (2019). "Is the glass half empty or half full? Latest results in the replication crisis in Psychology" (PDF). *Skeptical Inquirer*. Vol. 43, no. 2. pp. 5–6. <https://skepticalinquirer.org/2019/03/is-the-glass-half-empty-or-half-full-latest-results-in-the-replication-crisis-in-psychology/> 11.2.2022 tarihinde erişilmiştir.
- Yang, Y., Youyou, W., & Uzzi, B. (2020). Estimating the deep replicability of scientific findings using human and artificial intelligence. *Proceedings of the National Academy of Sciences*, 117(20), 10762-10768. <https://doi.org/10.1073/pnas.1909046117> 05.5.2023 tarihinde erişilmiştir.
- Yardımcı, A., & Erar, A. (2005). Gibbs sampling approach to variable selection in linear regression with outlier values. *GU Journal of Science*, 18(4), 603-611.

Extended Abstract

Introduction

The reproducibility and P-value crises that started in social sciences, especially psychology, have left their mark in the last 10 years. In addition to all these, discussions of theory and validity crisis, which started from psychology but spread to social sciences, started. All these debates have imposed compulsory reform in the field of methods, applied statistical analysis, sampling and research practices in all fields of science in general and in social sciences in particular.

Method

Within the scope of this study, the causes of these crises and the ways to get out of the crisis were discussed and suggestions were made. In this context, two historical transitions began to occur in data analysis in this short period of time. First, the use of effect size, confidence interval and meta-analysis in uncertainty estimation in classical hypothesis testing, this trend was named as “the New Statistics”. The second is the transition from classical statistics to Bayesian statistics.

While some researchers state that the crisis has been exaggerated (Gibson, 2021), most scientists have admitted that a crisis exists and blame poor methodological and statistical practices. All these discussions necessitated reform in the fields of method, applied statistical analysis, sampling and research practices in all fields of science in general and social sciences in particular. Two historical trends began to occur in data analysis in this short period of time. The first of these is “The New Statistics” (Cumming, 2014), which prioritizes the use of effect size, confidence interval and meta-analysis in estimating uncertainty in classical hypothesis tests. The other is the approach that advocates the transition from classical or frequentist statistics to Bayesian statistics (Kruschke ve Liddell, 2015).

Findings

New statistics

New statistics instead of hypothesis tests, H0 hypothesis (NHST); using effect size, confidence interval and meta-analysis in uncertainty estimation instead of significance test recommended (Cumming, 2014). Accordingly, the H0 hypothesis finds the significance test insufficient as it brings the understanding that everything about the current effect is either black or white. P value is very limited for interpreting and analyzing data. According to Cumming, effect size and confidence intervals provide much more comprehensive information than the H0 hypothesis significance test. In addition, meta-analysis studies allow the collection and interpretation of evidence from many studies. Using new statistical techniques will contribute

to the rapid progress of science by providing more information to researchers (Cumming, 2013).

Bayesian Statistics

With the increasing technological and methodological developments, more complex models are created for the solution of statistical problems and it is tried to reach the result that best reflects the model of the data. In the classical approach, the p-value is considered as the main decision maker while making decisions about the models. Due to the misuse of the P value and the difficulties it brings in interpretation, it has been used less in the analyzes in recent years (Benjamin et al., 2018). As an alternative to classical statistics in statistical parameter estimation, Bayesian statistics is widely used in model selection and hypothesis testing. The Bayesian approach is based on a very important concept such as probability in the scientific decision making process. The approach, which emerged from the concept of subjective probability put forward by Thomas Bayes in 1783, was named Bayesian approach because it was based on Bayes' theorem. Since the posterior distributions obtained by using the prior distribution are used in the application of the Bayesian approach, difficulties have been experienced in the application phase for many years. It takes a long time to obtain posterior distributions, and their general use has not become widespread due to the use of high-dimensional integrals. However, thanks to Markov Chain Monte Carlo methods (MCMC), Metropolis-Hasting algorithm, Gibbs sampling algorithm and developing computational technologies, this approach has been used primarily in physics and space sciences (König ve van de Schoot, 2017).

The Bayesian approach has many advantages over classical statistics. It is widely recommended in the literature to use Bayesian statistics instead of classical statistics in small samples (Wagenmakers, et al., 2008). Maximum Likelihood estimation method is used in parameter estimation in classical statistics and asymptotically normal distribution assumption is required. However, it is very difficult to achieve multivariate normality in fields such as medicine and psychology, where it is difficult to create a large sample, or in areas where categorical and missing data are available, such as social sciences. This may lead to biased results in parameter estimations (Rupp et al., 2004). The Bayesian approach brings a great innovation and benefit to the statistical dimension of the reproducibility crisis, especially with the effect of obtaining posterior distributions using a priori information. The Bayesian approach offers many advantages over classical statistics outside of the small sample context (Rupp et al., 2004):

- The Bayesian approach offers preference to choose a model in the light of the available data and other information, instead of telling which model is correct as in classical statistics. In

this way, it allows the selection of the best models for future predictions (Kruschke, Aguinis ve Joo, 2012).

- Bayesian hypothesis tests are not based on the null hypothesis as in the classical one. The range of testable hypotheses is wide and intuitive interpretations that are not based on the significance test can be made by providing information about the alternative hypothesis (Kruschke, Aguinis ve Joo, 2012).
- It provides more reliable results by providing useful statistics of the posterior distribution such as mean and percentile for the data, as it allows the use of real prior knowledge in addition to the existing data (Lee ve Song, 2004).
- Since the Bayesian approach requires less computation, more and new models can be easily analyzed (Lee ve Song, 2004).
- Due to the sequential nature of Bayesian estimation, uncertainties on parameters are reduced by inference (Makowski ve diğerleri,, 2019).
- In the Bayesian approach, many problems that do not have a solution in the classical approach are easily solved (eg. Behrens-Fisher problem) (Dienes ve Mclatchie, 2018).
- Bayesian approach is more reliable, gives more accurate results in mixed and dirty data, type 1 error proneness is very low, and gives more accurate and reliable results with the addition of a priori information (Andrews ve Baguley, 2013; Etz ve Vandekerckhove, 2016; Kruschke, Aguinis ve Joo, 2012; Wagenmakers et al., 2018).

Conclusion

As a result, preferring open science, being transparent in research, utilizing the opportunities offered by new methods and statistics will contribute to the exit from this crisis and the production of healthier scientific knowledge. In social sciences, where it is difficult to create large samples, categorical and missing data are common, and difficult to meet assumption of multivariate normality. Bayesian approach, especially with the effect of obtaining posterior distributions by using a priori information and giving more accurate results in small samples, It will bring great innovation and benefit to the statistical dimension of the crisis.

Statement of Conflict of Interest

“Due to the nature of the study, informed consent or ethics committee approval was not required.”