

EDİTÖR: **PROF. DR. ÖZLEM ALPU**

İstatistikte

Güncel Konular -II

EĞİTİM
yayınevi

İstatistikte Güncel Konular -II

Editör: Prof. Dr. Özlem Alpu

Yayınevi Grubu Genel Başkanı: Yusuf Ziya Aydoğan (yza@egitimyayinevi.com)

Genel Yayın Yönetmeni: Yusuf Yavuz (yusufyavuz@egitimyayinevi.com)

Sayfa Tasarımı: Kübra Konca Nam

Kapak Tasarımı: Eğitim Yayınevi Grafik Birimi

T.C. Kültür ve Turizm Bakanlığı

Yayıncı Sertifika No: 76780

E-ISBN: 978-625-385-183-5

1. Baskı, Nisan 2025

Kütüphane Kimlik Kartı

İstatistikte Güncel Konular -II

Editör: Prof. Dr. Özlem Alpu

IV+95 s., 160x240 mm

Kaynakça var, izin yok.

E-ISBN: 978-625-385-183-5

Copyright © Bu kitabın Türkiye'deki her türlü yayın hakkı Eğitim Yayınevi'ne aittir. Bütün hakları saklıdır. Kitabın tamamı veya bir kısmı 5846 sayılı yasanın hükümlerine göre kitabı yayımlayan firmanın ve yazarlarının önceden izni olmadan elektronik/mechanik yolla, fotokopi yoluyla ya da herhangi bir kayıt sistemi ile çoğaltılamaz, yayımlanamaz.

EĞİTİM
yayınevi

Yayınevi Türkiye Ofis: İstanbul: Eğitim Yayınevi Tic. Ltd. Şti., Atakent mah. Yasemen sok. No: 4/B, Ümraniye, İstanbul, Türkiye

Konya: Eğitim Yayınevi Tic. Ltd. Şti., Fevzi Çakmak Mah. 10721 Sok. B Blok, No: 16/B, Safakent, Karatay, Konya, Türkiye
+90 332 351 92 85, +90 533 151 50 42
bilgi@egitimyayinevi.com

Yayınevi Amerika Ofis: New York: Egitim Publishing Group, Inc.
P.O. Box 768/Armonk, New York, 10504-0768, United States of America
americaoffice@egitimyayinevi.com

Lojistik ve Sevkiyat Merkezi: Kitapmatik Lojistik ve Sevkiyat Merkezi, Fevzi Çakmak Mah. 10721 Sok. B Blok, No: 16/B, Safakent, Karatay, Konya, Türkiye
sevkiyat@egitimyayinevi.com

Kitabevi Şubesi: Eğitim Kitabevi, Şükran mah. Rampalı 121, Meram, Konya, Türkiye
+90 332 499 90 00
bilgi@egitimkitabevi.com

İnternet Satış: www.kitapmatik.com.tr
bilgi@kitapmatik.com.tr

EĞİTİM YAYINEVİ
GRUBU

EĞİTİM
yayınevi

SALON
yayıncılık

kitapmatik
eğitim yayıncılık

Kitapmatik
eğitim yayıncılık

EĞİTİM
KITABEVİ

İÇİNDEKİLER

TEK DENEKLİ DENEYSEL TASARIM	1
Nurhan Doğan, İsmet Doğan	
BİR YÖNLÜ VARYANS ANALİZİNE ALTERNATİF OLAN PARAMETRİK OLMAYAN k-ÖRNEKLEM TESTLERİ.....	17
Aslı Ceren Macunluoğlu, Gökhan Ocakoğlu	
HASSAS SORU YAKLAŞIMI.....	28
Robab Tetik, İlker Ercan	
DELPHİ YÖNTEMİ	49
İsmet Doğan, Nurhan Doğan	
İSTATİSTİKSEL PROSES KONTROL GRAFİKLERİ.....	69
Özlem Toluk, İlker Ercan	

ÖNSÖZ

Bu kitap, İstatistik alanında seçilmiş güncel konuları ele alan ikinci kitabımız olup, beş bölümden oluşmaktadır. Kitabın birinci bölümünde Nurhan Doğan ve İsmet Doğan sınırlamaların var olduğu deneysel çalışmalarda bir alternatif veya tamamlayıcı rol üstlenen tek denekli deneysel tasarıma odaklanmıştır. Aslı Ceren Macunluoğlu ve Gökhan Ocakoğlu parametrik testlerin kullanımı için gerekli varsayımlar ihlal edildiğinde, bu testlere alternatif olarak literatürde önerilen parametrik olmayan testlerden Kruskal-Wallis testi, Kruskal-Wallis testinin permütasyon testi temelli modifiye versiyonu, Mood'un Medyan testi, Van der Waerden testi ve Savage testini incelemiştir. Robab Tetik ve İlker Ercan hassas soru çalışmalarında ölçüm hatalarını azaltmayı ve insanların yanıtlarının güvenilirliğini artırmayı amaçlayan farklı stratejileri bir araya toplayıp, başlıca hassas soru yöntemleri olarak, randomize cevap yöntemi, aday yöntemi, gruplandırılmış cevap yöntemi, fasulye yöntemi, eşsiz-sayım yöntemi ve çapraz modelini ayrıntılı ele almışlardır. İsmet Doğan ve Nurhan Doğan doğrulanmış verilerin yokluğunda grup tahmininin mantıklı olduğu pek çok disiplinde karmaşık bir sorunu bir grup bireyin kolektif olarak ele almalarını sağlamak için kullanılan yöntem olarak karşımıza çıkan Delphi yönteminin tarihçesi, özellikleri, avantaj/dezavantajları, türleri ve yöntem içinde yer verilen kavramlarla sağlık araştırmalarında yöntemin kullanımı üzerinde durmuşlardır. Kitabın son bölümünde Özlem Toluk ve İlker Ercan istatistiksel süreç kontrol (İPK) grafiklerini kapsamlı bir şekilde ele almaktadır. İPK'nın sağlık hizmetlerindeki ve diğer alanlardaki uygulama alanlarını açıklayan bölüm, çeşitli kontrol grafik türlerini, hesaplama yöntemlerini ve kullanım amaçlarını detaylandırmaktadır. Nicel ve nitel veriler için farklı kontrol grafiklerinin yanı sıra, küçük değişimleri tespit etmede etkili olan CUSUM ve EWMA gibi gelişmiş tekniklere de değinilmektedir. Yazarlar, İPK'nın süreç performansını izleme, kontrol altında tutma ve iyileştirme amacını vurgulamaktadır.

Kitabın akademisyenlere, araştırmacılara ve tüm okuyuculara faydalı olmasını ve çalışmalarına katkı sağlamasını diliyorum.

Çalışmanın gerçekleşmesinde emeği geçen tüm bölüm yazarlarına teşekkür ederim.

Prof. Dr. Özlem ALPU

TEK DENEKLİ DENEYSEL TASARIM

Nurhan Doğan**, İsmet Doğan***

Giriş

Araştırmacılar, bir deney tasarlarlarken, genellikle tasarımı bir grup yaklaşımı çerçevesinde ele alır ve bu süreçte gruplar arası ya da grup içi karşılaştırmaları neredeyse otomatik olarak dâhil ederler. Bu şaşırtıcı değildir, çünkü bilimsel çalışmalar çoğunlukla çok sayıda katılımcıyla yapılan deneylere dayanmaktadır. Klinik müdahalelerin etkilerini test eden çoğu araştırmacı tarafından, müdahalelerinin etkililiği ve altta yatan mekanizmalar hakkındaki soruları yanıtlamak için rastgele kontrollü deneme tasarımları kullanılmaktadır. Rastgele kontrollü denemelerde, katılımcılar ya araştırma altındaki aktif bir müdahaleye ya da alternatif bir müdahale koşuluna rastgele dağıtılır ve belirli bir yapı üzerindeki ortalama etkiler koşullar arasında karşılaştırılır. Deneysel karakterleri ve büyük uygulanabilirlikleri nedeniyle, rastgele kontrollü denemeler, müdahale araştırması için altın standart olarak görülmektedir. Bu yaklaşım çok kullanışlı olsa da, grup tasarımlarında da bazı sınırlamalar söz konusudur. Grup verilerinden elde edilen sonuçlar önemli olmasına rağmen araştırmacılar bazen, bireysel olarak katılımcılar üzerindeki müdahalelerin etkililiğini daha büyük bir özgüllükle anlamaya çalışırlar ki grup çalışmaları tipik olarak buna uygun değildir. Sınırlamaların var olduğu durumlarda, tek denekli deneysel tasarım (TDDT)lar, grup tasarımlarına geçerli bir alternatif veya tamamlayıcı olabilir. Örneğin, geniş çaplı bir grup çalışmasının başlangıç aşamalarında pilot verilerin elde edilmesi için TDDT yöntemi kullanılabilir. Yinelemeli, müdahale ve teori odaklı tasarım çalışmaları, öğretim ve öğrenmeye yönelik etkili yaklaşımların geliştirilmesine önemli katkılarda bulunmakta, araştırmacıların yalnızca neyin işe yaradığına değil, "nasıl, ne zaman ve neden işe yaradığına" da bakmalarına olanak sağlamaktadır. TDDT çalışmaları, araştırmacıların bu karmaşık öğrenme ekolojilerini (bir müdahale veya yaklaşımın nasıl, ne zaman ve neden işe yaradığını incelemek, katılımcılarla ilişkili bireysel farklılıkları yakından

** Prof. Dr., Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Afyonkarahisar, Türkiye, nurhandogan@hotmail.com, ORCID: orcid.org/0000-0001-7224-6091

*** Prof. Dr., Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Afyonkarahisar, Türkiye, ismet.dogan@afsu.edu.tr, ORCID: 0000-0001-9251-3564

araştırmak vb.) ele almasına açık ve güçlü bir şekilde izin verir. Bunlara ek olarak, iyi yapıldığında yüksek iç geçerliliğe ve dış geçerliliğe sahip olan deneysel çalışmalardır (Plavnick ve Ferreri, 2013). Literatürde TDDT, tek durumlu deney tasarımları, tek denekli araştırma tasarımı, küçük N tasarımları, çoklu durum tasarımı, tek durum tasarımı, tek sistemli tasarımlar vb. farklı terimler ile ifade edilmektedir. Tek vaka, $N = 1$, N-of-1 ve tek hasta denemeleri, bir araştırmaya yalnızca bir deneğin dâhil edildiğini ima eder. TDDT'lar için bazen "denek içi tasarım" veya "denek içi tekrarlı tasarım" terimleri de kullanılmaktadır. Bu terimler, tasarımın genellikle aynı bireyin veya bireylerin zaman içindeki performansına odaklandığını doğru bir şekilde ifade etmektedir. Ancak, terimler tam olarak kesin değildir, çünkü bazı tasarımlar müdahalelerin bireyler üzerindeki veya bireyler arası etkilerini incelemeye dayanır. Genel olarak, TDDT'ler yalnızca bireyleri değil, büyük grupları da değerlendirmeye olanak tanıyan çeşitli deneysel düzenlemeler sunar. Terminoloji ne olursa olsun, tasarım çerçevesi temelde aynıdır: zaman içinde tek bir kişiyi veya küçük bir grup insanı ileriye dönük ve yoğun bir şekilde incelemek; çalışmanın tüm aşamalarında sonucu tekrar tekrar ve sıklıkla ölçmek ve müdahaleyi sırayla uygulamak veya geri çekmek (Kazdin, 2021). TDDT dört ana nedenden dolayı avantajlıdır.

1. Araştırmacılar, zamanla farklı perspektiflerden karmaşık organizasyonel olguları inceleyerek derin bir kavrayış geliştirebilirler.
2. Araştırmacıların dışarıdan kolayca gözlemlenemeyen bir olguya benzersiz bir erişim avantajı elde etmelerini sağlar.
3. Vaka, birden fazla vakanın mevcut olmayabileceği ve bir vakanın incelenmesinin yeni bir teori üretmek için yeterli olduğu nadir bir olgunun veya sürecin somut örneği olabilir. Bu yüzden, tek vaka araştırmasının mantığı aşağıdaki üç koşuldan en az birini karşılamalıdır:
 - Vakanın sıra dışı bir olgu olması,
 - Daha önce araştırmacıların erişimine açık olmamış olması,
 - Vakanın zaman içinde (boylamsal olarak) gözlemlenebilmesi.
4. Birden fazla vakayla veya büyük örneklemli istatistiksel çalışmalar gibi diğer yöntemlerle elde edilemeyen bir olgunun ayrıntı düzeyinde incelenmesidir (Özcan, Han ve Graebner, 2017).

TDDT'ler, bağımsız değişkenin araştırmacı tarafından değiştirildiği ve bağımlı değişkenin tekrar tekrar ölçüldüğü deneysel tasarımlar grubuna dâhildir. Tek vaka deneysel desenlerinde vaka, aynı zamanda müdahale ve analiz birimi olarak ele alınır. Her vakada yapılan ölçümler, bağımsız değişkenin uygulanmasından önce ve sonra gerçekleştirildiği için vaka genellikle bir kontrol unsuru olarak işlev görür. Tek durum tasarımlarının deneysel varyantları,

- Bağımsız değişkenin tanıtılması ve geri çekilmesi,
- Bağımsız değişkenin farklı aşamalar boyunca manipüle edilmesi,
- Bağımsız değişkenin tanıtılması yoluyla müdahalenin farklı zaman noktalarında kademeli bir şekilde tekrarlanması,

sebeplerinden dolayı nedensel bir ilişkinin belirlenmesi için bir temel sağlar (Garbacz ve Kratochwill, 2020). TDDT'ların ortak sınırlamaları vardır. Bir ölçüm, müdahale olmaksızın tekrarlanan testlerle değişirse, bu tür tepkiselliği azaltmak için daha göze çarpmayan izleme gibi adımlar atılmadığı sürece, TDDT için yararlı olmayabilir. Müdahalelerin etkilerinin zaman içinde değerlendirildiği göz önüne alındığında, sistematik çevresel değişiklikler veya olgunlaşma, tedavi ile sonuç arasındaki ilişkiyi etkileyebilir ve dolayısıyla tedavinin etkisini gizleyebilir. Bununla birlikte, tersine çevirme ve çoklu temel tasarımlardan oluşan tasarım mantığı, bu tür etkileri büyük ölçüde kontrol etmektedir. TDDT'lar, tekrarlanan ölçümlere ve tedavi ile sonuç arasındaki ilişkinin ayrıntılı bir çalışmasına dayandığından, sıklıkla örneklenemeyen bağımlı ölçümler kullanan çalışmalar, TDDT'lara aday değildir. Benzer şekilde, tedavinin başlatılması ile sonuçta değişiklik başlatılması arasında geçici bir ilişkinin belirlenmemesi, değişikliklerin müdahaleye atfedilmesini zorlaştırabilir. Müdahalenin etkilerine ilişkin uygunsuz kararlara neden olabilecek, müdahalenin getirilmesi veya kaldırılmasıyla karıştırıcı bir değişkenin ilişkilendirilmesi her zaman mümkündür. Bireylerin başına gelen tedaviyi bırakma veya kontrolsüz olaylar, TDDT'a kafa karıştırıcı değişkenler katabilir. Bu problemler TDDT'lara özgü değildir, aynı zamanda randomize kontrollü çalışmalarda da ortaya çıkmaktadır (Epstein ve Dallery, 2022). Müdahalelerin etkinliğini incelemek için kullanılabilecek TDDT'lar;

- AB tasarımları,
- Ters tasarımlar,

- Çoklu temel tasarımlar,
- Alternatif tedavi tasarımları,

olarak isimlendirilmektedir (Lobo ve ark., 2017). Tipik geri çekme tasarımı, 'B' harfiyle tedavi aşamasını ifade eden bir ABAB modelidir. Geri çekme, müdahale ile hedef davranış arasında işlevsel bir bağlantı olup olmadığını belirlemek amacıyla, çalışmanın bir veya daha fazla aşamasında müdahalenin geçici olarak kaldırılmasını içerir. Bu ifadeden de anlaşılacağı üzere tipik olarak, bir müdahalenin müdahale etkileri, müdahale koşulunun (B fazı) hiçbir müdahalenin yürürlükte olmadığı durumdaki başlangıç durumu (A fazı) değiştirilerek değerlendirilir. A ve B fazları, dört fazı tamamlamak için tekrarlanır. Çoklu temel tasarımında, araştırmacılar, katılımcılar genelinde bir müdahale koşulu ile bir temel durumu birleştirir. Çoklu temel desenler, tek denekli araştırmalarda en yaygın kullanılan tasarımlardan biridir. Bu desenlerde veriler, üç veya daha fazla davranış, üç veya daha fazla kişinin davranışları ya da üç veya daha fazla farklı durumda gerçekleşen davranışlar üzerinden ayrı ayrı toplanır. Müdahalenin etkileri, farklı zaman noktalarında farklı başlangıç noktalarına uygulanarak değerlendirilir. Alternatif tedavi tasarımı ise, çeşitli müdahale yöntemleri arasından en etkili olanı belirlemek isteyen araştırmacılar için oldukça faydalıdır. Bu tasarımın en büyük avantajlarından biri, araştırmacının müdahaleyi geri çekmek zorunda olmamasıdır. Ayrıca, her aşamanın süresi oldukça kısa olabilir, bu da araştırmacılara değişkenler arasındaki karşılaştırmaları daha hızlı yapma imkânı sunar. Bunun yanı sıra, AB tasarımlarından farklı olarak, temel veri toplama zorunluluğu yoktur. Eğer araştırmacı temel verileri toplarsa, müdahale uygulanmadan önce bu verilerin sabit olması gerekmez. Ancak, bu araştırma tasarımının bazı sınırlamaları da vardır. En önemli sınırlamalardan biri, önceki müdahale aşamalarından elde edilen sonuçların sonraki aşamalarındaki davranışı etkileyebileceği anlamına gelen çoklu tedavi müdahalesi veya aktarma etkisidir (Çakıroğlu, 2012). Görsel inceleme yoluyla deneysel etkiyi gösteren alternatif tedavi tasarımında faz yoktur. Bununla birlikte, ilgili her bir tedavi koşulundaki veriler ayrı ayrı incelenirse, her bir koşulun görece düzeyi ve eğilimi iki veri serisi arasında karşılaştırılabilir. Her tasarımın avantajları ve sınırlamaları vardır ve araştırmacılar, bağımlı ve bağımsız değişkenler arasında işlevsel bir ilişki gösterme olasılığı en yüksek olan tasarımı seçmelidir. TDDT terimi, az sayıda hasta üzerinde bir müdahalenin etkinliğini test etmek için kullanılabilen çeşitli deneysel yöntemleri ifade eder. TDDT'ler, yalnızca vaka raporlarından ibaret değildir;

aksine, müdahale başlamadan önce titizlikle tasarlanmış çalışmalardır ve bu nedenle gerçek anlamda "deneysel" bir tasarım niteliği taşırlar. TDDT metodolojisi, bir müdahalenin etkinliğini değerlendirmeyi veya iki ya da daha fazla müdahalenin göreceli etkinliğini karşılaştırmayı hedefler. Genel olarak, TDDT'ler daha büyük ölçekli deneylerin pratik ve tamamlayıcı bir parçası olarak kullanılmaktadır. Araştırmacının erken gelişim aşamasında olduğu kadar araştırma bulgularının bireysel hastalara uygulanmasında da faydalı olabilirler. Özellikle TDDT'lara uygun durumlar:

- Klinik izlenimlerden ziyade kanıtlara dayalı en iyi tedaviyi sağlamak için günlük klinik uygulamada belirli bir hasta için mevcut bir müdahalenin etkinliğini değerlendirmek,
- Tek veya birkaç hasta ile klinik rehabilitasyon ortamında araştırma yapmak,
- Yeni bir müdahalenin pilot uygulaması veya bilinen bir müdahalenin atipik bir vakaya veya müdahalenin orijinal olarak tasarlandığı başka bir duruma/tipine uygulanması/modifiye edilmesi,
- Müdahale paketinin hangi bölümünün etkili olduğunu araştırmak,
- Bir grup çalışması için hiçbir zaman yeterli hasta bulunamayacak nadir koşullar veya olağandışı müdahale hedefleriyle çalışmak,
- Bir grup çalışması için homojen bir hasta örneği elde etmenin imkânsızlığı,
- Zaman veya finansman sınırlılığı,

olarak, TDDT tasarlama ve uygulama adımları ise;

- Deney boyunca birçok kez tekrarlanabilen, uygulanması nispeten kısa olan ve müdahale hedefini yansıtan bir sonuç ölçüsü bulun.
- Kullanılacak diğer sonuç ölçütlerini seçin (kontrol ölçütleri, genelleştirme ölçütleri, uygulama ölçütleri, diğer standart ölçütler).
- Bir tasarım seçin.
- Tasarım ayrıntılarını ve/veya rastgeleleştirmeyi planlayın.
- Müdahale ayrıntılarını uyarlayın/açıklayın ve bir usule uygunluk ölçüsü seçin.
- Protokolü standartlara göre gözden geçirin.

- Beklenmedik olayları izleyerek, orijinal tasarımdan ve/veya müdahale içeriğinden herhangi bir sapma veya uyarılama hakkında bilgi toplayın. Prosedürel uygunluğu ve sonuç ölçüsünün değerlendiriciler arası güvenilirliğini değerlendirerek deneyi başlatın.
- Verilerin görsel olarak grafiğini çizin ve analiz edin.
- Görsel analizi doğrulamak için istatistikleri kullanın,

olarak belirtilmektedir. Bir TDDT'nin prensip olarak bir grup tasarımına uygun ve tercih edilir olduğuna karar verdikten sonra yanıtlanması gereken soru, bu tasarımın uygulanabilirliğidir. TDDT kullanımını engelleyebilecek ana neden, tekrar tekrar ölçülebilen geçerli ve güvenilir bir sonuç ölçüsü seçmenin zorluğudur (Krasny-Pacini ve Evans, 2018). TDDT çalışmaları, klinik ve eğitim ortamlarındaki müdahalelerin etkinliğini incelemek için onlarca yıldır kullanılmaktadır. Ancak TDDT metodolojisinin bazı yönleri için bir fikir birliği mevcutken, diğerleri için devam eden bir tartışma söz konusudur. En çok tartışılan konulardan biri, TDDT verilerinin nasıl analiz edilmesi gerektiğidir. TDDT verileri genellikle görsel olarak yorumlanmıştır. Görsel analiz, verileri analiz etmenin geleneksel yöntemidir ve birçok alanda TDDT'da bir etkinin olup olmadığını belirlemek için birincil yöntem olarak yaygın şekilde kabul görmektedir. Grafik gösterim ve görsel analizin avantajları;

- Gözlem periyodunun hemen ardından her bir davranış ölçüsünü bir diyagram üzerinde çizmek, bir araştırmacının bir katılımcının performansını sürekli olarak görsel bir kayıt üzerinde değerlendirmesini sağlar.
- Grafik gösterim, araştırmacının davranıştaki ilginç değişimleri olduğu gibi incelemesini sağlar.
- Diyagramlar davranış değişikliğinin istatistiksel analizlerini tanımlar; ancak grafiklerin görsel analizi daha az zaman alır ve bilgilerin deşifre edilmesini kolaylaştırabilir. Ek olarak, görsel analiz istatistiksel varsayımlara bağlı değildir.
- Görsel analiz, davranış değişikliğinin önemini belirlemek için muhafazakâr bir yöntem olarak kabul edilir. Bu nedenle, verilerdeki aralığı, değişkenliği, eğilimleri ve örtüşmeleri tespit eden bir grafik üzerinde çizilen veriler nedeniyle davranış değişikliği istatistiksel olarak anlamlı kabul edilir.

- Görsel analiz, davranış değişikliğine ilişkin bağımsız yargıları ve açıklamaları göstermede etkilidir.
- Görsel analiz aynı zamanda etkili bir geri bildirim kaynağı gösterir ve sunar,

olarak sıralanmaktadır. Sıklıkla kullanılmasına rağmen bu analizde, değerlendiriciler arası uyumun olmadığı ve çalışmaların Tip II hatalarla sonuçlanma eğiliminin söz konusu olduğu belirtilmektedir (Cooper, Heron ve Heward, 2007; Alqraini, 2017). TDDT verilerini yorumlamak için birçok istatistiksel prosedür mevcuttur (görsel analiz, yarı istatistiksel teknikler, geleneksel çıkarımsal istatistik testleri, rastgeleleştirme testleri) ve TDDT verilerinin istatistiksel analizi üzerine devam eden dinamik bir araştırma vardır. TDDT verilerinin analizinde hangi istatistiksel yöntemin kullanılacağı konusunda ortak bir görüş bulunmamaktadır (Manolov ve ark., 2014; Tate ve ark., 2016). TDDT'ler, bireysel düzeyde deneysel değerlendirme yapılmasına olanak tanıyan belirli aşamalardan oluşur ve çalışma süresi boyunca birey üzerinde tekrarlanan ölçümler temelinde yapılandırılır. Bu tasarım, müdahalenin bireysel ihtiyaçlara göre uyarlanmasına imkân tanır, yeni müdahalelerin değerlendirilmesine elverişlidir ve zaman içindeki süreç değişkenleri hakkında içgörüler sunar. Müdahale ile sonuçlar arasındaki işlevsel ilişkiler, aşamalar boyunca toplanan ileriye dönük veriler ışığında görsel olarak analiz edilir ve bu değerlendirme nicel etki ölçümleriyle desteklenebilir. TDDT, hem birey içindeki hem de bireyler arasındaki değişkenliği değerlendirme yeteneği sunar ve beklenen değişim mekanizmalarını analiz etmeye yardımcı olur (Zetterberg ve ark., 2022). İki ana özellik, TDDT'ların tanımlayıcı veya temel özellikleridir. Birincisi, tasarımlar zaman içinde devam eden veya sürekli değerlendirme gerektirir. Ölçümler, ayrı aşamalar içinde birçok durumda uygulanır. İkincisi, müdahale etkileri genellikle aynı katılımcı(lar) içinde zaman içinde tekrarlanır. Tasarımlar, müdahale etkilerinin tekrarlanma şekli bakımından farklılık gösterir, ancak her tasarım, zaman içinde tekrarlanan değerlendirmeden ve farklı koşullar altında performans değerlendirmesinden yararlanır (Kazdin, 2021). TDDT'lar tam olarak nedir? Literatürde bunlar "bazı deneysel tedavi koşullarının etkisini araştırmak için yalnızca bir katılımcı veya bir grup birey kullanan tasarımlar" olarak tanımlanmaktadır. N-of-1 klinik deneylerinde olduğu gibi, TDDT de teori ile pratik arasındaki boşluğu kapatmada etkili olabilir. Bu yaklaşımda, hasta rastgele sırayla farklı tedavilere tabi tutulur ve böylece kendi kontrolünü oluşturur. Bu yöntem, belirli bir hasta için en uygun

tedavinin hızla belirlenmesine, gereksiz yere uzun süren tedavilerden kaçınılmasına ve tedavinin etkilerine dair sistematik veri toplanmasına olanak tanır. Ayrıca, bu süreç hastanın tedaviye bağlılığını artırma potansiyeline sahiptir. Ancak, TDDT tasarlanmış bir müdahale içerirken, vaka çalışmaları herhangi bir müdahalenin bulunmadığı tanımlayıcı bir araştırma türüdür; dolayısıyla ikisi birbirine karıştırılmamalıdır (Bulte ve Onghena, 2008). TDDT'lar, bir veya daha fazla tedavi veya tedavi düzeyi ile biyolojik veya davranışsal sonuçlardaki değişiklikler arasındaki ilişkiyi incelemek için bir deneysel tasarım ailesini temsil eder. Erken dönem deneysel psikoloji araştırmalarında ortaya çıkan bu tasarımlar, temel ve uygulamalı davranış analizi alanlarında genişletilmiş ve resmileştirilmiştir. TDDT'lar, tıp, eğitim, danışmanlık psikolojisi, klinik psikoloji, sağlık davranışı ve nörobilim dâhil olmak üzere çeşitli alanlarda uygulanmıştır. TDDT'lar, hedef davranış veya semptomdaki değişikliklerin müdahalenin bir işlevi olup olmadığını belirlemek için bir çerçeve sağlar. TDDT'ların temelleri, tekrarlanan ölçümü, koşulların tekrarını ve kendi kontrolü olarak hizmet eden her bireyle ilgili etkilerin analizini içerir. Bu süreç, bir birey için en uygun tedaviyi belirlemek, nadir hastalıkları tedavi etmek ve erken evre translasyonel araştırmayı uygulamak için yararlı olabilir. Bu şekilde kullanıldıklarında, TDDT'lar "kişiselleştirilmiş N-of-1 denemeler" olarak adlandırılabilir. TDDT'lar, tek bir kişi hakkında büyük miktarda verinin toplandığı, söz konusu kişinin onun kontrolü olarak görev yaptığı ve deneysel kontrolü geliştirmek için koşulların sırasının rastgele seçilebildiği çok aşamalı deneysel tasarımlardır. Yani, bir kişinin bir aşamadaki sonuçları başka bir aşamadaki sonuçlarla karşılaştırılır. Tipik bir çalışmada, replikasyonlar birkaç kişi içinde ve/veya arasında elde edilir. Bu, davranış ile tedavi veya bunların seviyeleri arasındaki nedensellik hakkında güçlü çıkarımlara izin verir (Epstein ve Dallery, 2022). TDDT'lar en az bir bağımsız değişkenin farklı düzeylerde belirli bir süre boyunca tekrar tekrar gözlemlendiği deneylere uygulanan araştırma desenlerini ifade eder. TDDT'lar, komorbiditeleri veya eşzamanlı tedavileri olan bireylerin dışlanması nedeniyle sıklıkla eleştirilen ve klinik uygulamayla yeterince bağlantılı olmadığı düşünülen paralel grup tabanlı rastgele klinik çalışmaların temel sınırlamasını aşmaktadır (Vlaeyen ve ark., 2020). Çok çeşitli alanlardaki araştırmacılar, müdahalelerin veya tedavilerin etkilerini incelemek için TDDT'ı uygulamışlardır. Bu tür bir metodoloji, otizm, farklı engeller veya davranış bozuklukları olan bireyler gibi özel popülasyonlarda belirli müdahalelerin farklı ilgi sonuçları üzerindeki etkilerini incelemek için davranışsal, eğitimsel, psikolojik ve rehabilitasyon araştırmalarının farklı

alanlarında sıklıkla uygulanmıştır. Bu tür tasarımlarda, bir veya daha fazla durum, zaman içinde bir sonuç üzerinde tekrar tekrar ölçülür. TDDT'ler, bağımlı değişkenin tekrarlanan ölçümelerini en az iki farklı koşul altında – genellikle bir taban çizgisi ve bir müdahale koşulu – karşılaştırarak, tek bir birey üzerinde müdahalenin etkisini değerlendirmek için kullanılabilir (Jamshidi ve ark., 2022). TDDT'lar, müdahalelerin bireysel vakaların davranışları üzerindeki etkilerini incelemek için kullanılan bir dizi yöntemdir. Vakalar, bireysel kişilere, ortak bir bağlamdaki birey gruplarına veya hatta birden fazla ortamda gözlemlenen tek bir bireye karşılık gelebilir. TDDT'ın temel özelliği, farklı tedavi koşulları altında tekrarlanan davranış ölçümüdür; bu ölçümler, genellikle bir başlangıç temel aşaması ve ardından bir tedavi aşaması gibi zamansal aşamalar halinde yapılandırılır. Bu özellikler, her bir vakadaki aşamalar veya koşullar arasında sonuçların karşılaştırılmasıyla müdahale etkilerinin tanımlanmasına izin verir (Chen ve Pustejovsky, 2024). Psikopatolojiyi azaltmada eşit derecede geniş terapötik değişken çeşitliliğinin etkilerini inceleyen klinik araştırmacılar tarafından çok çeşitli basit ve karmaşık deneysel TDDT'lar geliştirilmiştir. Grup tasarımında bir tedavi, bir kontrol prosedüründen istatistiksel olarak daha büyük bir etki yaratıyorsa işe yarar. Bu tür bulgular temel araştırmalarda teorik nedenlerle çok önemli iken, klinik araştırmalarda hastalardaki bu değişiklikler istatistiksel olarak anlamlı ancak klinik olarak yararsız olacak kadar küçük olabilir. TDDT'larda, belirli hastalardaki davranış değişikliğinin boyutu kolayca gözlemlenir ve bu da klinik fayda konusunda karar vermeyi kolaylaştırır. TDDT'lardan elde edilen veriler, sonuçlar zayıf veya net değilse istatistiksel olarak analiz edilebilse de, pratikte bu nadiren gereklidir (Barlow ve Hersen, 1973).

Uygulama

Bu yazının amacı, rehabilitasyon alanında görev yapan profesyonelleri TDDT metodolojisiyle tanıştırmak, klinik rehabilitasyon uygulamalarında bir TDDT'nin nasıl tasarlanıp hayata geçirileceğine dair pratik öneriler sunmak ve klinik araştırmalarda kullanılan TDDT'leri inceleyerek örnek bir veri seti üzerinde nasıl uygulanacağını göstermektir. TDDT'ler, eğitim ve psikolojinin farklı alanlarında, özellikle davranış analizi, klinik psikoloji ve okul psikolojisi gibi disiplinlerde yaygın olarak kullanılmıştır. TDDT'lar az sayıda katılımcıyla, hatta tek bir katılımcıyla uygulanabilir. Bu da onları, yeterince büyük ve geniş bir veri tabanı oluşturmanın zor olabileceği düşük insidanslı engelli bireyler ve gruplar arası deneysel bir tasarım yürütmek için yeterince büyük ve homojen bir örneklem oluşturmanın zor olabileceği diğer popülasyonlarda müdahaleleri incelemek için çekici bir araç haline getirir.

Ayrıca, bireysel katılımcılar için müdahalelerle TDDT'lar uygulanabilir ve müdahale stratejilerinin sürekli olarak iyileştirilmesine olanak tanır (Chen ve Pustejovsky, 2024). Özel gereksinimi olan bireylerin sayısındaki artış ve bu bireylere nitelikli eğitim sunma ihtiyacı, özel gereksinimi olan bireylerle yürütülen deneysel araştırmalarda denk gruplar oluşturmadaki zorluklar, kontrol grubunda yer alan katılımcıların bir süreliğine de olsa deneysel uygulamalardan mahrum kalması gibi etik ve yönetsel sınırlılıklar araştırma yöntemlerindeki arayışları artırmıştır. Bu arayışlarla birlikte, özel gereksinimi olan bireyler için geliştirilmiş uygulamaların etkililiğine yönelik sınamaların güçlü şekilde gerçekleştirilmesi ve özel gereksinimi olan bireylere sunulan uygulamaların kanıta-dayalı olması gereksinimi gibi pek çok neden tek denekli araştırma yöntemlerine olan ilginin ve kullanımının artmasını sağlamıştır (Şahin, 2022).

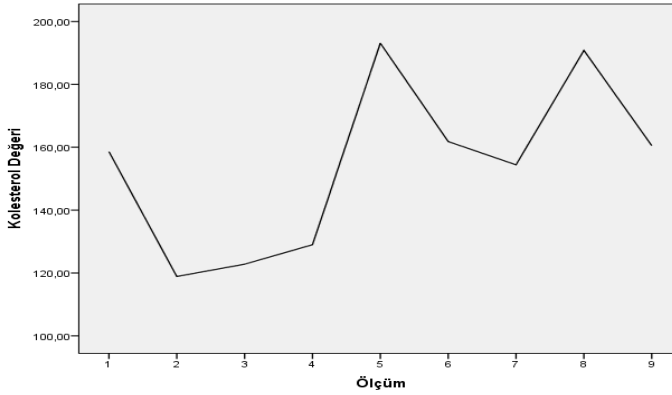
Örnek

Bu örnekte TDDT yöntemlerinden biri olan alternatif tedavi tasarımı için, bir hastanın birbirini ardışık olarak takip eden dokuz farklı zamandaki serum total kolesterol değerleri kullanılmıştır. Tedavi süresince hastanın değerleri gözlemlenerek tedavinin etkili olup olmadığı takip edilmiştir. Hastaya ait serum total kolesterol değerleri Tablo 1'de verilmiştir.

Tablo 1. Örnek Veri

Ölçüm	Total Kolesterol Değeri
1	158.6
2	118.9
3	122.8
4	129.0
5	193.1
6	161.8
7	154.4
8	190.8
9	160.5

Hastaya ait serum total kolesterol değerleri için elde edilen alternatif tedavi tasarımı grafiği Şekil 1'de verilmiştir.



Şekil 1. Alternatif tedavi tasarım grafiği

Şekil 1’den hastaya uygulanan tedavinin etkin bir tedavi olmadığı söylenebilir. Çünkü hastaya ait serum total kolesterol değeri tedavinin sonunda hastanın başlangıçtaki değerinin üzerinde kalmıştır.

SONUÇ

TDDT’lar bireysel bir hasta için en uygun tedaviyi belirlemek üzere kişiselleştirilmiş tıpta kullanılabilecek esnek, titiz ve uygun maliyetli yaklaşımlar sağlar. Geniş bir tasarım yelpazesini temsil eder ve kişiselleştirilmiş (N-of-1) tasarımlar, özellikle tıpta öne çıkan bir örnektir. Bu tasarımlar, rastgele kontrollü denemelere dahil edilebilir ve meta-analiz teknikleri kullanılarak entegre edilebilir. Bireysel vakalar için maksimum etkileri gösteren yeni nesil müdahalelerin yanı sıra klinik pratiğe erken aşama translasyon araştırmalarına yol açabilirler (Epstein ve Dallery, 2022). Tüm bilimsel araştırma tasarımlarında olduğu gibi, tek denekli deneysel araştırmalarda da araştırmanın geçerlik ve güvenilirliğinin sağlanması elde edilen bulguların inandırıcılığı, anlamlılığı ve doğruluğu açısından son derece önemlidir. Bu durumun, özellikle, araştırmada ele alınan uygulamanın bilimsel-dayanaklı bir uygulama olup olmadığı konusunda karar verebilmek açısından önemi tüm bilimsel araştırma tasarımlarında çok büyüktür (Aydın, Tekin-İftar ve Rakap, 2019). TDDT’ların ortak sınırlamaları vardır. Müdahale olmaksızın tekrarlanan testlerle bir ölçüm değiştiğinde, bu tür reaktiviteyi azaltmak için adımlar atılmaz ise, TDDT yararlı olmayabilir. Müdahalelerin etkilerinin zaman içinde değerlendirildiği göz önüne alındığında, sistematik çevresel değişiklikler, bir tedavi ile sonuç arasındaki ilişkiyi etkileyebilir ve böylece bir tedavinin etkisini belirsizleştirebilir. Bununla birlikte, tersine çevirme ve çoklu temel tasarımların tasarım mantığı, bu tür etkileri büyük

ölçüde kontrol eder. TDDT'lar tekrarlanan ölçümlere ve tedavi ile sonuç arasındaki ilişkinin ayrıntılı bir çalışmasına dayandığından, sıklıkla örneklenemeyen bağımlı ölçümler kullanan çalışmalar TDDT'lar için aday değildir. Benzer şekilde, tedavinin başlatılması ile sonuçta değişikliğin başlatılması arasında zamansal bir ilişkinin belirlenememesi, değişikliklerin müdahaleye atfedilmesini zorlaştırabilir. Müdahalenin etkileri hakkında uygun olmayan kararlara neden olabilecek, müdahalenin başlatılması veya kaldırılmasıyla karıştırıcı bir değişkenin ilişkisi her zaman mümkündür. Bireylerin başına gelen bırakma veya kontrolsüz olaylar, TDDT'na kafa karıştırıcı değişkenler getirebilir. Gerçekte bu problemler sadece TDDT'lara özgü değildir, rastgele kontrollü denemelerde de ortaya çıkabilir (Epstein ve Dallery, 2022). Belirli bir TDDT çalışması, bireysel katılımcılar üzerinde bir müdahalenin etkilerine dair klinik açıdan anlamlı ve değerli kanıtlar sunabilir. Ancak, tek bir çalışmadan elde edilen bulguların daha geniş bir popülasyona ne kadar genellenebileceğini belirlemek güçtür. Bununla birlikte, TDDT'lerin yaygın olarak kullanıldığı bazı araştırma alanlarında, benzer müdahaleleri inceleyen çok sayıda çalışma bulunmaktadır. Meta-analiz yöntemleri, bu çalışmaların sonuçlarını bir araya getirerek müdahalenin genel etkilerini, etkilerdeki değişkenliği ve daha büyük ya da daha küçük etkilerle ilişkili faktörleri belirlemeye yardımcı olabilir. Böylece, daha geniş kapsamlı ve dış geçerliliği yüksek sonuçlara ulaşmak mümkün hale gelir (Chen ve Pustejovsky, 2024). Avantajlarına karşın tek denekli araştırmaların araştırma sonuçlarının genellenmesiyle ilgili eleştiriler aldığı, grafik üzerinde görsel olarak gösterilen verilerin farklı araştırmacılar tarafından farklı şekillerde yorumlanabileceği, bir durumu derinlemesine anlamayı gerektiren ve değişkenler arasındaki ilişkiyi ortaya koymayı amaçlayan araştırma sorularını yanıtlamada yeterli olmayabileceği gibi sınırlılıklara sahip olduğu ifade edilmektedir. Tek denekli araştırmaların dış geçerlikle ilgili aldığı eleştirilerin sürekli olarak bireysel müdahalelere yer vermesi ve verilerin bireysel olarak toplanmasından kaynaklandığı belirtilmektedir. Ancak tek denekli araştırma katılımcılarının bireysel özelliklerinin ayrıntılı şekilde tanımlanması, araştırmaların farklı katılımcılar ve farklı koşullar altında tekrarlı ölçümlerle yinelenmesi ve bu araştırmaların dâhil edildiği meta-analiz çalışmalarının gerçekleştirilmesiyle, tek denekli araştırma sonuçlarının genellenebileceğini ve bu eleştirilerin ortadan kaldırılabilceğini söylemek mümkündür (Şahin, 2022). TDDT'lerin en büyük sorunlarından biri, küçük örneklem boyutu nedeniyle genellenebilirliğin sınırlı olmasıdır. Bu sorunu aşmak için araştırmacılar, TDDT'yi farklı vakalar, ortamlar veya davranışlar üzerinde

tekrar ederek genellenebilirliği artırmaya çalışır. Ayrıca, meta-analitik yöntemler kullanarak bu tekrarların sonuçlarını nicel olarak sentezleyebilir ve en iyi uygulamalara yönelik güçlü kanıtlar sunabilirler. Bu tasarımlardan elde edilen sonuçları özetlemek için, TDDT'ların zaman içindeki ilgi sonucundaki doğrusal büyüme ve seri bağımlılık veya otokorelasyon gibi özelliklerini hesaba katan birçok ölçüm ve teknik önerilmiştir. Bunlar, parametrik olmayan ve parametrik metrikleri ve teknikleri ve farklı TDDT çalışmalarında etki büyüklüklerini özetlemek için regresyon ve regresyona dayalı olmayan yaklaşımları içerir. Birden fazla TDDT çalışmasından elde edilen verilerin toplanmasına yönelik ana yaklaşımlar, basit bir ortalama, medyan veya etki boyutları aralığının hesaplanmasını, etki boyutlarının regresyon analizini veya ham verilerin veya etki boyutlarının çok düzeyli analizini içerir. Bu metrik ve analiz yaklaşımlarının bolluğundan kaynaklanan önemli bir soru, hangi yaklaşımların uygun olduğudur. Son on yılda, TDDT'larının meta-analizine olan ilgi artmıştır. Meta-analizin kullanımı gerçekten de bireysel vakalar hakkında bilgi kaybetmeden genel tedavi etkisi hakkında çıkarımlar yapmak ve moderatör değişkenlerin dikkate alınması dâhil olmak üzere çok sayıda fırsat sunar. Meta-analizlerin gerçek etki boyutlarını ve çalışmalar arasındaki bu etki boyutlarındaki sistematik değişkenliği belirlemede özellikle yardımcı olabileceği, nedensel çıkarımlara temel oluşturmak için TDDT'ların önemine rağmen, bu tasarımları meta-analiz etme yollarının tartışmalarla dolu olduğu belirtilmektedir (Jamshidi ve ark., 2022). TDDT'lar, tek bir birim üzerinde zaman içinde tekrarlanan ölçümleri içerdiğinden, her durum için veri serisinin otokorelasyon sergilemesi olasıdır. Pozitif otokorelasyon varlığında, etki büyüklüğü tahminlerinin örnekleme varyansları aşağı doğru yanlı olacaktır, bu da müdahale etkileri hakkında şişirilmiş Tip I hata oranlarına ve yanlış istatistiksel çıkarımlara yol açabilir. Bununla birlikte, bu etki boyutlarının çok düzeyli meta-analizi için otokorelasyonun etkileri henüz incelenmemiştir (Chen ve Pustejovsky, 2024). Görsel analiz, TDDT sonuçlarını yorumlama sürecinin önemli bir bileşeni olmaya devam ederken, TDDT çalışmaları için istatistiksel analiz tekniklerinin kullanımına yönelik argümanların çok fazla olduğu ifade edilmektedir. En azından tüm TDDT raporları, meta-analizi kolaylaştırmak için ham verileri sunmalıdır. Rastgele N-of-1 denemelerinin, bir müdahalenin etkinliği için en yüksek düzeyde kanıt sunduğu kabul edilmiştir ve bu nedenle mükemmel iç geçerliliğe sahiptir. Bununla birlikte, dış geçerlilik replikasyona bağlıdır ve ham verilerin sağlanması, uygulama standartlarının ve klinik kılavuzların geliştirilmesini desteklemek için sağlam bir kanıt temeli oluşturmada kritik olan meta-analizlere dâhil edilecek etki

boyutlarının hesaplanmasına izin verir. Bununla birlikte, istatistiksel analizin önemli olduğu kabul edilmesine rağmen, TDDT verileri için hangi istatistiksel analizin en uygun olduğu konusundaki tartışmalar hâlâ devam etmektedir (Manolov ve ark., 2014).

TDDT'lar, müdahale etkilerini tek bir katılımcı düzeyinde test etmek için güçlü bir yöntemdir. Çıkarımların bireysel düzeyden ziyade grup düzeyinde yapıldığı grup karşılaştırma tasarımlarına göre çeşitli avantajlar sunar. Tek bir vaka hakkında ayrıntılı bilgilerin ortaya çıkmasına izin verir. Tekrarlanan ölçümler, araştırmacıların davranışın veya sonuç ölçüsünün zaman içinde nasıl değiştiğini değerlendirmesine olanak tanıyan aşamalar içinde gerçekleştiği için bağımsız değişkenin bireyi nasıl etkilediğine dair detaylı bilgiler ortaya çıkmakta ve konu içi eğilimler takip edilebilmektedir. Rastgele kontrollü denemeler için değerli bir alternatif veya gerekli bir başlangıç adımı olarak kabul edilen bu tasarım, rastgele kontrollü denemeler yapılmadan önce yenilikçi müdahalelerin etkilerini test etmede ilk adım olarak yürütülebilir. TDDT'lar, ilgilenilen sorunun nadiren ortaya çıktığı ve/veya çok sayıda katılımcı bulmanın zor olduğu durumlarda ve heterojenliği veya spesifik komorbiditesi yüksek popülasyonlarda kullanılabilir. Böylelikle, iki tasarım birbirini tamamlayabilir. TDDT, klinik araştırmacıların hastalarına yönelik klinik bakımı iyileştirmeleri için kullandıkları yöntemlerin standart bir parçası haline gelmelidir.

Kaynaklar

Kitap

- Cooper, J., Heron, T., Heward, W., (2007), *Applied Behavior Analysis*. 2nd edition. Upper Saddle River, NJ: Prentice Hall.
- Özcan P, Han S, Graebner ME. (2017) Single cases: the what, why, and how. In: Mir, RA, Sanjay J, (eds.) *The Routledge companion to qualitative research in organization studies*. New York: Routledge, pp. 92-112. <https://wrap.warwick.ac.uk/96715/>

Dergi Makalesi

- Alqraini, F., (2017), Single-case experimental research: a methodology for establishing evidence-based practice in special education. *International Journal of Special Education*, 32(3), 551-566. <https://files.eric.ed.gov/fulltext/EJ1184160.pdf>
- Aydın, O., Tekin-İftar, E., Rakap, S., (2019), Bilimsel dayanaklı uygulamaları belirlemede “tek-denekli deneysel araştırmaların niteliksel göstergeleri” yönergesinin matematik becerileri öğretimi örneğinde ele alınışı. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 20(3), 597-628. <https://doi.org/10.21565/oezegitimdergisi.421952>
- Barlow, D.H., Hersen, M., (1973), Single-case experimental designs. Uses in applied clinical research. *Archives of General Psychiatry*, 29(3), 319-325. <https://doi.org/10.1001/archpsyc.1973.04200030017003>
- Bulte, I., Onghena, P., (2008), An R package for single-case randomization tests. *Behavior Research Methods*, 40(2), 467-478. <https://doi.org/10.3758/brm.40.2.467>
- Chen, M., Pustejovsky, J.E., (2024), Multilevel meta-analysis of single-case experimental designs using robust variance estimation. *Psychological Methods*, 29(3), 537-560. <https://doi.org/10.1037/met0000510>
- Çakıroğlu, O., (2012), Single-subject research: applications to special education. *British Journal of Special Education*, 39(1), 21-29. <https://doi.org/10.1111/j.1467-8578.2012.00530.x>
- Epstein, L.H., Dallery, J., (2022), The family of single-case experimental designs. *Harvard Data Science Review*, 4(SI3). <https://doi.org/10.1162/99608f92>
- Garbacz, S.A., Kratochwill, T.R., (2020), *Single-Case Experimental Designs*. 1st ed. Oxford: Oxford University Press, <https://doi.org/10.1093/obo/9780199828340-0265>
- Jamshidi, L., et al. (2022), A systematic review of single-case experimental design meta-analyses: characteristics of study designs, data, and analyses. *Evidence-Based Communication Assessment and Intervention*, <https://doi.org/10.1080/17489539.2022.2089334>
- Kazdin, A.E., (2021), Single-case experimental designs: characteristics, changes, and challenges. *Journal of the Experimental Analysis of Behavior*, 115(1), 56-85. <https://doi.org/10.1002/jeab.638>
- Krasny-Pacini, A., Evans, J., (2018), Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of Physical and Rehabilitation Medicine*, 61(3), 164-179. <https://doi.org/10.1016/j.rehab.2017.12.002>
- Lobo, M.A., Moeyaert, M., Baraldi Cunha, A., Babik, I., (2017), Single-case design, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy*, 41(3), 187-197. <https://doi.org/10.1097/NPT.0000000000000187>
- Manolov, R., Gast, D.L., Perdices, M., Jonathan, J., Evans, J.J., (2014), Single-case experimental designs: reflections on conduct and analysis. *Neuropsychological Rehabilitation*, 24(3-4), 634-660. <https://doi.org/10.1080/09602011.2014.903199>
- Plavnick, J.B., Ferreri, S.J., (2013), Single-case experimental designs in educational research: a methodology for causal analyses in teaching and learning. *Educational Psychology Review*, 25(4), 549-569. <https://doi.org/10.1007/s10648-013-9230-6>
- Şahin, V., (2022), Kitap incelemesi: tek denekli araştırma yöntemleri: özel eğitim ve davranış bilimlerinde uygulamalar. *Turkish Journal of Special Education Research and Practice*, 4(1), 92- 99. <https://doi.org/10.37233/TRSPED.2022.124>
- Tate, R.L., et al., (2016), The single-case reporting guideline in behavioural interventions (SCRIBE) 2016 statement. *Archives of Scientific Psychology*, 4(1), 1-9. <https://doi.org/10.2522/ptj.2016.96.7.e1>

- Vlaeyen, J.W.S., et al., (2020), From Boulder to Stockholm in 70 years: single case experimental designs in clinical research. *Psychological Record*, 70(4), 659-670. <https://doi.org/10.1007/s40732-020-00402-5>
- Zetterberg, H., Flink, I., Spöndly-Nees, S., Wagner, S., Karlsten, R., Åsenlöf, P., (2022), Behavioral medicine physiotherapy in the context of return to work for chronic pain: a single-case experimental design study. *International Journal of Environmental Research and Public Health*, 19, 1509. <https://doi.org/10.3390/ijerph19031509>

BİR YÖNLÜ VARYANS ANALİZİNE ALTERNATİF OLAN PARAMETRİK OLMAYAN K-ÖRNEKLEM TESTLERİ *

Aslı Ceren Macunluoğlu **, Gökhan Ocakoglu ***

Giriş

İstatistiksel veri analizinde, kullanılacak olan test yönteminin seçimi analiz sürecinin en önemli adımlarından birisi olarak kabul edilmektedir. Parametrik testler; verilerin aralıklı ya da oransal ölçekle ölçülmüş olmasını gerektirir ve uygulanabilmeleri için gözlemlenen verilerin, belirli varsayımlara dayalı olarak seçilmiş bir anakütleden geldiği kabulüne dayanır. Bu varsayımlar arasında genellikle verilerin normal dağılım göstermesi ve varyansların homojen olması gibi koşullar yer alır. Dolayısıyla, parametrik testler, araştırılan problemlerin çözümünde ve elde edilen sonuçların yorumlanmasında, bu varsayımların sağlandığı durumlarda tercih edilen yöntemlerdir.

Parametrik testlerin uygulanabilmesi için belirli varsayımların sağlanması gerekmektedir. Bu varsayımlar; gözlem değerlerinin bağımsızlığı, ortalamaları karşılaştırılacak örneklemelerin ait olduğu anakütlelerin normal dağılım göstermesi ve anakütle varyanslarının homojen olması şeklinde sıralanabilir (Moder, 2010). Parametrik testlerin doğru şekilde uygulanabilmesi için, bu varsayımların kontrol edilmesi süreci son derece önem taşımaktadır (Leech ve Onwuegbuzie, 2002). Normal dağılım veya varyansların homojenliği varsayımlarının ihlal edilmesi durumunda, testin Tip I hata olasılığı ve gücü olumsuz etkilenmektedir. Bu nedenle, varsayımların sağlanamadığı durumlarda istatistiksel analizlerde hatalı sonuçlar elde edilmemesi adına parametrik olmayan testlerin kullanılması gerekmektedir.

Parametrik olmayan test yöntemleri, gözlemlenen verilerin doğrudan kullanılması yerine, bu verilerin küçükten büyüğe doğru sıralanarak her birine sıra numarası atanması prensibine dayanır. Bu yöntem, hesaplamaların sıra

* Bu çalışma Prof. Dr. Gökhan Ocakoglu danışmanlığında 01.08.2019 tarihinde tamamladığımız “Bir Yönlü Varyans Analizine Alternatif Olan Parametrik ve Parametrik Olmayan K-Örneklem Test Prosedürlerinin Performanslarının Karşılaştırılması” başlıklı yüksek lisans tezi esas alınarak hazırlanmıştır (Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü, Bursa, Türkiye, 2019).

** Doktora öğrencisi, Bursa Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı, Bursa, Türkiye
cerenmacunluoglu@gmail.com, ORCID: orcid.org/0000-0002-6802-5998

*** Prof. Dr., Bursa Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Bursa, Türkiye
ocakoglu@uludag.edu.tr, ORCID: orcid.org/0000-0002-1114-6051

numaraları üzerinden gerçekleştirilmesi nedeniyle parametrik testlere kıyasla daha esnek bir istatistiksel yaklaşım sunar. Parametrik olmayan testler, parametrik testlerin varsayımlarının sağlanmadığı durumlarda tercih edilmektedir. Bu testlerin en belirgin avantajlarından biri, parametrik testlere kıyasla daha az varsayıma sahip olmalarıdır (McSweeney ve Katz, 1978). Örneğin, verilerin normal dağılım göstermesi gerekmediği gibi, sıralayıcı veya sınıflayıcı ölçekle ölçülmüş verilere de uygulanabilirler. Ayrıca, öğrenilmesi, uygulanması ve yorumlanması parametrik testlere göre genellikle daha kolaydır. Ancak, parametrik olmayan testlerin bazı dezavantajları da bulunmaktadır. Eğer parametrik testlerin uygulanması için gerekli varsayımlar sağlanıyorsa, parametrik olmayan testlerin kullanımı veri kaybına veya bilgi israfına yol açabilir. Bu nedenle, eldeki veri setinin yapısına ve incelenen değişkenlerin özelliklerine en uygun test yönteminin seçilmesi büyük önem taşımaktadır.

Bir yönlü varyans analizi (ANOVA), üç veya daha fazla grubun ortalamalarının karşılaştırılmasında kullanılan en önemli parametrik testlerden biridir (Luepsen, 2017). ANOVA, F testi temelinde çalışır ve farklı gruplara ait ortalamaların istatistiksel olarak anlamlı bir farklılık gösterip göstermediğini belirlemek için F istatistiğini kullanır. ANOVA'nın uygulanabilmesi için belirli varsayımların sağlanması gerekmektedir. Bunlar; gruplardaki gözlem değerlerinin bağımsız olması, her grubun ait olduğu anakütlenin normal dağılıma uygunluk göstermesi ve grupların varyanslarının homojen olmasıdır. Eğer bu varsayımlardan biri veya birkaçı ihlal edilirse, ANOVA'nın sonuçları güvenilirliğini kaybedebilir ve testin Tip I hata olasılığı artabilir. Bu tür durumlarda, bir yönlü varyans analizinin parametrik olmayan test alternatifleri olan k-örneklem testleri tercih edilmelidir.

Bu bölümde, bir yönlü varyans analizinin parametrik olmayan alternatifi olarak literatürde yer alan Kruskal-Wallis testi, Kruskal-Wallis testinin permütasyon testi üzerine temellenen modifiye versiyonu, Mood'un Medyan testi, Van der Waerden ve Savage testlerine yer verilmiştir.

Bir Yönlü Varyans Analizi (ANOVA Testi)

Bir yönlü varyans analizi, ikiden fazla anakütle ortalamasının karşılaştırılması amacıyla kullanılan, en önemli istatistiksel yöntemlerden birisidir (Luepsen, 2017).

X_{ij} , i. gruptaki, j. örnekleme göstermek üzere; $i=1, \dots, k$; $j=1, \dots, n_i$;

$N = \sum n_i$ olarak gösterilen tüm gruplardaki örneklem büyüklüğü olsun.

X_{ij} 'ler birbirinden bağımsız, μ_i ve σ_i^2 parametrelerine sahip normal dağılıma uygunluk gösterdikleri varsayılır. Bu parametrelere ait en iyi doğrusal yansız tahmin edicileri Eşitlik (1) ve Eşitlik (2) yardımıyla elde edilir (Patrick, 2007).

$$\bar{X}_{i.} = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i} \quad (1)$$

$$S_i^2 = \frac{\sum (X_{ij} - \bar{X}_{i.})^2}{n_i - 1} \quad (2)$$

Denemeler sonucunda elde edilen veriler $Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$ şeklinde tek faktörlü bir modelle tanımlansın. Burada;

Y_{ij} : i. grup için j. gözlemindeki cevap değişkeni,

μ : Genel ortalamayı,

α_j : j. gözlemin etkisini,

ε_{ij} : deneme hatasını

ifade eder (Buning,1997).

ε_{ij} 'ler normal dağılıma sahip, ortalaması 0 ve σ_i^2 ortak varyansa sahiptir. ε_{ij} 'lerin birbirlerinden bağımsız oldukları varsayılır.

Grup ortalamaları arasında fark olup olmadığını araştırmak için kullanılan hipotez ailesi aşağıdaki gibidir;

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k = 0 \quad i=1, 2, 3, \dots, k$$

$$H_1: \text{en az bir } \mu_i \neq 0$$

H_0 hipotezi, “en az bir ortalama farklıdır” şeklinde olan hipoteze karşı sınanır.

Ortalamalar arasındaki farkları analiz edebilmek için genel kareler toplamı (GKT), gruplar arası kareler toplamı (GAKT) ve grup içi kareler toplamının (GİKT) hesaplanması gerekir. Bu hesaplamalar Eşitlik (3), Eşitlik (4) ve Eşitlik (5) ile gösterilmiştir.

$$GKT = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})^2 \quad (3)$$

$$GAKT = \sum_{i=1}^k n_i (X_{i.} - \bar{X}_{..})^2 \quad (4)$$

$$GİK = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2 \quad (5)$$

Bir yönlü varyans analizi için gerekli olan test istatistiğini hesaplamak için gruplar arası kareler ortalaması (GAKO) ve grup içi kareler ortalaması (GİKO) hesaplanır. GAKO ve GİKO, GAKT ve GİKT' nin serbestlik derecelerine bölünmeleriyle elde edilir. Bu hesaplamalar Eşitlik (6) ve Eşitlik (7) ile ifade edilmiştir. Test istatistiği ise Eşitlik (8) ile elde edilir.

$$GAKO = \frac{GAKT}{k-1} \quad (6)$$

$$GİKO = \frac{GİK}{N-k} \quad (7)$$

$$F = \frac{GAKO}{GİKO} \sim F_{\alpha; k-1, N-k} \quad (8)$$

Eğer $F > F_{\alpha; k-1, N-k}$ ise belirlenen yokluk hipotezi reddedilir.

Bir Yönlü Varyans Analizine Alternatif Parametrik Olmayan K-Örneklem Testleri

Kruskal-Wallis Testi

Tek yönlü varyans analizi testinin parametrik olmayan alternatiflerinden bir tanesi Kruskal-Wallis (KW) testidir. KW testi, bağımsız üç veya daha fazla grubu karşılaştırmak için kullanılan parametrik olmayan bir test yöntemidir (Marozzi, 2013). Gerçek gözlem değerleri yerine gözlem değerlerine verilen sıra numaraları kullanılarak yürütülmektedir. Test istatistiğini hesaplamak için, veriler küçükten büyüğe doğru sıralanır ve her birine sıra numarası verilir (Kruskal ve Wallis, 1952).

KW testi, konum parametresi olarak medyan değerinin gruplar arasında farklılık gösterip göstermediğini test etmektedir ve hipotezler;

$$H_0: \text{med}_1 = \text{med}_2 = \dots = \text{med}_k \text{ ve}$$

$$H_1: \text{med}_i \neq \text{med}_j \text{ şeklinde kurulur.}$$

Veri setinde iki ya da daha fazla gözlem değeri birbirine eşit ise (tekrar eden gözlemler varsa) verilen sıra numaralarının ortalaması alınır. Elde edilen sonuç, eşit gözlemlerin yeni sıra numarası olur (Kruskal ve Wallis, 1952). KW testine ait test istatistiği Eşitlik (9) ile hesaplanmaktadır.

$$KW = \frac{1}{s^2} \left(\sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right) \quad (9)$$

Burada,

n_i : i. grubun örneklem büyüklüğünü ($i=1, 2, \dots, k$),

N : Tüm gruplardaki örneklem büyüklüğünü ($\sum n_i$),

$R_i = \sum_{j=1}^{n_i} R_{ij}$, i. örneklemin sıra numaralarının toplamını

S^2 : Sıra numaralarına ait varyansı ifade etmektedir.

Eşitlik (9)'da yer alan sıra numaralarına ait varyans (S^2) Eşitlik (10) ile elde edilir.

$$S^2 = \frac{1}{N-1} \left(\sum_{i=1}^k \sum_{j=1}^{n_i} r_{ij}^2 - \frac{N(N+1)^2}{4} \right) \quad (10)$$

Eğer veri setinde tekrar eden gözlemler yoksa S^2 , sıra numaralarına ait varyans olmak üzere Eşitlik (11) ile elde edilir.

$$S^2 = \frac{N(N+1)}{N12} \quad (11)$$

Veri setinde tekrar eden gözlemler yoksa KW testine ait test istatistiği Eşitlik (12)'deki gibi hesaplanır.

$$KW = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (12)$$

Elde edilen test istatistiği $KW > \chi_{k-1}^2$ ise belirlenen yokluk hipotezi reddedilir.

Kruskal-Wallis Testinin Permütasyon Testi Üzerine Temellenen Modifiye Versiyonu

Permütasyon testi, randomizasyon testi olarak da bilinir (Efron ve Tibshirani 1994; Good, 1994). Permütasyon testi, verilerin rasgele yeniden atandığı, p-değerinin permütasyon verilerine dayanılarak hesaplandığı bir test yöntemidir (Pesarin, 2001).

Kruskal-Wallis (KW) testinin hesaplanmasında permütasyon ve sıra numarası dönüşümleri olmak üzere iki yöntem mevcuttur. KW testinin permütasyon testi üzerine temellenen modifiye versiyonu, F istatistiğine dayanan permütasyon yönteminin, sıra numaraları yöntemi ile birleştirilmesi sonucunda elde edilir (Kruskal ve Wallis, 1952).

Permütasyonların elde edilmesi sürecine, test istatistiği T ve α anlamlılık seviyesi seçilerek başlanır ve aşağıda yer verilen adımlar uygulanır:

1- $\pi_1, \pi_2, \dots, \pi_n$ araştırmada belirlenen, veri sıralarının tüm farklı permütasyon kümesi oluşturulur.

2- Permütasyon testi için veriler küçükten büyüğe doğru sıralanır ve her birine sıra numarası verilir.

3- KW test istatistiği hesaplanır ($H_1=t_0$).

4- Küçükten büyüğe doğru sıralanan her bir veri için farklı permütasyon (π_i) değeri elde edilir.

5- Elde edilen permütasyon (π_i) değerleri için KW test istatistiği elde edilir ($H_i=H(\pi_i)$).

6- Bu adımlar i ($i= 2, 3, \dots, M$) sefer tekrar edilir.

KW testinin permütasyon testi üzerine temellenen modifiye versiyonuna ait test istatistiği ise Eşitlik (13)'teki gibi hesaplanır (Odiase ve Ogbonmwan, 2005);

$H_0: med_1= med_2= \dots = med_k$ ve

$H_1: med_i \neq med_j$ olmak üzere

$$p_0 = p(H \leq H_i) = \frac{1}{M} \sum_{i=1}^M \psi(t_0 - H_i) \quad (13)$$

Burada

$$\psi(.) = \begin{cases} 1, & \text{eğer } t_0 \geq H_i \\ 0, & \text{eğer } t_0 < H_i \end{cases} \text{dır.}$$

Elde edilen test istatistiği $p_0 \leq \alpha$ ise belirlenen yokluk hipotezi reddedilir.

Mood'un Medyan Testi

Mood'un Medyan (MM) testi, iki bağımsız gruba ait verilerinin test edilmesinde kullanılan medyan testinin, üç ve daha fazla örneklem karşılaştırması için kullanılan genelleştirilmiş versiyonudur (Brown ve Mood, 1951). Bir yönlü varyans analizine alternatif olan bir testtir (Levy, 1979).

MM testine ait test istatistiğini hesaplamak aşağıdaki adımlar uygulanır:

1- Tüm verilere ait ortak bir medyan (M) değeri hesaplanır.

2- Her örneklem için, kaç tane gözlemin, hesaplanan medyan değerinden daha büyük olduğu ve kaç tanesinin medyan değerine eşit veya medyan değerinden daha küçük olduğu belirlenir.

3- Bu adımlar sonucunda $2 \times k$ boyutunda sıklık tablosu elde edilir (Tablo 1).

Burada O_{1j} : j. örneklemede medyan değerinden büyük değerlerin sayısını, O_{2j} : j. örneklemede medyan değerine eşit veya daha küçük değerlerin sayısını ifade eder.

Tablo 1: Mood'un Medyan Testi İçin Gerekli Olan 2xk Boyutunda Sıklık Tablosu

	1	2	...	n_k	Toplam
Medyan değerinden büyük değerler	O_{11}	O_{12}		O_{1k}	A
Medyan değerine eşit ve küçük değerler	O_{21}	O_{22}		O_{2k}	B
Toplam	n_1	n_2		n_k	N

4-Elde edilen sıklık tablosundan beklenen değerler hesaplanır.

$$1. \text{ satırdaki j. örneklemin beklenen değeri: } E_{1j} = \frac{n_1 A}{N}$$

$$2. \text{ satırdaki j. örneklemin beklenen değeri: } E_{2j} = \frac{n_2 B}{N}$$

5-Grup medyanları arasında fark olup olmadığını incelemek için kullanılan hipotez ailesi aşağıdaki şekilde oluşturulur;

$$H_0: \text{med}_1 = \text{med}_2 = \dots = \text{med}_k \text{ ve}$$

$$H_1: \text{med}_i \neq \text{med}_j \text{ (En az bir medyan farklıdır.)}$$

6-Beklenen değerler yardımıyla MM test istatistiği Eşitlik (14) ile hesaplanır.

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (14)$$

Elde edilen test istatistiği, $\chi^2 > \chi^2_{(i-1)*(j-1)}$ ise belirlenen yokluk hipotezi reddedilir.

Van der Waerden Testi

Van der Waerden (VW) testi, parametrik bir test olan bir yönlü varyans analizinin ve parametrik olmayan Kruskal-Wallis testinin avantajlarına sahiptir (van der Waerden, 1953). VW testi, örneklemelerin geldiği anakütlelerin normal dağılıma uygunluk göstermesi durumunda, neredeyse bir yönlü varyans analizi kadar güçlü bir testtir ancak normal dağılıma uygunluk göstermemesi durumunda da Kruskal-Wallis testinin sağlamlılığını göstermektedir (Conover, 1999).

VW test istatistiğini hesaplamak için ilk önce Kruskal-Wallis testinde olduğu gibi veriler, küçükten büyüğe doğru sıralanır ve her birine sıra numarası verilir. Sıra numaraları elde edildikten sonra normal skorlar hesaplanır (van der Waerden, 1953).

Normal skorlar Eşitlik (15) ile hesaplanır.

$$A_{ij} = \phi^{-1} \left(\frac{R_{ij}}{N+1} \right) \quad (15)$$

Burada,

n_i : i. grubun örneklem büyüklüğünü ($i=1, 2, \dots, k$),

N : Tüm gruplardaki örneklem büyüklüğünü ($\sum n_i$),

R_{ij} : i. gruptaki, j. örneklem sıra numaralarının toplamını ve

Φ^{-1} , standart normal dağılım için kümülatif dağılım fonksiyonunu göstermektedir.

Her i. grup için normal skorların ortalamaları Eşitlik (16) ile hesaplanır.

$$\bar{A}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} A_{ij} \quad (16)$$

S^2 , skorların örneklem varyansı olmak üzere Eşitlik (17) ile hesaplanır.

$$S^2 = \frac{1}{N-1} \sum_{i=1}^k \sum_{j=1}^{n_i} A_{ij}^2 \quad (17)$$

Grup medyanları arasında fark olup olmadığını araştırmak için kullanılan hipotezler;

H_0 : $\text{med}_1 = \text{med}_2 = \dots = \text{med}_k$ ve

H_1 : $\text{med}_i \neq \text{med}_j$ (en az bir medyan farklıdır) şeklinde kurulur.

Van der Waerden (VW) test istatistiği Eşitlik (18) ile elde edilir.

$$V = \frac{1}{S^2} \sum_{i=1}^k n_i \bar{A}_i^2 \quad (18)$$

Elde edilen test istatistiği $V > \chi_{k-1}^2$ ise belirlenen yokluk hipotezi reddedilir.

Savage Testi

Savage testi, konum parametreleri arasındaki farklılıkları değerlendirmek için kullanılan bir yönlü ANOVA'ya parametrik olmayan bir alternatif olarak literatürde yer almaktadır. Bu test, özellikle üstel dağılıma uygunluk gösteren veri setlerinde güçlüdür ve ölçek farklılıklarını veya aşırı değer dağılımındaki konum farklılıklarını karşılaştırmada etkilidir (Hajek, 1969).

Savage test istatistiği, Savage skorları kullanılarak hesaplanır. Bu skorları elde etmek için, öncelikle Kruskal-Wallis testinde olduğu gibi veriler küçükten büyüğe doğru sıralanır ve her bir gözleme sıra numarası atanır. Sıra numaraları belirlendikten sonra, Savage skorları, Eşitlik (19) kullanılarak hesaplanır (Desu ve Raghavarao, 2003).

Savage skorları, üstel dağılımdaki sıra istatistiklerinin beklenen değerleridir. Skorların 0 civarında merkezlenmesi için, Eşitlik (19)'da verilen ifadeden 1 değeri çıkartılmaktadır.

$$S_{ij} = \sum_{i=1}^k \frac{1}{N-i+1} - 1 \quad (19)$$

Her i. grup için Savage skorlarının ortalaması Eşitlik (20) ile hesaplanır.

$$\bar{S}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} S_{ij} \quad (20)$$

S_E^2 , skorların örneklem varyansı olmak üzere Eşitlik (21) ile hesaplanır.

$$S_E^2 = \frac{1}{(N-1)} \sum_{i=1}^k \sum_{j=1}^{n_i} S_{ij}^2 \quad (21)$$

Burada,

n_i : i. grubun örneklem büyüklüğü ($i=1, 2, \dots, k$),

N : Tüm gruptaki örneklem büyüklüğüdür ($\sum n_i$).

Hesaplanan Savage skorları ile grup medyanları arasında fark olup olmadığını araştırmak için kullanılan hipotezler;

H_0 : $\text{med}_1 = \text{med}_2 = \dots = \text{med}_k$ ve

H_1 : $\text{med}_i \neq \text{med}_j$ (en az bir medyan farklıdır) şeklinde kurulur.

Bu hipotezler altında Savage test istatistiği Eşitlik (22) ile elde edilir (Desu ve Raghavarao, 2003).

$$T_E = \frac{1}{S_E^2} \sum_{i=1}^k n_i \bar{S}_i^2 \quad (22)$$

Elde edilen test istatistiği $T_E > \chi_{k-1}^2$ ise belirlenen yokluk hipotezi reddedilir.

SONUÇ

Kullanılacak test yönteminin parametrik mi yoksa parametrik olmayan mı olduğuna karar vermek istatistiksel analiz sürecinin en önemli adımlarından bir tanesidir. Parametrik testlerin uygulanabilmesi için gözlem değerlerinin bağımsız olması, karşılaştırılacak örneklemelerin ait olduğu anakütlelerin

normal dağılıma uygunluk göstermesi ve anakütle varyanslarının homojen olması gibi varsayımların sağlanması gerekmektedir. Bu varsayımlar ihlal edildiğinde, testin Tip I hata olasılığı ve gücü olumsuz olarak etkilenmekte, dolayısıyla araştırmacılar hatalı sonuçlar elde edebilmektedir. İstatistiksel analiz sürecinin sonunda elde edilecek sonuçların doğru ve güvenilir olması için, araştırmacıların veri yapısına ve incelenen değişkenlere uygun test yöntemini seçmesi büyük önem taşımaktadır. Eğer parametrik testlerin uygulanabilmesi için gerekli varsayımlar sağlanmıyorsa, parametrik olmayan test yöntemlerinin kullanılması gerekmektedir. Bir yönlü varyans analizi (ANOVA), parametrik bir test olduğu için varsayımlar ihlal edildiğinde kullanılması hatalı sonuçlara yol açabilir. Bu nedenle, bu teste alternatif olarak literatürde parametrik olmayan testlerden Kruskal-Wallis testi, Kruskal-Wallis testinin permütasyon testi temelli modifiye versiyonu, Mood'un Medyan testi, Van der Waerden testi ve Savage testi önerilmektedir.

Kaynaklar

Kitap

- Brown, G. W., Mood A. M. (1951). *On Median Tests for Linear Hypotheses*. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press. pp: 159-166.
- Conover, W. (1999). *Practical Nonparametric Statistics*. 3rd Edition, Wiley, pp. 396-406.
- Desu, M., Raghavarao D. (2003). *Nonparametric Statistical Methods For Complete and Censored Data*. 5th edition, CRC Press, Florida, pp:2 -384.
- Efron, B., Tibshirani, R. (1994). *An Introduction to the Bootstrap*. 1st edition, New York: CRC Press, pp:3-456
- Good, P. (1994). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. 2nd edition, New York: Springer pp:1-219.
- Hajek, J. (1969). *A Course in Nonparametric Statistics*. San Francisco: Holden-Day, 83.
- Leech, N. L., Onwuegbuzie A. (2002). *A Call for Greater Use of Nonparametric Statistics*. Speeches/Meeting Papers.
- Pesarin, F. (2001). *Multivariate Permutation Test With Application to Biostatistics*. 1st edition, New York: Wiley, pp: 1:432.

Dergi Makalesi

- Buning, H. (1997). Robust Analysis of Variance. *Journal of Applied Statistics* 24: 319-332.
- Kruskal, W. H., Wallis, A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47:583-621.
- Levy, K. J. (1979). Pairwise Comparisons Associated With the K Independent Sample Median Test. *The American Statistician*, 33: 138-139.
- Luepsen, H. (2017). Haiko Luepsen Comparison of nonparametric analysis of variance methods: A vote for van der Waerden. *Communications in Statistics - Simulation and Computation*, DOI: 10.1080/03610918.2017.1353613.
- Marozzi, M. (2013). Nonparametric Simultaneous Tests for Location and Scale Testing: A Comparison of Several Methods. *Communications in Statistics - Simulation and Computation* 42:1298-1317.
- McDonald, P (1999). Power, Type I and Type III Error Rates of Parametric and Nonparametric Statistical Tests. *The Journal of Experimental Education*, 67:367-379.
- McSeeley, M., Katz, B. (1978). Nonparametric statistics; Use and nonuse. *Perceptual and Motor Skills*. *SAGE Journals* 46:1023-1032.
- Moder, K. (2010) Alternatives to F-Test in One Way ANOVA in case of Heterogeneity of Variances (a simulation study). *Psychological Test and Assessment Modeling*, 52(4): 343-353.
- Odiase, J., Ogbonmwan, S. (2005). JMASM20: Exact Permutation Critical Values for The Kruskal-Wallis One-Way ANOVA. *Journal of Modern Applied Statistical Methods*, Article 28, 4(2): 609-620.
- van der Waerden, B. (1953). Order Tests for The Two-Sample Problem, II, III, Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen. Serie A, 564, pp 303-310 ve pp 311-316.

Yayınlanmamış Tez

- Patrick, J. (2007). *Simulations to Analyze Type I Error and Power in The Anova F Test and Nonparametric Alternatives* (Unpublished Master's thesis,). University of West Florida, USA.

HASSAS SORU YAKLAŞIMI

Robab Tetik**, İlker Ercan ***

Giriş

Son zamanlarda, davranışları inceleyen araştırmalar, hassas konularda yoğunlaşmaya başlamıştır. Bu kapsamda inançlar, cinsel davranışlar, maddi durumlar, siyasi görüşler, yasal olmayan eylemler ve etnik köken gibi konular genellikle hassas başlıklar arasında yer almaktadır. Hassas konulara ilişkin ilk çalışmalar, 1971 yılında Amerika Birleşik Devletleri'nde federal hükümet tarafından madde bağımlılığı yaygınlığını tahmin etmek amacıyla yürütülmüştür.

Bunun dışında, kadın katılımcıların küretaj yaptırma durumlarını ya da yetişkin bireylerin seçimlerde oy kullanma durumlarını sorgulayan farklı çalışmalar da mevcuttur. Bu tür araştırmalarda en önemli hususlardan biri, katılımcıların verdikleri yanıtların doğruluğudur. Bu nedenle, veri toplama sürecindeki yöntemin sonucu doğrudan etkileyebileceği ifade edilmektedir (Tourangeau ve ark., 2007).

Hassas konulara ilişkin doğrudan sorular yöneltildiğinde, katılımcıların gerçek cevabı vermekten kaçındıkları ya da soruları yanıtsız bıraktıkları gözlemlenmiştir. Katılımcılara özel ve hassas konular hakkında sorular sorulduğunda, sosyal tercih edilebilirlik yanlılığı adı verilen bir etkenin anket araştırmalarını etkilediği yaygın olarak kabul edilmektedir. Bu yanlılık, katılımcıların kendilerini sosyal olarak kabul görülmeyen bir özelliğe sahip olarak göstermekten kaçındıkları durumlarda ortaya çıkan bir ölçüm hatasıdır. Örneğin, suç işlemiş olmak ya da ırkçı görüşlere sahip olmak gibi konular hakkında doğru bilgi vermekten kaçınabilirler (Philips ve Clancy, 1972). Sosyal tercih edilebilirlik yanlılığını azaltmayı hedefleyen araştırmalar, yanıtlayıcıların yasa dışı veya utanç verici faaliyetlerini ve davranışlarını açıklamak konusunda isteksiz olduklarını ortaya koymaktadır (Nuno ve ark., 2015).

** Dr., Bursa Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Bursa, Türkiye, robabahmadian@outlook.com, ORCID: orcid.org/0000-0003-1550-639X

*** Prof. Dr., Bursa Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Bursa, Türkiye iercan@msn.com, ORCID: orcid.org/0000-0002-2382-290X

Araştırmacılar, toplum tarafından kabul görülmeyen davranışlarla ilgili anket sonuçlarının güvenilirliği konusunda endişe duymaktadırlar. Hassas konulara yönelik anketlerden elde edilen veriler incelendiğinde, gerçekçi olmayan raporlama oldukça yaygındır. Bu durumun, raporlanan konunun katılımcının zorlandığı bir konu olma durumuna ve anketin nasıl tasarlandığına da bağlıdır. Yanlış raporlamanın önemli bir hata kaynağı olduğu yapılan metodolojik araştırmalarda gösterilmiştir (Tourangeau ve ark., 2007).

Sosyal açıdan hassas kabul edilen konulara yönelik yapılan araştırmalarda, literatürde farklı yöntemler üzerinde çalışılmıştır. Geliştirilen yöntemler, sosyal tercih edilebilirlik yanlılığını azaltarak ölçüm hatalarını en aza indirmeyi ve verilen yanıtların güvenilirliğini artırmayı hedeflemektedir. Aynı zamanda, bu teknikler katılımcıya güven ve gizlilik sağlama amacını da taşımaktadır. Bu bağlamda, dolaylı soru sorma teknikleri kullanılarak sosyal tercih edilebilirlik yanlılığı sorununa yönelik çözümler geliştirilmektedir (Nuno ve ark., 2015).

Bilimsel araştırmanın temel bileşenleri arasında deneyler, teoriler ve veriler yer almaktadır. Genel anlamda veri, belirli bir sonuca ulaşmak amacıyla toplanan bilgileri ve belgeleri ifade eder. Araştırma sonucunda elde edilen veri ise, araştırmada incelenecek olan hipotezleri test etmek için araştırma sürecinden elde edilen her türlü bilgiye verilen isimdir. Araştırma sürecinde, veri olarak kabul edilen bilgiler değerlendirilerek yorumlanmakta ve sonuçlar ortaya çıkarılmaktadır. Veri toplama süreci çeşitli yöntemler aracılığıyla gerçekleştirilebilir. Görüşmeler, dokümanların veya kayıtların incelemesi, gözlemler, deneyler, ölçekler ve anketler en yaygın kullanılan veri toplama yöntemleri arasındadır.

Görüşme: Yüz yüze ya da telefon aracılığıyla katılımcılara belirli sorular yöneltilir ve alınan yanıtlar yazılı olarak kaydedilir veya sesli/görüntülü olarak kayıt altına alınabilir. Sorular genellikle önceden hazırlanır ancak görüşmenin akışına göre ek sorular eklenebilir (Mantha, 2016).

Doküman İncelemesi: Araştırma için gerekli olan belgeler ve kayıtlar hâlihazırda mevcut olup, doğrudan bireylerden veri toplama gerektirmez. Yayınlanmış istatistikler, veri tabanları, günlük kayıtlar ve gazete arşivleri bu yöntemin örnekleri arasında sayılabilir (Mantha, 2016).

Gözlem: Sistematik bir şekilde izleme ve kayıt tutma esasına dayanan bu yöntem, doğrudan bilgi sağlamaktadır. Gözlem yapan kişinin konuyla ilgili

yeterli bilgi sahibi olması ve gözlemde tarafsız yaklaşımı büyük önem taşımaktadır (Mantha, 2016).

Deney: Kontrollü bir ortamda gerçekleştirilen ve bir grup denek üzerinde uygulanan bu yöntem, özellikle tıp ve psikoloji gibi alanlarda yaygın olarak kullanılır. Araştırma konusunun belirli bir grup üzerindeki etkisini ölçmek amacıyla tercih edilir (Mantha, 2016).

Ölçekler: Bireylerin belirli davranışsal özelliklerini ölçmek için oluşturulan, güvenilirliği ve geçerliliği test edilmiş ölçme araçlarıdır (Ercan ve Kan, 2004).

Anket: Katılımcılardan bilgi toplamak için kullanılan ve kapalı veya açık uçlu sorulardan oluşan bir yöntemdir. Anketler yüz yüze, telefonla, bilgisayar aracılığıyla ya da posta yoluyla uygulanabilir. Açık uçlu sorular, katılımcılara düşüncelerini daha detaylı açıklama fırsatı sunarken, kapalı uçlu sorular verilerin daha yapılandırılmış bir şekilde analiz edilmesine olanak tanır. Anketler, geniş katılımcı gruplarından hızlı ve maliyet açısından verimli bir şekilde veri toplanmasını sağlar (Sapsford ve Jupp, 1996).

Anketlerin Hazırlanışı

Başarılı bir araştırmanın ilk adımı, amaca uygun ve etkili bir anket hazırlamaktır. Bunu başarabilmek için araştırma konusunun, hedefin ve örneklem grubunun detaylı bir şekilde belirlenmesi gereklidir. Anketi yanıtlayacak bireylerin eğitim, ekonomik durum, sosyal ve politik yapıları iyi analiz edilmelidir. Soruların, katılımcıların değerleriyle çatışmaması, anlama seviyelerine uygun olması ve ilgi alanlarına hitap etmesi önemlidir. Anketler, bireylerin inançlarını, düşüncelerini, tutumlarını, tercihlerini ve bilgi seviyelerini ölçmeyi amaçlayabilir (Brace, 2018).

Anket soruları genellikle şu kategorilere ayrılmaktadır:

Demografik Sorular: Katılımcıların yaş, eğitim düzeyi veya sosyo-ekonomik durum gibi kriterlere göre değerlendirmek amacıyla sorulur kullanılır (Brace, 2018).

Davranışsal Sorular: Bilişsel, duyuşsal ve psikomotor olmak üzere üç farklı davranış türünü kapsar. Bilişsel davranışlar bilgi edinme ve akıl yürütme ile ilgilidir ve genellikle soru sorularak ölçülür. Duyuşsal davranışlar sevgi, nefret veya önyargılar gibi duygusal durumları içerir; bu tür veriler doğrudan sorular veya gözlemler yoluyla elde edilebilir. Psikomotor davranışlar ise

fiziksel hareketleri kapsadığı için doğrudan gözlenebilir ve ölçülebilir (Bacanlı, 1999; Peker, 2000).

Tutum ve İnanç Soruları: Katılımcının belirli bir konudaki düşüncelerini veya hislerini belirlemek için sorulan sorulardır. Tutumların ölçülmesi genellikle birden fazla soru içeren ölçeklerle yapılır ve çoğunlukla kapalı uçlu sorular tercih edilir. Ancak, bazen de açık uçlu sorular sorulabilir, bu durumda yanıtlar kodlanarak analiz edilebilir (Brace, 2018).

Bilgi Soruları: Katılımcıların belirli bir konudaki bilgi seviyelerini, bu bilgiyi nereden ve ne zaman edindiklerini anlamaya yönelik sorulardır (Brace, 2018).

Hassas Sorular: Kişisel veya sosyal açıdan mahrem kabul edilen, hatta yasadışı davranışlarla ilgili konuları içeren sorulardır. Özellikle sağlık, psikolojik ve sosyolojik araştırmalarda sıkça yer alır. Hassas sorulara doğru yanıtlar alabilmek anket çalışmalarında önemli bir zorluk oluşturur (Barnett, 1998).

Hassas Soru

Araştırmalarda, üreme geçmişi, cinsel yaşam, küretaj, HIV enfeksiyonu, yasa dışı madde kullanımı, aile içi şiddet, gelir düzeyi, her türlü istismar, hırsızlık, sosyal güvenlik dolandırıcılığı, ehliyet olmadan araç kullanımı, evlilik dışı doğum, vergi kaçırma ve sınavlarda kopya çekme gibi konular hassas kabul edilmektedir. Bu konulara ilişkin sorular, diğer konulara göre daha yüksek oranda yanıtsız bırakılma veya ölçüm hatalarına yol açma eğilimindedir. Bu nedenle, bir araştırmada sosyal açıdan hoş karşılanmayan bir durumun sorulması ya da katılımcının bir toplumsal normu ihlal ettiğini kabul etmesinin beklenmesi durumunda, bu tür sorular hassas soru olarak değerlendirilir. Hassasiyet derecesi, katılımcının verdiği yanıtlarla belirlenir; örneğin, oy kullanma hakkında sorulan bir soru, oy vermiş bir kişi için hassas kabul edilmezken, oy vermeyen biri için hassas olabilir. Hassas sorular, kişisel bilgilerin açığa çıkmasına, sosyal onay beklentisine veya belirli tehditlere yol açtığına hassas olarak nitelendirilir (Tourangeau ve Yan, 2007).

Hassas sorular yalnızca sosyal normlara aykırı davranışları ortaya çıkaran sorularla sınırlı değildir; aynı zamanda katılımcılar tarafından özel ve mahrem olarak algılanan, ifşa edilmesi halinde olumsuz sonuçlar doğurabilecek soruları da içerir. Örneğin, kokain bulundurmak sadece sosyal açıdan kabul görmeyen bir durum olmakla kalmaz, aynı zamanda yasa dışıdır. Bu nedenle, insanlar yalnızca kötü bir izlenim bırakmaktan kaçınmak için değil, aynı

zamanda hukuki sonuçlardan korunmak amacıyla bu tür araştırmalarda yanlış beyanda bulunabilirler (Tourangeau ve Yan, 2007).

Hassas soruların yanıtlanma oranlarını değerlendiren bir araştırmada, Amerikan Ulusal Aile Büyüme Araştırması (NSFG) kapsamında toplanan küretaj verileri ile merkezlerden temin edilen veriler karşılaştırılmıştır. Bu çalışmada, 15 ile 44 yaş arası kadınlardan oluşan ulusal bir örneklem kullanılmış ve hem anket verileri hem de sağlık kuruluşlarından alınan kayıtlar üzerinden toplam kürtaj sayısı tahmin edilmiştir. Sonuçlar, anketlerde bildirilen kürtaj vakalarının yalnızca %52'sinin gerçeği yansıttığını göstermiştir (Fu ve ark., 1998).

Hassas konulara ilişkin sorular sorulduğunda, katılımcılar sosyal olarak kabul gören davranışları olduğu gibi bildirirken, toplum tarafından onaylanmayan davranışları eksik bildirme eğiliminde olabilirler (Barnett, 1998; Lee, 1993; Rasinski ve ark., 1999; Singer ve ark., 1995; Tourangeau ve ark., 2000; Tourangeau ve ark., 2007). Bu tür yanıltıcı beyanlar, hassas davranışların toplumda gerçek yaygınlığının olduğundan düşük tahmin edilmesine neden olur. Ayrıca, sosyal uygunluk yanlılığı, hassas konulara ilişkin çalışmaların geçerliliğini düşürebilir. Bu tür anketlerde reddetme yanlılığı ve yanıt yanlılığı gibi sorunlardan dolayı doğru çıkarımlar yapılması zorlaşmaktadır (Fisher, 1993). Bu tür yanlılıkların kaynağı, soruların hassasiyet derecesi, soru yapısı, veri toplama yöntemi, katılımcının ve anketörün özellikleri gibi çeşitli faktörlere dayanmaktadır. Araştırmacılar yanlılıkları mümkün olduğunca ortadan kaldırmak ve dolayısıyla anakütle parametresini en az hata ile tahmin edebilmek için birçok yöntem geliştirmiştir. Bu yöntemlerin bir kısmı, katılımcıların güvenliğini artırmayı ve yanıtlarının gizliliğini sağlamayı hedeflemektedir (Lee, 1993). Hassas soru çalışmalarında, bu kapsamda kullanılan başlıca yöntemler arasında randomize yanıt yöntemi, aday yöntemi, gruplandırılmış yanıt yöntemi, fasulye yöntemi, eşsiz-sayım yöntemi ve çapraz model bulunmaktadır.

Hassas Soru Yöntemleri

Rasgele Yanıt Yöntemi (Randomized Response Technique)

Hassas konularla ilgili soruların doğurduğu zorlukların üstesinden gelmek amacıyla geliştirilen ilk görüşme tekniklerinden biri, rastgele yanıt yöntemidir (Warner, 1965). Bu yöntem (RRT), katılımcının gizliliğini koruyarak, rastgele bir mekanizmanın (örneğin zar veya çark) sonucuna göre hassas bir soruyu yanıtlamasını sağlar. Warner'ın önerdiği bu teknik sayesinde, bireyler hassas

konular hakkında bilgi verebilirken, yanıtlarının doğrudan belirlenmesi engellenmiş olur.

Bu yöntemde, katılımcılardan bir rastgele karar mekanizmasına bağlı olarak yanıt vermeleri istenir. Böylece yanıtlayan kişi, gerçek cevabını açıklamadan soruya karşılık verebilir. RRT, katılımcı mahremiyetini korumanın yanı sıra istatistiksel doğruluğu artırmak amacıyla çeşitli metodolojik yeniliklerle geliştirilmiştir (Höglinger ve ark., 2016).

Yöntemin uygulanışına örnek olarak, katılımcının iki zar atması ve kendisine verilen kapalı bir zarf içindeki hassas soruyu yanıtlaması gösterilebilir. Örneğin:

Katılımcının cinsel deneyiminde vajinismus problemi yaşayıp yaşamaması bir soru şeklinde düzenlenir ve “Evet” ve “Hayır” seçenekleri sunulur.

Daha sonra katılımcıya şu talimatlar verilir:

Attığın zarların rakamlarını topla ve bu değer 5 ile 10 arasında olursa, soruyu dürüstçe yanıtlayın.

Bu durumun olasılığı 0,75’tir.

Attığın zarların rakamlarını topla ve bu değer 2, 3 veya 4 ise zarfı açmadan "Evet" seçeneğini işaretle.

Bu durumun olasılığı 0,167’tir.

Attığın zarların rakamlarını topla ve bu değer 11 veya 12 ise zarfı açmadan "Hayır" seçeneğini işaretle.

Bu durumun olasılığı 0,083’tür.

Katılımcı zarlardan elde edilen rakamları görüşmeciyle paylaşmadığı için, verilen yanıtın gerçek mi yoksa rastgele mi olduğu ayırt edilemez. Böylece, hem mahremiyet korunur hem de hassas verilerin toplanması sağlanır.

Formül 1’de Hox ve Lensvel’in (2004) hassas davranışların olasılığı verilmiştir.

$$\pi_{\text{RRT}} = \frac{\lambda - \theta}{s} \quad (1)$$

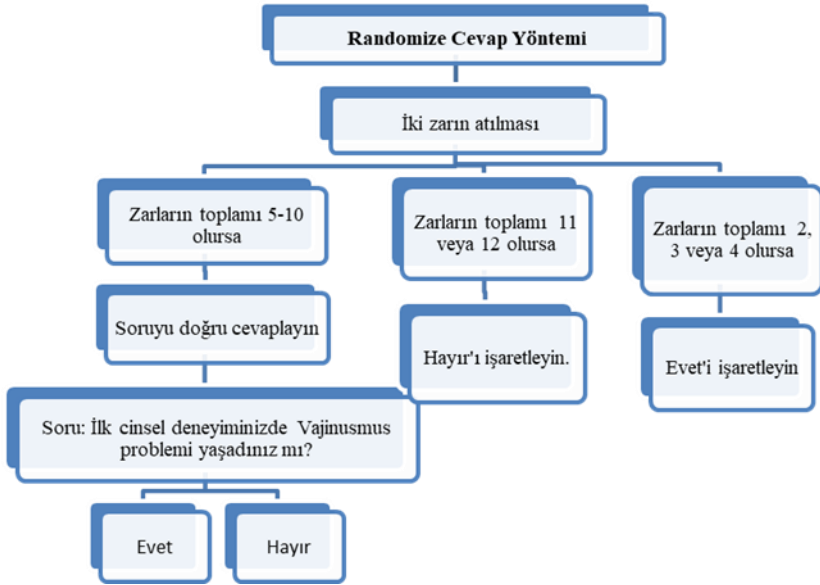
π : hassas davranışın tahmini oranı,

λ : elde edilen 'Evet' yanıtların oranı,

θ : cevabın 'zorla Evet' olma oranı ($\theta=0,167$),

s: hassas soruyu doğru bir şekilde cevaplama oranı ($s=0,75$).

Şekil 1’de Randomize cevap yöntemine ait bir diyagram çizilmiştir.



Şekil 1. Randomize cevap yöntemin algoritması

Rastgele Yanıt Yönteminin Uygulama Alanları ve Eleştirileri

Rastgele yanıt yöntemi, çeşitli hassas konuların araştırılmasında kullanılmaktadır. Bu yöntemin uygulandığı çalışmalardan bazıları şunlardır: Vieira (2009) tarafından yasadışı kürtaj üzerine yapılan araştırma, Böckenholt ve Vander Heijden'in (2007) sosyal güvenlik dolandırıcılığına yönelik çalışması, Simon ve arkadaşlarının (2006) yasadışı uyuşturucu kullanımıyla ilgili araştırması ve Blank ile Gavin'in (2009) illegal avlanma üzerine yaptığı inceleme. Ayrıca, okuryazarlık seviyesi düşük bireyler için de erişilebilir bir yöntem olduğu belirtilmiştir (Razafimanahaka ve ark., 2012; Solomon ve ark., 2007).

Bu yöntemin temel özelliklerinden biri, rastgele bir mekanizma kullanılması nedeniyle ek bir değişkenlik kaynağına sahip olmasıdır. Bu durum, analizlerin güvenilirliği açısından daha büyük örneklem gruplarının kullanılmasını gerektirir (Lensvelt-Mulders ve ark., 2005).

Aday Yöntemi (Nominative)

Aday yöntemi, aynı zamanda ağ örnekleme tekniği ismiyle de bilinmekte olup, esnek bir örnekleme yöntemidir (Sirken MG, 1972). Bu yaklaşım, ilk önce seyrek görülen hastalıklar ve ortaya çıkan olayları araştırmak için Monroe Sirken tarafından 1975 yılında geliştirilen anket çalışmalarının bir parçası olarak ortaya çıkmıştır. Aday yönteminde, katılımcılardan yalnızca kendi davranışları ya da yaşadıkları hakkında değil, aynı zamanda arkadaşları veya aile bireyleri gibi yakın çevrelerindeki kişilerin belirli davranışları ya da yaşadıkları hakkında da bilgi vermeleri istenir.

Aday yönteminin hassas konulara yönelik ilk uygulamalarında, katılımcılardan arkadaşlarının ne kadarının belirli bir hassas davranışta bulunduğunu tahmin etmeleri istenmiştir. Daha sonra, bu tahminlerden elde edilen yüzdelerin ortalaması alınarak, ilgili hassas davranışın toplumda ne kadar yaygın olduğu hesaplanmıştır (Sudman ve ark., 1988).

Aday yöntemi, tüm katılımcıların benzer sayıda arkadaşına sahip olduğu varsayımına dayanır. Ayrıca, nüfustaki her bireyin, hassas davranışta bulunan arkadaşlarının sayısını doğru bir şekilde bildirdiği kabul edilir. Bu bilgiler, uygun bir düzeltme yöntemiyle işlenerek, hassas davranışların gerçek yaygınlığına dair daha doğru bir tahminde bulunulabilir.

Bu düzeltme süreci, bir kişinin hassas davranışını bildiren birden fazla arkadaşının olabileceğini hesaba katar. Örneğin, belirli bir toplulukta bir bireyin hassas bir davranışta bulunduğunu düşünelim. Bu bireyin n kadar yakını, onun bu davranışını bildiğini ifade ettiğinde, aynı davranış birden fazla kez raporlanmış olur. Bu durumda, her bir raporun gerçek sayıyı şişirmesini önlemek adına, her rapor $1/n$ oranında ağırlıklandırılarak hesaplanmalıdır.

Başka bir ifadeyle, araştırmacıların bir bireyin hassas davranışı hakkında verilen tüm raporları toplarken, her bir raporun sadece $1/n$ 'lik bir kısmını dikkate alması gerekir. Böylece, toplamda gerçek sayıya daha yakın bir tahmin elde edilebilir. Aday yöntemi için önerilen çoğaltma düzeltmesi formülü de bu mantık üzerine kurulmuştur. Formülde paydada 1 eklenerek, hiç kimse ilgili hassas davranış hakkında bilgi bildirmediğinde paydanın sıfır olmasının önüne geçilmiştir (Sudman ve ark., 1988).

$$\frac{1}{1+B_{ij}} \quad (2)$$

Aday yönteminin erken versiyonlarında, katılımcılara iki temel soru yöneltilmiştir:

A. Yakın çevrenizde, ilk cinsel deneyiminde vajinismus problemi yaşayan kaç arkadaşınız var?

Eğer katılımcı, vajinismus sorunu yaşamış bir veya daha fazla arkadaş bildirmişse, her biri için şu ek soru sorulmuştur:

B. Bu kişinin dışında, onun bu probleminden haberdar olan başka kaç arkadaşı var?

Ancak, pratik uygulamalarda birden fazla vajinismus sorunu yaşayan yakın arkadaşını rapor edilmesi durumunda, bu kişilerden biri rastgele seçilmelidir. A sorusunda belirtilen hassas durumu yaşayan bireyler arasından yalnızca birinin rastgele belirlenmesi, yöntemin sağlıklı işleyişini korumak için gereklidir (Miller, 1985).

Pratik aday yöntemi kapsamında, katılımcılardan belirli bir hassas konu hakkında bilgi sahibi oldukları arkadaşlarının sayısını belirtmeleri istenir. Sonrasında, bu kişilerden rastgele birini aday göstermeleri talep edilir. Ardından, aday gösterilen kişinin bu hassas durumu hakkında bilgi sahibi olan kaç kişi olduğu sorulur.

Bu süreç, aynı kişiyi birden fazla katılımcının rapor etmesiyle oluşabilecek veri fazlalığını önlemek amacıyla çoğaltma düzeltmesi uygulanarak işlenir. Böylece, araştırılan hassas konunun toplumdaki yaygınlığına dair daha doğru bir tahmin yapılabilir (Miller, 1985). Bu şekilde formül 3'te hassas sorunun yaygınlık oranı verilmiştir.

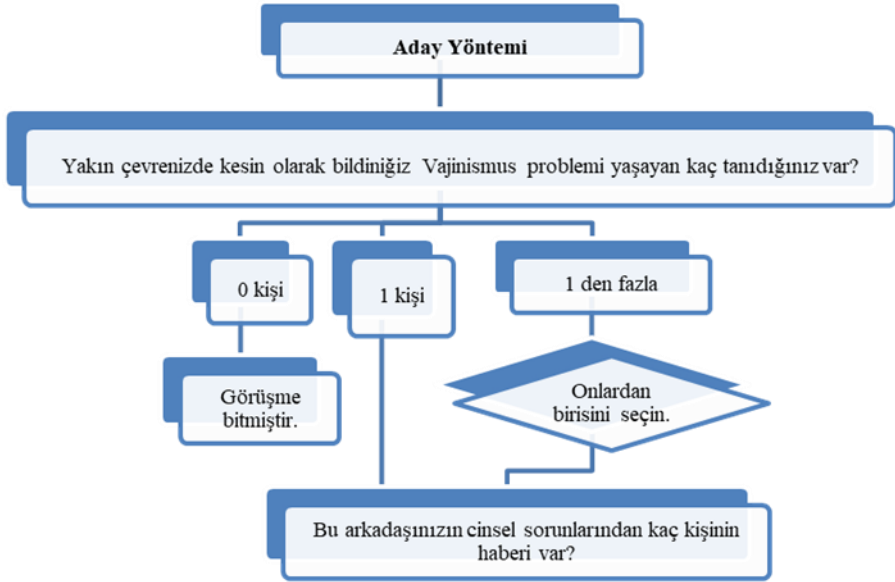
$$\pi_{NT} = \sum_{j=1}^n \frac{A_j}{1+B_j} \quad (3)$$

π_{NT} : n boyutlu bir örnekleme hassas davranışa sahip kişilerin sayısı,

A_j : j. Katılımcı tarafından bilinen hassas davranışta bulunanların sayısı,

B_j : Hassas davranışta bilgisi bulunan kişi sayısıdır.

Şekil 2'de Aday yöntemine için bir diyagram çizilmiştir.



Şekil 2. Aday yöntemin algoritması

Aday Yönteminin Uygulama Zorlukları ve Kullanım Alanları

Aday yöntemi uygulanırken, hassas bir davranış hakkında bilgi sahibi olan arkadaşlarının sayısını eksik bildirilebilir. Bundan dolayı, bu yöntemle yapılan tahminler dikkatli bir şekilde değerlendirilmelidir. Ayrıca, aday yönteminin genellikle yüksek tahminler verme eğiliminde olduğu bilinmektedir (Miller, 1985).

Bu yöntem, başlangıçta Amerika Birleşik Devletleri Ulusal Uyuşturucu Bağımlılığı Enstitüsü tarafından eroin kullanımı üzerine yapılan araştırmalar için geliştirilmiştir. İlgili araştırmada üç farklı dönemde eroin kullanımının yaygınlığını belirlemek amacıyla yapılmıştır. Her defasında, aday yöntemiyle elde edilen sonuçlar, anonim bir şekilde kendi başına katılımcıların doldurdukları anketlere göre eroin kullanımının daha yaygın olduğunu göstermiştir.

Aday yönteminin en büyük avantajlarından biri, uygulanmasının oldukça pratik ve kolay olmasıdır (Miller, 1985).

Eşsiz Sayım Yöntemi (Unmatched-Count Method)

Eşsiz sayım yöntemi (UCT), ilk kez Dalton, Wimbush ve Daily (1994) tarafından uygulanmıştır. Yöntem, bireylerin hassas davranışlarını doğrudan

işşa etmeden, bu davranışların genel yaygınlık tahminini yaparak tam anonimlik sağlar. Son 30 yıl içinde, cinsel riskli davranışlar araştırmasında (Hubbard ve ark., 1989), tehlikeli sürüş araştırmasında (Sheppard ve Earleywine, 2013), ırkçı önyargılar olasılığında (Blair ve Imai, 2012) ve yasa dışı avlanmalar (Nuno ve ark., 2013) gibi hassas konularda yaygın olarak kullanılmıştır.

Bu yöntemde, katılımcılara, kendilerinin hassas davranışlarıyla ilgili doğrudan sorular yöneltilmez. Bunun yerine, aynı anda sosyal olarak kabul edilen veya tarafsız bir dizi soruya ek olarak, hassas davranışa ilişkin bir soru sorulur. Bu durumda, hassas davranışın yaygınlık tahminini yapabilmek için, diğer tarafsız soruların yaygınlık oranı kullanılır. Bu nedenle, eşsiz sayım yöntemi, hassas olmayan davranışlara dair referans sorularını ve ek olarak hassas davranış içeren soruyu gerektirir.

Yöntemin uygulamasında iki ayrı soru listesi hazırlanır. Bu listeler arasındaki tek fark, birinin hassas soruyu içermesi nedeniyle daha uzun olmasıdır. Katılımcılardan, yalnızca hangi davranışların kendilerine ait olduğunu belirtmeleri istenir, ancak hangi davranışların söz konusu olduğu hakkında herhangi bir bilgi vermezler. Hesaplama, hassas olmayan grup ile hassas soru içeren grup arasındaki ortalama davranış farkından yapılır. Bu şekilde, katılımcıların anonimliği korunarak, hassas davranışların yaygınlığı tahmin edilir. Yapılan araştırmalar, eşsiz sayım yönteminin, çalışan istismarı, hırsızlık ve riskli cinsel davranışlar gibi hassas davranışları daha yüksek oranda tahmin ettiğini göstermektedir (Droitcour ve ark., 1991).

Bir örnek olarak, Şekil 3'te, katılımcılara yöneltilen hassas olmayan sorular şu şekilde sıralanmıştır: " idrar yolu enfeksiyonu yaşamadım", "Kadın doğum hekimi muayenesinde özel muayene tercih ederim", "Doğum kontrol hapları kullandım", "Düzenli Pap smear testi yaptırırım". Hassas soru ise, "İlk cinsel deneyimimde vajinismus problemi yaşadım" şeklindedir. Katılımcılar, rastgele bir şekilde kontrol ve tedavi gruplarına ayrılır. Kontrol grubu, hassas sorular bulunmayan bir liste alırken, tedavi grubu, hassas olmayan maddelere ek olarak hassas bir madde içeren listeyi alır. Katılımcılardan, hangi maddelerin kendilerine uyduğunu belirtmeleri istenir. Hassas davranışın yaygınlığı, tedavi grubundaki "evet" yanıtlarının oranından, kontrol grubundaki "evet" yanıtlarının oranı çıkarılarak hesaplanır.

Bu yöntemin temel noktası, katılımcıların hangi ifadelerin kendilerine ait olduğunu değil, sadece hangi ifadelerin geçerli olduğunu saymalarıdır. Bu sayede, katılımcıların anonimliği tamamen korunur (Tsuchiya ve ark., 2007).

Yanıtlayıcı, hassas soruyu içeren listedeki tüm maddelere aynı şekilde cevap verirse, yanıt vermekten kaçınabilir. Bu durumda, madde listesinde, yaygınlığı düşük ve yüksek olan soruların bulunması gerekmektedir. Ayrıca, hassas olmayan soruların, hassas soruyla alakalı konulardan seçilmesi önemlidir, çünkü tamamen farklı konularda sorular seçilmesi, şüpheye yol açabilir (Tsuchiya ve ark., 2007). Eşsiz sayım yönteminde yaygınlık tahmini formül 4’te verilmiştir:

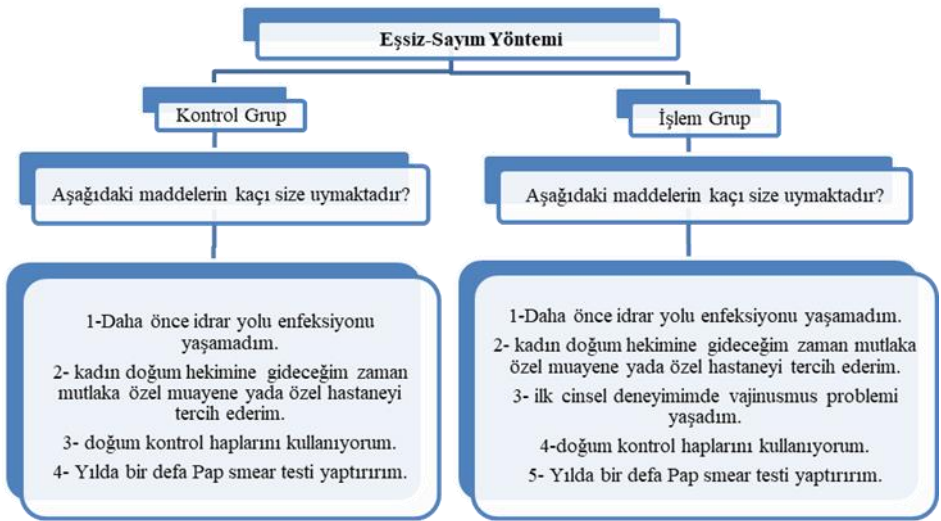
$$\pi_{UCT} = p_{H+} - p_{H-} \quad (4)$$

π_{UCT} : tahmin edilen hassas davranış oranı,

p_{H+} : hassas soruyu yanıtlayan grupta ‘Evet’ yanıtlarının oranı,

p_{H-} : hassas soruyu yanıtlamayan grupta ‘Evet’ yanıtlarının oranı.

Şekil 3’te eşsiz-sayım yöntemine ait bir diyagram ise çizilmiştir.



Şekil 3. Eşsiz-Sayım yönteminin algoritması

Eşsiz Sayım Yönteminin Avantajları ve Sınırlamaları

Bazı araştırmalar, eşsiz sayım yönteminin hassas davranışların yaygınlık tahmininde, doğrudan sorulan anketlere kıyasla daha etkili olduğunu göstermektedir (Sheppard ve Earleywine, 2013; Tsuchiya ve ark., 2007). Ayrıca, eşsiz sayım yöntemi, randomize cevap yöntemine göre daha basit ve anlaşılır bir yaklaşım olarak değerlendirilmiştir. Ancak, bu yöntemin nadiren

görülen davranışlarla ilgili uygulamalarında, elde edilen verilerde yüksek standart hatalar bulunabilmektedir, bu da bir olumsuzluk olarak kabul edilir.

Eşsiz sayım yöntemi, randomize cevap yöntemi ile karşılaştırıldığında önemli bir avantaja sahiptir; çünkü herhangi bir randomizasyon aracına ihtiyaç duymaz. Bu, yöntemin güvenilirliğini artırabilir ve daha kısa süre içinde uygulanmasını sağlar. Ancak, uygulamada anlaşılabilirlik sorunları yaşanabileceği için, daha eğitimli katılımcılarla yapılan araştırmalar için daha uygun bir yöntem olarak önerilmektedir. Ayrıca, daha doğru tahminler elde edebilmek için büyük örneklem gruplarının kullanılması gerekmektedir (Tsuchiya ve ark., 2007).

Fasulye Yöntemi (Bean Method)

Fasulye Yöntemi (BM), hassas konu araştırmalarında kullanılan oldukça basit tekniktir. Renkleri farklı olan fasulyeler (ya da benzeri nesneler) içeren iki kavanoz katılımcılara verilir: biri büyük, diğeri küçük. Kavanozlar arasındaki fasulye miktarı, katılımcıların kavanozlar arasındaki fasulye hareketlerini gözlemleriyle tespit edilemeyecek kadar olmalıdır. Katılımcılara, eğer cevabı evet ise, belirli bir renkteki fasulyeyi küçük kavanozdan büyük kavanoza aktarmaları, hayır cevabı verdiklerinde ise farklı renkte bir fasulye taşımaları söylenir. Tüm katılımcılar uygulamayı tamamladıktan sonra, fasulye bileşimindeki değişiklikler analiz edilerek hassas sorunun yaygınlığı tahmin edilir (Lau ve ark., 2011). HIV ile ilgili risk davranışlarını belirlemek amacıyla yapılan bir ankette, fasulye yöntemi uygulandıktan sonra hassas davranış oranı, yüz yüze yapılan risk davranış anketine göre daha yüksek bulunmuştur (Lau ve ark., 2011).

Fasulye yöntemi, randomize cevap yönteminden (RRT) farklıdır çünkü katılımcılardan doğrudan hassas soruyu cevaplamaları istenmez. Bu yöntem grup ortamlarında uygulanır ve katılımcılar yanıtlarının izlenebileceğinden endişe edebilirler. Teknik olarak basit olan bu yöntem, katılımcılar tarafından genellikle iyi karşılanmaktadır. Fasulye yöntemi, yüksek hassasiyet gerektiren davranışların tahmininde etkili bir şekilde kullanılabilir. Bu yöntem, seçilen örneklemden topluma ait bir tahminin ortaya çıkmasından ziyade, doğru yaygınlık hesaplamasına olanak tanır (Lau ve ark., 2011).

Fasulye yönteminin adımları, Şekil 4'teki örnekteki gibi, ilk olarak "İlk cinsel deneyiminizde vajinismus problemi yaşadınız mı?" sorusuyla hassas bir soru sorulmasıyla başlar. Katılımcıya iki kavanoz verilir ve verilen talimatları görüşmecinin göremeyeceği bir şekilde uygulaması istenir. Eğer cevabı evet

ise, büyük kavanozdan siyah olmayan bir fasulyeyi küçük olan kavanoza aktarması istenir; hayır cevabında ise siyah bir fasulye taşımaları beklenir. Fasulye yönteminde hassas sorunun tahmini formül 5'te verilmiştir:

$$\pi_{BM} = \frac{\text{Büyük kavanozdan taşınan siyah dışında diğer fasulyelerin sayısı}}{\text{Büyük kavanozdan taşınan bütün fasulyelerin sayısı}} \quad (5)$$

Şekil 4'te fasulye yöntemine ait bir diyagram çizilmiştir.



Şekil 4. Fasulye yönteminin algoritması

Gruplandırılmış Cevap Yöntemi (Grouped Answer Method)

Gruplandırılmış cevap yöntemi, 1990 yılında kayıtsız veya yasadışı göçü tahmin etmek amacıyla geliştirilmiştir (GAO, 1999; Droitcour ve Larson, 2002). Bu yöntemde, katılımcılara hassas bir soru doğrudan yöneltilmez. Bunun yerine, katılımcılardan belirli cevap seçenekleri arasından bir tercih yapmaları istenir. Ancak, bu yöntemde cevaplar "kutular" ya da setler halinde gruplanır. Katılımcılar iki gruba ayrılır ve her gruba farklı bir flaş kart verilir: Kart-1 veya Kart-2. Kart-1'de, içinde hassas da bulunan beş soru iki kutuda gruplanmaktadır, Kart-2'de ise aynı sorular farklı bir şekilde gruplanmaktadır (Droitcour ve Larson, 2002). Tablo 1'de görüldüğü gibi, hassas soru her iki kart içeriğinde Kutu-B'de yer alır. Katılımcılardan, cevaplarına dayalı olarak Kutu-A veya Kutu-B'yi seçmeleri istenir. Genellikle, çoğu katılımcının hassas soruyu içermeyen kutuyu seçmesi beklenir. Ancak, ikinci grupta soru grupları değiştirilmiştir. Her katılımcı kesinlikle bir kategoriye aittir. Gruplandırılmış cevap yöntemi, iki grup katılımcı tarafından verilen cevaplar ile hassas sorunun yaygınlığını dolaylı yoldan tahmin etmeye olanak tanır, ancak hiçbir katılımcıya hassas sorunun yanıtı açıklanmaz (Droitcour ve Larson, 2002).

Tablo 1'deki gibi, gruplandırılmış cevap yönteminde sorulan sorular arasında "idrar yolu enfeksiyonu yaşamadım", "Kadın doğum muayenesinde

özel muayene tercih ederim", "Doğum kontrol haplarını kullandım", "Düzenli bir şekilde Pap smear testi yaptırıyorum" gibi hassas olmayan sorular bulunurken, hassas soru olarak "İlk cinsel deneyimimde vajinismus problemi yaşadım" yer almaktadır. Bu sorular, her iki kartın kutularına dağıtılır. Formül 6'da gruplandırılmış cevap yönteminde kullanılan formül verilmiştir.

$$\pi_{GAT} = \pi_{A1} - \pi_{B2} \quad (6)$$

π_{GAT} : tahmin edilen hassas davranış oranı,

π_{A1} : kart 1'de A yanıtı oranı,

π_{B2} : kart 2'de B yanıtı oranı.

Tablo 1'de gruplandırılmış cevap yöntemine ait kartlar verilmiştir.

Tablo 1. Gruplandırılmış cevap yönteminin kartları

Kart 1	A	- İdrar yolu enfeksiyonu yaşamadım. - Kadın doğum muayenesinde özel muayene tercih ederim. - Doğum kontrol haplarını kullandım.
	B	- Düzenli bir şekilde Pap smear testi yaptırıyorum. - İlk cinsel deneyimimde vajinismus problemi yaşadım
Kart 2	A	- Doğum kontrol haplarını kullandım. - Düzenli bir şekilde Pap smear testi yaptırıyorum.
	B	- İdrar yolu enfeksiyonu yaşamadım. - Kadın doğum muayenesinde özel muayene tercih ederim. - İlk cinsel deneyimimde vajinismus problemi yaşadım.

Çapraz Model (Crosswise Model)

Çapraz modeli (CWM), hassas soru yöntemlerinin bir parçası olarak öne sürülmüştür. Bu modelin temel amacı, katılımcıları harici bir randomizasyon sürecine (örneğin zar atma) gerek kalmadan, dolaylı bir şekilde hassas davranışların yaygınlığını tahmin etmektir. Çapraz yöntemin, diğer yöntemlere kıyasla kolay ve daha anlaşılabilir bir sorgulama tekniği olduğu vurgulanmıştır (Yu ve ark., 2008). Bu modelde, cevap seçeneklerinin hiçbiri katılımcının hassas bir özelliğe sahip olup olmadığını doğrudan ortaya koymaz, bu da katılımcıların şüpheye düşme riskini ortadan kaldırır (Jann ve ark., 2012).

Çapraz yöntemi ile doğrudan sorgulama yöntemlerini karşılaştıran çalışmalar, çapraz modelinin daha yüksek yaygınlık tahminleri sağladığını göstermiştir (Coutts ve ark., 2011; Jann ve ark., 2012). Son dönemde yayınlanan iki çalışmada, kontrol grubu olmadan çapraz yöntemi uygulanmış ve benzer sonuçlar elde edilmiştir (Eslami ve ark., 2013; Vakilian ve ark., 2014). Örneğin, Sırbistan'daki küçük ve orta ölçekli firmalarda vergi

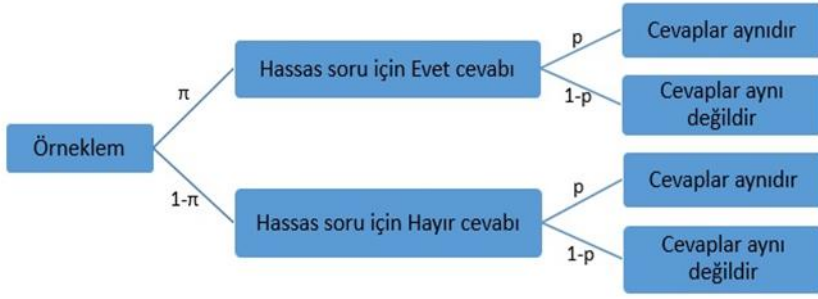
kaçakçılığının yaygınlığı, çapraz yöntemi kullanılarak incelenmiş ve doğrudan sorulan sorulara kıyasla daha yüksek tahminler elde edilmiştir (Kundt ve ark., 2013). Nakhaee ve ark. (2013) sporcular arasında anabolik steroid kullanımının yaygınlığı araştırmak için çapraz yöntemi kullanıldığında doğrudan sorulan sorulara göre iki kat daha fazla çıkmıştır. Jann ve ark. (2012), çapraz yöntemin sosyal olarak kabul edilebilirlik yanlılığını azaltmada başarılı olduğunu belirtmiştir.

Çapraz modelin uygulanması sırasında, katılımcılara iki soru yöneltilir, bu sorulardan biri hassas soru ve diğeri hassas soru değildir. Daha sonra katılımcılardan bu sorulara verdiği cevapların "aynı mı" yoksa "farklı mı" olduğunu belirtmesi istenir. Sorular birbiriyle doğrudan ilişkili değildir. Örneğin, hassas soru olarak "İlk cinsel deneyiminizde vajinismus problemi yaşadınız mı?" sorulabilirken, ilişkisiz soru olarak "Doğum gününüz Ocak, Şubat veya Mart ayında mıdır?" gibi bir soru yöneltilebilir. Katılımcılar, her iki soruya da "hayır" veya "evet" cevabı verirlerse "A" kutusunu, yalnızca bir soruya "evet" cevabı verirlerse "B" kutusunu işaretler (Jann ve ark., 2012). Hassas olmayan sorunlar için gerçekleşme olasılığı 0,50 değerinden farklı olmalıdır. Örnekte verilen hassas olmayan sorunun olasılığı 0,25'tir. Çapraz yöntemine ait detaylar Tablo 2'de sunulmuştur.

Tablo 2. Çapraz yönteminin kartı

Soru 1: Doğum gününüz Ocak, Şubat veya Mart ayında mıdır?	
Soru 2: İlk cinsel deneyiminizde vajinismus problemi yaşadınız mı?	
A veya B birini seçiniz.	A Her iki soruya cevabım aynıdır
	B Sorulara cevabım farklıdır

Şekil 5'te Çapraz modeli için bir diyagram çizilmiştir. π hassas soruya verilen "Evet" cevaplarının oranını belirtir; p , hassas olmayan sorunun bilinen olasılığını gösterir.



Şekil 5. Çapraz modeline ait ağaç diyagramı

Araştırmacı, hassas olmayan bir soruya "Evet" yanıtı verilme olasılığını önceden bilmektedir.

Katılımcılar, her iki soruya verdikleri cevaplar aynı olduğunda (ikisi de "evet" veya ikisi de "hayır") "A" seçeneğini, cevaplar farklı olduğunda ise (biri "evet" diğeri "hayır") "B" seçeneğini işaretler. Tüm cevaplar toplandıktan sonra, hassas sorunun yaygınlığı, çapraz yöntemi kullanılarak Formül 7 ile hesaplanır.

$$\pi_{CWM} = \frac{\lambda + p - 1}{2p - 1} \quad p \neq 0,5 \quad (7)$$

π_{CWM} : tahmin edilen hassas davranışın oranı,

λ : A'yı seçen deneklerin oranı,

p: hassas olmayan sorunun olasılığıdır.

Formül 8'te Çapraz yöntemi için hassas soru yaygınlığının varyansı verilmiştir:

$$Var(\pi_{CVM}) = \frac{\lambda(1-\lambda)}{n(2p-1)^2} \quad (8)$$

Formül 7, hassas olan ve olmayan soruların birbirinden bağımsız olması temel alınarak geliştirilmiştir. Hassas soru için " π " olasılığı, hassas olmayan soru için ise " p " olasılığı kullanılmıştır. Tablo 3'te gösterildiği gibi, bir katılımcının hem hassas hem de hassas olmayan soruya "hayır" cevabı verme olasılığı, bu iki sorunun birbirinden bağımsız olması nedeniyle $(1-p)(1-\pi)$ şeklinde hesaplanır. Diğer olasılıklar da benzer şekilde tabloda belirtilmiştir. Çapraz yönteminde, hassas olan ve olmayan soruların birleşik cevapları karmaşık bir şekilde değil, yalnızca toplanan "evet" ve "hayır" cevaplarının sayısı üzerinden belirlenir. "A" seçeneğine verilen "evet" cevaplarının toplamı (λ), Formül 9 ile ifade edilir (Yu ve ark., 2008).

$$\lambda = (1-p)(1-\pi) + \pi p \quad (9)$$

Tablo 3. Çapraz yöntemin kontenjans tablosu

Hassas soru		Hassas olmayan soru		
		Hayır	Evet	
	Hayır	$(1-p)(1-\pi)$	$p(1-\pi)$	$1-\pi$
	Evet	$(1-p)\pi$	$p\pi$	π
		$1-p$	p	

Her iki soruya da "evet" veya her ikisine de "hayır" cevabı verildiğinde "A" seçeneği işaretlenir ve bu durumda elde edilen sayı λ olarak ifade edilir. Denklemden λ ve p değerleri bilindiğinden, bilinmeyen değişken olan π (hassas sorunun olasılığı) için çözüm yapıldığında, çapraz yöntemine ait formül ortaya çıkar.

Çapraz modelin uygulanmasında, düşük yaygınlık (prevalans) ve küçük örneklem boyutları söz konusu olduğunda, çapraz model için Uludağ düzeltmesinin kullanılması önerilir (Ahmadian ve Ercan, 2023).

SONUÇ

Hassas konular üzerine yapılan araştırmalarda, randomize cevap yöntemi oldukça etkili sonuçlar vermesine rağmen, karşılıklı görüşmeler ve randomize bir cihaz gerektirmesi nedeniyle uygulaması daha zorlu görülmektedir. Aday yöntemi ise, incelenen hassas konunun yaygınlığının düşük olduğu durumlarda kullanılması tavsiye edilen bir yöntemdir. Hassas konuların yaygınlık düzeyinin yüksek olduğu durumlarda, aday yöntemi, diğer veri toplama yaklaşımlarına kıyasla daha düşük bir performans sergilemektedir.

Buna karşılık, eşsiz sayım ve gruplandırılmış cevap yöntemleri, özellikle geniş örneklem büyüklüklerinde benzer biçimde tatmin edici sonuçlar üretmektedir.

Çapraz model ise, diğer yöntemlere oranla daha yüksek düzeyde başarı sağlamanın yanı sıra, randomize cevap tekniğinde olduğu gibi birebir görüşme ya da rastgeleleştirme cihazlarının kullanımına gerek duymaz.

Kaynaklar

Kitap

- Bacanlı, H. (2006). *Gelişim ve öğrenme*. Nobel Yayın Dağıtım.
- Brace, Ian. (2018). *Questionnaire design: How to plan, structure and write survey material for effective market research*. Kogan Page Publishers.
- Davis, J.A; Smith, T.W. (2005). *General social surveys, 1972–2004: Cumulative codebook*. National Opinion Research Center.
- Enzmann, D. (2017). *Die Anwendbarkeit des Crosswise-Modells zur Prüfung kultureller Unterschiede sozial erwünschten Antwortverhaltens*. In *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung*. Springer.
- Fishburne, P.M; Abelson, H.I; Cisin, I.H. (1980). *National survey on drug abuse: Main findings 1979*. National Institute on Drug Abuse, Office of Medical and Professional Affairs.
- G.A.O. (1999). *Survey methodology: an innovative technique for estimating sensitive survey items*. Washington, DC.
- Hubbard, M.L; Casper, R.A; Lessler, J.T. (1989). *Respondents' reactions to item count lists and randomized response*. Proceedings of the Survey Research Section', American Statistical Association, Washington, DC, 544-448.
- Miller, J.D. (1985). *The nominative technique: A new method of estimating heroin prevalence*. NIDA Research Monograph.
- Peker, M.R. (2000). *Gelişim psikolojisine genel bakış*. Akınoğlu matbaacılık.
- Sapsford, R; Jupp, V(Eds). (1996). *Data collection and analysis*. Sage.
- Sirken, M.G. (1975). *Network surveys of rare and sensitive conditions*. Advances in Health Survey Research Methods NCHSR Research Proceedings Series Washington, DC: Supt of Docs, US Govt Print Off.
- Tourangeau, R; Rips, L.J; Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Dergi Makalesi

- Ahmadian, R; Ercan, I.(2023). Evaluating the sensitive question methods; recommended Uludağ Adjustment for the Crosswise Model. *Communications in Statistics-Simulation and Computation* 52 (12), 5759-5772
- Barnett, J. (1998). Sensitive questions and response effects: an evaluation. *Journal of Managerial Psychology* 13: 63-76. <https://doi.org/10.1108/02683949810369138>.
- Blair, G; Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis* 20: 47-77. <https://doi.org/10.1093/pan/mpr048>.
- Blank, S.G; Gavin, M.C. (2009). The randomized response technique as a tool for estimating non-compliance rates in fisheries: a case study of illegal red abalone (*Haliotis rufescens*) fishing in Northern California. *Environmental Conservation* 36: 112-119. <https://doi.org/10.1017/S037689290999004X>.
- Böckenholt, U; VanderHeijden, P.G. (2007). Item randomized-response models for measuring noncompliance: Risk-return perceptions, social influences, and self-protective responses. *Psychometrika* 72: 245. <https://doi.org/10.1007/s11336-005-1495-y>.
- Coutts, E; Jann, B. (2011). Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods & Research* 40: 169-193. <https://doi.org/10.1177/0049124110390768>.
- Dalton, D.R; Wimbush, J.C; Daily, C.M. (1994). Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior. *Personnel Psychology* 47: 817-829. <https://doi.org/10.1111/j.1744-6570.1994.tb01578.x>.
- Droitcour, J; Caspar, R.A; Hubbard, M.L. (2004). The item count technique as a method of indirect questioning: A review of its development and a case study application. *Measurement Errors in Surveys*, 185-210. <https://doi.org/10.1002/9781118150382.ch11>.
- Droitcour, J.A; Larson, E.M. (2002). An innovative technique for asking sensitive questions: The three-card method. *Bulletin of Sociological Methodology. Bulletin de Méthodologie Sociologique* 75: 5-23. <https://doi.org/10.1177/075910630207500103>.
- Ercan, I; Kan, I. (2004). Ölçeklerde güvenilirlik ve geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30(3), 211-216.

- Eslami, B; Sundin, Ö; Macassa, G. (2013). Anxiety, depressive and somatic symptoms in adults with congenital heart disease. *Journal of psychosomatic research* 74: 49-56. <https://doi.org/10.1016/j.jpsychores.2012.10.006>.
- Fisher, R.J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research* 20: 303-315. <https://doi.org/10.1086/209351>.
- Fu, H; Darroch, J.E; Henshaw, S.K. (1998). Measuring the extent of abortion underreporting in the 1995 National Survey of Family Growth. *Family Planning Perspectives*, 128-138. <https://doi.org/10.2307/2991627>.
- Hoffmann, A; Diedenhofen, B; Verschuere, B. (2015). A strong validation of the crosswise model using experimentally-induced cheating behavior. *Experimental Psychology*. <https://doi.org/10.1027/1618-3169/a000304>.
- Hoffmann, A; Musch, J. (2016a). Assessing the validity of two indirect questioning techniques: A Stochastic Lie Detector versus the Crosswise Model. *Behavior Research Methods* 48: 1032-1046. <https://doi.org/10.3758/s13428-015-0628-6>.
- Hoffmann, A; Musch, J. (2019b). Prejudice against women leaders: Insights from an indirect questioning approach. *Sex Roles* 80: 681-692. <https://doi.org/10.1007/s11199-018-0969-6>.
- Hox, J; Lensvelt-Mulders, G. (2004). Randomized response analysis in Mplus. *Structural Equation Modeling* 11: 615-620. https://doi.org/10.1207/s15328007sem1104_6.
- Höglinger, M; Jann, B; Diekmann, A. (2016). Sensitive questions in online surveys: An experimental evaluation of different implementations of the randomized response technique and the crosswise model. In *Survey Research Methods* Vol. 10, No. 3, pp. 171-187. <https://doi.org/10.18148/srm/2016.v10i3.6703>.
- Jann, B; Jerke, J; Krumpal, I. (2012). Asking sensitive questions using the crosswise model: an experimental survey measuring plagiarism. *Public opinion quarterly* 76: 32-49. <https://doi.org/10.1093/poq/nfr036>.
- John, F.A.S; Edwards-Jones, G; Gibbons, J.M. (2010). Testing novel methods for assessing rule breaking in conservation. *Biological Conservation* 143: 1025-1030. <https://doi.org/10.1016/j.biocon.2010.01.018>.
- Kundt, T.C; Misch, F; Nerré, B. (2017). Re-assessing the merits of measuring tax evasion through business surveys: An application of the crosswise model. *International Tax and Public Finance* 24: 112-133. <https://doi.org/10.1007/s10797-015-9373-0>.
- Lau, J.T; Yeung, N.C; Mui, L.W. (2011). A simple new method to triangulate self-reported risk behavior data—the bean method. *Sexually transmitted diseases* 38: 788-792. <https://doi.org/10.1097/OLQ.0b013e318218cc66>.
- Lee, R.M. (1993). Doing research on sensitive topics. Sage.
- Lensvelt; Mulders, G.J; Hox, J.J; VanDerHeijden, P.G. (2005). How to improve the efficiency of randomised response designs. *Quality and Quantity* 39: 253-265. <https://doi.org/10.1007/s11135-004-0432-3>.
- Mantha, B.R; Menassa, C.C; Kamat, V.R. (2016). A taxonomy of data types and data collection methods for building energy monitoring and performance simulation. *Advances in Building Energy Research* 10: 263-293. <https://doi.org/10.1080/17512549.2015.1103665>.
- Nakhaee, M.R; Pakravan, F; Nakhaee, N. (2013). Prevalence of use of anabolic steroids by bodybuilders using three methods in a city of Iran. *Addiction and Health* 5: 77.
- Nuno, A; Bunnefeld, N; Naiman L.C. (2013). A novel approach to assessing the prevalence and drivers of illegal bushmeat hunting in the Serengeti. *Conservation Biology* 27: 1355-1365. <https://doi.org/10.1111/cobi.12124>.
- Nuno, A; Freya, A.V; S.t. John. (2015). How to ask sensitive questions in conservation: A review of specialized questioning techniques. *Biological Conservation* 189: 5-15. <https://doi.org/10.1016/j.biocon.2014.09.047>.
- Ozan, C. ve Köse, E. (2014). Eğitim programları ve öğretim alanındaki araştırma eğilimleri. *Sakarya University Journal of Education*, 4(1), 116-136.
- Phillips, D.L; Clancy, K.J. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology* 77(5), 921-940. <https://doi.org/10.1086/225231>.
- Rasinski, K.A; Willis, G.B; Baldwin, A.K. (1999). Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and*

- Cognition* 13: 465-484. [https://doi.org/10.1002/\(SICI\)1099-0720\(199910\)13:5<465::AID-ACP609>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-0720(199910)13:5<465::AID-ACP609>3.0.CO;2-Y).
- Razafimanahaka, J.H; Jenkins, R.K; Andriafidison, D. (2012). Novel approach for quantifying illegal bushmeat consumption reveals high consumption of protected species in Madagascar. *Oryx* 46: 584-592. <https://doi.org/10.1017/S0030605312000579>.
- Schnapp, P. (2019). Sensitive Question Techniques and Careless Responding: Adjusting the Crosswise Model for Random Answers. *Methods, Data, Analyses*, 13(2), 13. <https://doi.org/10.12758/mda.2019.03>.
- Sheppard, S.C; Earleywine, M. (2013). Using the unmatched count technique to improve base rate estimates of risky driving behaviours among veterans of the wars in Iraq and Afghanistan. *Injury Prevention* 19: 382-386. <https://doi.org/10.1136/injuryprev-2012-040639>.
- Simon, P; Striegel, H; Aust, F. (2006). Doping in fitness sports: estimated number of unreported cases and individual probability of doping. *Addiction* 101: 1640-1644. <https://doi.org/10.1111/j.1360-0443.2006.01568.x>.
- Simsek, S. ve Demirel, İ. (2020). Ortaokul Türkçe ders kitaplarının anlatıma dayalı etkinlikler yönünden karşılaştırılması. *Kastamonu Eğitim Dergisi*, 28(5), 2125-2135. DOI: 10.24106/kefdergi.731886.
- Singer, E; Hippler, H.J., Schwarz, N. (1992). Confidentiality assurances in surveys: Reassurance or threat?. *International Journal of Public Opinion Research* 4(3), 256-268. <https://doi.org/10.1093/ijpor/4.3.256>.
- Singer, E; Von, Thurn, D.R; Miller, E.R. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly* 59: 66-77. <https://doi.org/10.1086/269458>.
- Sirken, M.G. (1972). Stratified sample surveys with multiplicity. *Journal of the American Statistical Association* 67: 224-227. <https://doi.org/10.2307/2284732>.
- Smith, J; Noble, H. (2014). Bias in research. *Evidence-Based Nursing*, 17(4), 100-101. <https://doi.org/10.1016/j.annemergmed.2004.07.454>
- Solomon, J; Jacobson, S.K; Wald, K.D. (2007). Estimating illegal resource use at a Ugandan park with the randomized response technique. *Human Dimensions of Wildlife* 12: 75-88. <https://doi.org/10.1080/10871200701195365>.
- Sudman, S; Sirken, M.G; Cowan, C.D. (1988). Sampling rare and elusive populations. *Science* 240: 991-996. <https://doi.org/10.1126/science.240.4855.991>.
- Tourangeau, R; Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin* 133: 859. <https://doi.org/10.1037/0033-2909.133.5.859>.
- Tsuchiya T, Hirai Y, Ono S. (2007). A study of the properties of the item count technique. *Public Opinion Quarterly* 71: 253-272. <https://doi.org/10.1093/poq/nfm012>.
- Vakilian, K; Mousavi, S.A; Keramat, A. (2014). Estimation of sexual behavior in the 18-to-24-years-old Iranian youth based on a crosswise model study. *BMC research notes* 7: 28. <https://doi.org/10.1186/1756-0500-7-28>.
- Vieira, E.M. (2009). Frequency and characteristics of induced abortion among married and single women in São Paulo, Brazil. *Cadernos de Saúde Pública* 25: 179-187. <https://doi.org/10.1590/S0102-311X2009000100019>.
- Xing Yang and Nabendu Pal. (2010). Estimation of a population size through capture-mark-recapture method: a comparison of various point and interval estimators. *Journal of Statistical Computation and Simulation*. 80 (3): 335-354. <https://doi.org/10.1080/00949650802635165>.
- Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60: 63-69. <https://doi.org/10.1080/01621459.1965.10480775>.
- Yu, J.W; Tian, G.L; Tang, M.L. (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika* 67: 251. <https://doi.org/10.1007/s00184-007-0131-x>.

Giriş

Anket araştırmaları bir popülasyon ya da örnekleme, çalışmada dikkate alınan değişkenler arasındaki ilişki, sıklık ve dağılım hakkında bilgi edinmek için yapılır. Anket, katılımcıların kendi beyanlarını veya yanıtlarını değerlendirmek amacıyla anket ya da mülakatlarda kullanılan ve "madde" olarak adlandırılan bir dizi soru veya ifadeden oluşur. Yazılı veya sözlü olarak uygulanabilen anket araştırmaları, bireyleri veya grupları ölçmek, tanımlamak ya da karakterize etmek için anketlerin kullanımını içerir. Anket araştırma tasarımları:

- Tanımlayıcı,
- Boylamsal,
- İlişkisel/karşılaştırmalı

olarak sınıflandırılmaktadır. Yaygın olarak kullanılan ve anket araştırmasının bir parçası olarak kabul edilen bir araştırma yöntemi de Delphi yöntemidir (Shang, 2023). Delphi yöntemi, politika geliştirme amacıyla grup görüşünü ortaya çıkarmak için kullanılan geleneksel yöntemlerin sınırlılıkları nedeniyle geliştirilmiştir ve istatistiksel model tabanlı kanıtların bulunmadığı, bilginin belirsiz ve eksik olduğu ve uzmanlara ait yargının bireysel görüşten daha iyi olduğu sorunlu alanlarda pratik bir yaklaşımdır. Fikir birliği sağlanamadığında, belirsizlik bulunduğunda veya yeterli kanıt olmadığında tercih edilebilecek uygun bir yöntemdir. İstatistiksel ya da model tabanlı prosedürlere meydan okumak için değil, daha ziyade istatistiksel yöntemlerin pratik ya da mümkün olmadığı durumlarda kullanılmak üzere, bir grup uzmandan en güvenilir fikir birliğini elde etmek için tasarlanmıştır (Trevelyan ve Robinson, 2015). Literatürde Delphi yöntemi;

** Prof. Dr., Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Afyonkarahisar, Türkiye ismet.dogan@afsu.edu.tr, ORCID: orcid.org/0000-0001-9251-3564

*** Prof. Dr., Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Afyonkarahisar, Türkiye nurhandogan@hotmail.com, ORCID: orcid.org/0000-0001-7224-6091

- Bir grup bireyin karmaşık bir problemi birlikte çözmesini sağlamak amacıyla, grup içi iletişim sürecini etkin bir şekilde yönlendiren yapılandırılmış bir teknik (Trevelyan ve Robinson, 2015),
- Uzmanların kullanımı ve grup görüşü hakkında bilgi sağlayan, kontrollü geribildirim ile serpiştirilmiş bir dizi anket kullanımı (Nasa, Jain ve Juneja, 2021),
- Uzmanlardan oluşan bir panele ilgili bir konu hakkında görüşlerini sorma, toplu yanıtlarını özetleyip sunma ve bu süreci belirli sayıda tur boyunca tekrarlama prosedürü (Spranger ve ark., 2022),
- Hakkında belirsiz ve eksik bilgi bulunan karmaşık konuların uzmanlar tarafından yinelemeli ve yapılandırılmış bir süreçte değerlendirildiği grup tartışma yöntemi (Sablatzky, 2022),
- Bir alandaki uzmanlardan oluşan bir panel arasında birkaç tur sorgulama yoluyla fikir birliğine varmak için kullanılan bir anket tekniği (Shang, 2023),
- Karmaşık bir sorun üzerinde bilinçli bir grup konsensüsüne ulaşmak amacıyla uzman bilgi ve görüşlerini sistematik olarak birleştirmek için kullanılan nitel bir yöntem (Donohoe ve Needham, 2009), vb. farklı şekillerde tanımlanmaktadır.

Doğrulanmış verilerin yokluğunda grup tahmininin mantıklı olduğu finans, ekonomi, kalkınma planlaması, sağlık hizmetleri, tıp, hemşirelik, sosyal politika, turizm ve sürdürülebilirlik bilimi dâhil olmak üzere birçok disiplinde karmaşık bir sorunu bir grup bireyin kolektif olarak ele almalarını sağlamak için kullanılan yerleşik bir yöntemdir. Bireysel görüşleri toplayarak, bunları kolektif zekâ yardımıyla istatistiksel olarak şekillendirilmiş bir fikir birliğine dönüştüren bir istatistiksel yöntem haline gelmiştir (Nasa, Jain ve Juneja, 2021). Fikir birliği, bir grup üyeleri arasında ortak bir görüşün oluşması veya görüşlerin istikrara kavuşması olarak tanımlanır. Bu süreç, kontrollü geri bildirim içeren yinelemeli turlar ve uzman görüşlerinin değerlendirilmesiyle desteklenen sıralı anketler aracılığıyla sağlanır. Aynı zamanda, farklı değerleri ve deneyimleri düzenleyen bir mekanizma sunarak çeşitli bakış açılarının fikir birliğine dâhil edilmesini kolaylaştırır (Donohoe ve Needham, 2009). Belirli bir konuya dair görüşleri toplamak için kullanılan bu yöntem, birleşik zekânın bireysel muhakemeyi geliştirdiği ve uzmanların ortak görüşünü yansıttığı fikrine dayanır (Villiers ve ark., 2005). Bunu, istatistiksel grup yanıtları da dâhil olmak üzere kontrollü geri bildirimlerle serpiştirilmiş bir dizi anket

aracılığıyla yapmaya çalışır. Panelistler arasındaki anonimlik, bireysel hâkimiyetin ortaya çıkmasını önler. Ayrıca panelistlerin, diğer panelistler tarafından sunulan argümanlara dayanarak, bunu yaptıklarını kamuya açıklamadan fikirlerini değiştirmelerine olanak tanır. Bu avantajların, fikir birliğinin güvenilirliğini artırdığı kabul edilmektedir (Veugeliers ve ark., 2020). Yöntem,

- Odak grup tartışmaları gibi yüz yüze ortamlarda çözülmesi zor olan tartışmalı konularda fikir birliği oluşturmak,
- Farklı yargılara yol açan varsayımları veya bilgileri araştırmak veya ortaya çıkarmak,
- Katılımcıların deneyimleri aracılığıyla veri boşluklarını doldurmak veya modelleri doğrulamak,
- Sınırlı bir süre içinde farklı disiplinlerden veya coğrafi konumlardan girdilerin bir araya getirilmesini gerektiren karmaşık konuları ele almak, politikaları formüle etmek veya değerlendirmek,

için kullanılabilir (Mukherjee ve ark., 2015). Delphi çalışmalarının hedefleri:

- Bir grup uzmandan bilgi toplamak ve damıtmak,
- Geleceğe ilişkin bir öngörüü destekleyecek kesin bilgilerin bulunmadığı karmaşık konularda fikir birliğine varmak ve/veya yargıda bulunmak,
- Fikirleri güvenilir ve yaratıcı bir şekilde araştırmak,
- Kritik karar verme süreçleri için uygun bilgiler üretmek,

olarak belirtilmektedir (Naisola-Ruiter, 2022). Delphi çalışmasının temel mantığı, kolektif grup yanıtlarının bireysel yanıtlardan daha üstün olduğu varsayımına dayanır. Bu teori, tek bir bireyin en iyi fikirlere sahip olmayabileceğini, ancak kolektif değerlendirme sürecinde en güçlü fikirlerin öne çıkma eğiliminde olduğunu savunur (Sablatzky, 2022). Bu yöntem:

- Etik, siyasi, hukuki veya sosyal ikilemlerin ekonomik veya teknik ikilemlere baskın geldiği,
- Mali, coğrafi veya zaman kısıtlamaları ve/veya demokratik katılımı ilgili endişeler nedeniyle yüz yüze temasın mümkün olmadığı veya arzu edilmediği,

- Analitik tekniklerin kullanılmadığı, kesin bilginin olmadığı ve grup konsensüsüyle yönetilen öznel görüşlerin toplanmasının gerekli olduğu,
- İlgili uzmanların farklı alanlarda ve/veya mesleklerde olması ve doğrudan iletişim halinde olmadığı,

durumlarda son derece uygundur (Donohoe ve Needham, 2009).

Tarihçe

1950'lerden bu yana profesyonel kılavuzlar geliştirmek için yaygın olarak kabul gören bir araç olarak başlangıçta askeri tahminler için geliştirilen bu yöntem, kamu politikalarından sağlık hizmetlerine kadar çeşitli disiplinlerde yaygın olarak kullanılmaktadır. Ampirik kanıtların yetersiz veya çelişkili olduğu durumlarda, yapılandırılmış anketler yoluyla yürütülen ve grup kolaylaştırıcılığını içeren yinelemeli bir süreçten oluşmaktadır (Shi ve ark., 2024). Orijinal Delphi yöntemi, 1950'lerde Rand Corporation'dan Olaf Helmer, Norman Dalkey ve Nicholas Rescher tarafından, teknolojinin savaş üzerindeki etkisini öngörmek amacıyla geliştirilmiştir. O günden bu yana, sağlık hizmetleri, pazarlama, eğitim, bilgi sistemleri, ulaşım ve tamamlayıcı-alternatif tıp gibi alanlarda rehberler oluşturmak ve bir müdahalenin temel bileşenlerini belirlemek için kullanılmaktadır (Trevelyan ve Robinson, 2015). Belirsizlikle mücadele etmek, insan deneyiminin ayrılmaz bir parçasıdır ve tarih boyunca araştırmacılar, teorisyenler ve filozoflar, bu alandaki insan yetkinliğini geliştirmeye çaba göstermiştir. Antik Yunan'da krallar ve generaller, devletleri ve kariyerleri adına belirsizlikle başa çıkabilmek için bir kâhine, Pythia'ya ya da daha yaygın olarak bilinen adıyla Delphi Kâhini'ne danışırldı. Kâhin, kehanetlerinin doğrudan Yunan tanrısı Apollo'dan geldiği rivayet edilen ünlü bir figürdü. 2500 yıldan fazla bir süre sonra, Soğuk Savaş'ın zirvesinde, ABD Hava Kuvvetleri de askeri belirsizliklerle karşı karşıyaydı ve bu konuda RAND Corporation adlı düşünce kuruluşuna danıştı. RAND Corporation'daki araştırmacılar, antik kâhinin adını kullanarak Delphi yöntemini geliştirdi. Bu yöntem, birkaç askeri uzmanın işe alınmasını, her birine olası bir Sovyet saldırısının ihtimali, sıklığı ve şiddeti hakkında sorular yöneltilmesini ve uzmanlardan anonim geri bildirim vermelerinin istenmesini içeriyordu. Zamanla, bu yöntem gizliliği kaldırılarak askeri alanın ötesine geçmiş ve çeşitli sağlık bilimlerinde de kullanılmaya başlanmıştır (Shang, 2023). RAND Corporation, 1950'ler ve 1960'larda Soğuk Savaş döneminde ulusal güvenlik tehditlerine karşı hazırlık amacıyla bu yöntemi geliştirmiştir. Zamanla, pazarlama, reklamcılık, eğitim ve tıp gibi birçok farklı alanda araştırma yöntemi olarak kullanılmaya başlanmıştır. 1970'lerden itibaren

kütüphane ve bilgi bilimi gibi diğer disiplinler de bu yöntemi araştırmalarına entegre etmiştir (Sablatzky, 2022).

Özellikleri

Tipik bir Delphi çalışması, seçilmiş uzmanlara gönderilen birkaç aşamalı yazılı anketlerden oluşur. Her tur sonunda veriler toplanır, analiz edilir ve uzmanlara özet bir rapor sunulur. Uzmanlar bu raporu değerlendirerek diğer katılımcıların görüşlerine katılıp katılmadıklarını belirtirler. Bu süreç, fikir birliği sağlanana kadar tekrarlanır (Sablatzky, 2022). Delphi'nin genel olarak:

- Anonimlik,
- Yineleme,
- Kontrollü geri bildirim,
- Grup yanıtının istatistiksel toplanması,
- Uzman girdisi,

şeklinde belirtilen 5 temel özelliği vardır (Landeta, 2006). Başarılı bir Delphi anketinin temelini, fikir birliğine ulaşmayı sağlayan yinelemeli süreç ve kontrollü geri bildirim oluşturur. Delphi yönteminin gücü, anket turlarında panelistlerin anonim kalabilmesi, geri bildirimin yapılandırılmış şekilde sunulması ve tekrar eden tartışmaların yürütülmesidir. Anonim anket turları, yüz yüze veya grup içi tartışmalara kıyasla baskın görüşlerin etkisini azaltma ve grup uyumunu sınırlama açısından avantaj sağlar. Katılımcılar, belirsiz veya çözüme ulaşmamış konularda anonim görüşlerini paylaşırken kendilerini daha rahat hissederler. Ancak, anonim Delphi turlarında maddelerin yorumlanması bazen kritik bir mesele haline gelebilir ve fikir birliği sürecini etkileyebilir (Nasa, Jain ve Juneja, 2021). Anonimlik, katılımcıların araştırmacılar dışında çalışmaya kimlerin dâhil olduğunu bilmemesi gerektiği anlamına gelir ve anonimleştirilmiş anketler kullanılarak elde edilir ve benzersiz bir gizli kod atanarak geliştirilebilir. Ayrıca, birbirlerinden anonim olmak, potansiyel olarak kendilerinden daha üstün veya daha uzman olarak görülen diğer katılımcıların haksız etkisini önler, görüşlerin özel olarak ifade edilmesini ve değiştirilmesini sağlar. Tekrarlar, katılımcıların önceki değerlendirmelerini gözden geçirmelerine olanak tanıyan, birden fazla turdan oluşan geri bildirim sürecini ifade eder. Delphi yönteminin bir diğer temel özelliği ise kontrollü geri bildirimdir. "Kontrollü" olarak adlandırılmasının nedeni, moderatörün verilen yanıtlar ve açık yorumlar doğrultusunda geri bildirim sürecini yönlendirmesidir. Bu süreç, katılımcıların anonim

meslektaşlarının yanıtlarından haberdar olmasını sağlarken, kolektif görüş ve yargıların yalnızca baskın sesler tarafından değil, tüm katılımcılar tarafından şekillendirilmesine imkân tanır. Kontrollü geri bildirim ile yineleme, katılımcılar arasında iletişim kurulmasına, bakış açılarının paylaşılmasına ve katılımcıların görüşlerini değiştirmelerine olanak tanır. İstatistiksel toplama, araştırmacı tarafından grup yanıtının istatistiksel özetinin sunulmasını içerir; bunlar genellikle merkezi eğilim ölçütleridir ve aynı zamanda nihai sonuçların istatistiksel analize tabi tutulmasına olanak tanır ve yarı nesnellik duygusuna yol açar. Grup yanıtlarının istatistiksel olarak toplanması verilerin analiz edilmesine ve yorumlanmasına olanak tanır. Her anket turundan sonra, elde edilen veriler analiz edilir ve tüm panel uzmanlarına kolayca yorumlanabilir bir biçimde sunulur. Uzman girdisi, her katılımcının araştırılan alanda bilgi sahibi olduğu anlamına gelir (Trevelyan ve Robinson, 2015; Niederberger ve Spranger, 2020; Spranger ve ark., 2022; Shang, 2023).

Avantaj ve Dezavantajları

Delphi yöntemi 1950'lerin başından bu yana ordu, iş dünyası, eğitim, sosyal bilimler ve sağlık hizmetleri de dâhil olmak üzere çok çeşitli alanlarda kullanılmaktadır. Tahmin yapma, sorunları tanımlama ve önceliklendirme, sıralama, politika oluşturma ve politika çerçevesi geliştirme gibi çeşitli karmaşık araştırma amaçları doğrultusunda kullanılabilir. Ancak, bu yöntemin esnekliği hem bir avantaj hem de bir dezavantaj olarak görülebilir. Uygulayıcılar, Delphi yöntemini kendi karar alma ve tahmin süreçlerine uyarlamaya genellikle açık, hatta bazen istekli davranırlar. Bazı durumlarda bu uyarlamalar faydalı olabilir ve yöntemin daha iyi anlaşılmasına katkı sağlayabilir. Bazı durumlarda ise keyfidir ve dolayısıyla kalite ve güvenilirliğe zarar verir (Veugelers ve ark., 2020). Delphi yönteminin,

- Delphi yöntemi, bir grup uzmanın belirli bir dizi karmaşık sorunu ele almasına ve bir dizi temel konu üzerinde fikir birliğine varmasına olanak tanır.
- Elektronik olarak uygulanan anketler, yüz yüze uzman panellerinde ortaya çıkabilecek baskın bireylerin etki riskini azaltır. Bu durum, yanıtları katılımcı grubun diğer üyelerine açıklanmadığı için bireylerin görüşlerini herhangi bir engelleme olmaksızın ifade etmelerine olanak tanır.
- Kontrollü (anketin bir dizi turda gerçekleştirilmesi ve sonuçların her turun sonunda tüm katılımcılara sunulması) geri bildirim söz konusudur.
- Grup yanıtının istatistiksel tanımı, belirli konular ve sorunlar üzerinde varılan fikir birliğinin derecesini tanımlamak için kullanılabilir. Delphi

yönteminin güvenilirliği, katılımcı grubunun büyüklüğüne ve yürütülen tur sayısına bağlı olarak belirtilen avantajları vardır (Bettcher, Sapirie ve Goon, 1998).

Delphi yöntemi, belirli bir alandaki uzmanların görüşlerini derlemek için esnek bir çerçeve sunar. Katılımcıların, diğer görüşleri değerlendirdikten sonra kendi düşüncelerini yeniden gözden geçirebilmesi, tek seferlik görüşmelerde veya odak gruplarında bulunmayan bir refleksiyon süreci sağlar. Delphi çalışmalarının temelini oluşturan uzman grupları arasındaki anonimlik, katılımcılar arasında dürüstlüğü teşvik eder ve grubun baskın veya yüksek profilli üyelerinin görüşlerine ekstra itibar verme riskini azaltır. Ancak Delphi çalışmaları, doğası gereği karmaşık ve zaman alıcı olabilir. Katılımcıların birden fazla turu tamamlaması gerekliliği, çalışmanın geçerliliğini etkileyen yüksek oranda ayrılma oranlarına yol açabilir. Katılımcıların her turda görüşlerini değiştirme veya revize etme yeteneği, hem avantaj hem de dezavantaj içeren bir durumdur. Bu süreç, onlara ek bilgiler doğrultusunda pozisyonlarını gözden geçirme ve yeniden değerlendirme imkânı sunar. Ancak bunun tersine, bu esnekliğin önyargıya yol açması ve katılımcıların çoğunluk görüşü olarak gördükleri şeye uymak için yanıtlarını değiştirmesi tehlikesi vardır. Delphi çalışmaları, fikir birliğinin ne anlama geldiği konusunda netlik olmaması nedeniyle eleştirilebilir. Delphi çalışmalarındaki esneklik ve refleksiflik seviyesine rağmen, bir uzman grubunun her konuda tam bir fikir birliğine ulaşması yine de pek mümkün değildir. Ancak, fikir birliği bir Delphi çalışmasının bir gereksinimi olduğundan, bu noktaya ne zaman ulaşılacağına dair bir yargıya varılması gerekir. İşte çalışmalar ve yazarlar arasındaki tutarsızlığın ortaya çıktığı nokta burasıdır. Önerilen fikir birliği oranı %51 ile %100 arasında değişkenlik göstermektedir. Ayrıca, bazı alanlarda fikir birliğinin, araştırma yönteminin bir parçası olarak önceden belirlenmediği de gözlemlenmiştir. Ayrıca, nesnel düzeyde bir fikir birliğinin belirlenmesi yalnızca ölçülebilir veriler toplandığında mümkündür. Bazı nitel Delphi çalışmalarında elde edilen fikir birliği, araştırmacı için daha öznel bir değerlendirme gerektirebilir ve bu da belirli önyargılara açık olmasına neden olabilir. Ayrıca, Delphi yöntemleri doğası gereği genellikle yalnızca uzman görüşlerine dayanarak bulgular üretir. Bu nedenle, bir diğer sınırlama, uzman görüşünün sağlık müdahaleleri hakkında yargılarda bulunmak için zayıf bir temel sağladığı düşünülen kanıt kalitesiyle ilgilidir. Bu durum, Delphi çalışmalarının bulgularının güvenilir veya geçersiz olduğu anlamına gelmez. Ancak, araştırmacıların, araştırma sorularının uzman görüşüyle yanıtlanmaya uygun olup olmadığını

değerlendirmeleri ya da sistematik literatür incelemesi gibi alternatif yaklaşımların daha uygun olup olmadığını göz önünde bulundurmaları gerektiğini gösterir (Barrett ve Heale, 2020). Ancak, bu yöntem birçok çalışmada öne çıkan önemli metodolojik zayıflıklardan tamamen arınmış değildir. Bunlar arasında;

- Temel bilgi kaynağı,
- Gerçeğe yaklaşmanın bir yolu olarak fikir birliğinin kullanılması,
- Yazılı ve kontrollü geri bildirimde yer alan etkileşimin sınırlı olması,
- Gruba bireysel katkı için sosyal telafi olasılığının kısıtlanması,
- Uzmanların sorumsuz eylemleri açısından anonimliğin sağladığı dokunulmazlık,
- Çalışmayı yürüten kişi tarafından ilgili manipülasyonun metodolojisinde bulunan kolaylık,
- Yöntemin doğruluğunu ve güvenilirliğini kontrol etmenin zorluğu,
- Bunu gerçekleştirmek için gereken zaman,
- Katılımcılar tarafından gereken çaba ve tahmin edilen olaylar arasındaki olası karşılıklı ilişkilerin dikkate alınmaması,

yer almaktadır. Bu metodolojik zayıflıklara ek olarak, Delphi yöntemi kendisinden ziyade, uzmanların seçiminde titizlik eksikliği, ilerleme ve ayrılmanın açıklanmaması, kötü formüle edilmiş sorular ve problemler, sonuçların yetersiz analizi vb. gibi zayıf uygulamaları nedeniyle de eleştirilmektedir (Landeta, 2006).

Türleri

Geleneksel (klasik) Delphi yaklaşımına ek olarak başka Delphi varyantları da vardır. En önemlileri;

- Gerçek Zamanlı Delphi,
- Politika Delphi,
- Argümantatif Delphi,
- Grup Delphi,
- Bulanık Delphi,
- Modifiye Delphi

- Karar Delphi
- İnternet Delphi

olarak bilinmektedir (Varndell, Fry ve Elliott, 2021; Spranger ve ark., 2022). Delphi yöntemlerinin tanımları konusunda pek çok tartışma bulunmaktadır. Bu durum, metodolojik titizlik eksikliği, araştırmaya yardımcı olacak mevcut rehberlerin azlığı ve Delphi'lerin nasıl yürütüldüğüne dair geniş çeşitlilikle ilgili olarak yönetime yönelik önemli bir eleştiri olan metodolojik karışıklığa katkıda bulunmaktadır. Klasik Delphi, Modifiye Delphi, Politika Delphi ve e-Delphi terimleri genellikle birbirinin yerine kullanılmaktadır. Ancak, bu farklı fakat ilişkili Delphi yöntemlerini kesin çizgilerle ayıran net tanımlar veya kılavuzlar henüz mevcut değildir. Günümüzde Klasik Delphi, genellikle sadece "Delphi" olarak anıldığından, bu terim büyük ölçüde eskimiş kabul edilmektedir ve en yaygın kullanılan yöntem olma özelliğini taşımaktadır. Tipik olarak 2-3 turdan oluşan Klasik Delphi'nin ilk aşamasında, büyük miktarda veri elde etmek amacıyla nitel açık uçlu anketler veya görüşmeler gerçekleştirilir. Devam eden turlarda ise, verilerin analiz edilmesi için merkezi eğilim ve dağılım ölçütlerine dayalı nicel anketler kullanılır. Modifiye Delphi inanılmaz derecede çeşitli şekillerde kullanılır ve neredeyse hiçbir Modifiye Delphi aynı şekilde yürütülmez. Genel olarak, çoğu Modifiye Delphi, genellikle ilk turda yapılan görüşmeler veya odak grupları gibi yüz yüze bir yönü içerir. Buna rağmen, Modifiye Delphi terimi, Klasik Delphi'den farklılaşan çeşitli yöntemleri kapsayan genel bir kavram gibi görünmektedir. Geleneksel Delphi yaklaşımından dikkat çekici bir ayrışma ise, fikir birliğinin hedeflenmediği ve görüş ayrılıklarının teşvik edildiği Politika Delphi'sidir. Bu nedenle, Politika Delphi yöntemi doğrudan karar verme sürecine yönelik değildir; bunun yerine, politika konularında analitik bir araç olarak kullanılır. Heterojen bir şekilde yürütölmelerine rağmen, politika Delphileri konuların formüle edildiği, katılımcılardan maddeler hakkındaki görüşlerinin sorulduğu, ancak aynı zamanda anlaşmazlıkların nedenlerinin araştırıldığı ve ardından seçeneklerin yeniden değerlendirildiği klasik Delphi ile benzer adımları içörmö eğilimindedir. Tamamen çevrimiçi ortamda gerçekleştirilen ve Klasik Delphi olarak sınıflandırılan e-Delphi, sunduğu benzersiz kolaylık, zaman ve maliyet tasarrufu ile giderek daha popüler hale gelmektedir. Ayrıca, veri yönetimini büyük ölçüde kolaylaştırması da önemli bir avantajdır. Özellikle, e-Delphi yöntemi uzmanların coğrafi konum veya saat dilimi kısıtlamalarından bağımsız olarak katılımına imkân tanıyarak, genellikle daha hızlı yanıt süreleri elde edilmesini sağlamaktadır (Shang, 2023).

Süreç

Delphi sürecinin dört aşaması bulunmaktadır. İlk aşamada araştırılan konu incelenir ve katılımcılara konuyla ilgili olduğunu düşündükleri bilgilere katkıda bulunma fırsatı verilir. İkinci aşamada ise, grubun konuya dair genel bir bakış açısı geliştirmesi amaçlanır. Eğer belirgin bir görüş ayrılığı tespit edilirse, üçüncü aşamada bu anlaşmazlığın nedenleri incelenerek farklılıkların kaynağı belirlenir. Son olarak, dördüncü aşama, elde edilen tüm bilgilerin nihai değerlendirmesini kapsar (Stitt-Gohdes ve Crews, 2004).

Tur Sayısı

Delphi çalışmalarında sıkça sorulan sorulardan biri tur sayısıdır. Tur sayısı, fikir birliğinin bir durdurma kriteri olarak kabul edilip edilmediğine ya da tur sayısının önceden belirlenip belirlenmediğine göre değişiklik gösterebilir. Klasik Delphi'de dört tur olmasına rağmen, birçok çalışma a priori üç tur kriteri belirlemiştir, ancak bazıları sadece iki tur içermektedir. Sadece iki turun dâhil edilmesinin bir sınırlaması, istikrarın teyit edilememesidir. Yıpranmanın her turda artması muhtemel olduğundan, katılımcı yorgunluğuna karşı koruma ve sonuçların anlamlı olmasını sağlamak için üç tur en uygun seçenek olarak görülmektedir (Trevelyan ve Robinson, 2015). Teorik olarak, Delphi süreci fikir birliğine varılana kadar yinelenebilir. Çoğu durumda gerekli bilgilerin toplanması ve fikir birliğine varılması için üç yineme yeterli olmaktadır. Ancak geleneksel bir Delphi çalışması tasarımı en az dört tur içerir (Hsu ve Sandford, 2007; Nasa, Jain ve Juneja, 2021).

1.Tur: Delphi süreci genellikle açık uçlu bir anketle başlar. Bu anket, Delphi katılımcılarından belirli bir konu hakkında ayrıntılı bilgi toplamak için temel bir araç görevi görür. Katılımcılardan gelen yanıtlar toplandıktan sonra, araştırmacılar bu verileri analiz ederek daha yapılandırılmış bir soru formuna dönüştürmelidir. Hazırlanan bu anket, ikinci turda veri toplama sürecinde kullanılacak ana araç haline gelir.

2.Tur: İkinci turda, Delphi katılımcılarına ilk turdan elde edilen bilgiler doğrultusunda hazırlanan yeni bir anket gönderilir. Katılımcılardan, araştırmacılar tarafından özetlenen maddeleri gözden geçirmeleri beklenir. Bu aşamada, maddeleri belirli bir öncelik sırasına göre derecelendirmeleri veya sıralamaları talep edilebilir. İkinci turun sonucunda, fikir birliği sağlanan ve görüş ayrılıklarının bulunduğu alanlar belirlenir. Bazı durumlarda, panelistlerden yaptıkları sıralamanın gerekçesini açıklamaları istenebilir. Bu

tur, fikir birliğinin oluşmaya başladığı aşama olup, katılımcıların yanıtları doğrultusunda elde edilen veriler sunulabilir.

3. Tur: Üçüncü turda, Delphi panelistlerine önceki turda araştırmacılar tarafından özetlenen maddeleri ve bunların derecelendirmelerini içeren bir anket sunulur. Katılımcılardan, önceki yargılarını gözden geçirmeleri veya fikir birliğine katılmamaları durumunda bunun nedenlerini açıklamaları istenir. Bu aşama, panelistlere hem mevcut bilgileri değerlendirme hem de maddelerin göreceli önemine dair görüşlerini daha da netleştirme fırsatı tanır. Ancak, önceki tura kıyasla fikir birliği oranında yalnızca küçük bir artış beklenir.

4. Tur: Dördüncü ve genellikle son turda, panelistlere kalan maddelerin listesi, bunların derecelendirmeleri, azınlık görüşleri ve üzerinde uzlaşılan noktalar sunulur. Bu aşama, katılımcılara fikirlerini gözden geçirmeleri için son bir fırsat tanır. Ayrıca, Delphi sürecindeki tur sayısının, araştırmacıların hedeflediği uzlaşma düzeyine bağlı olarak değiştiği ve genellikle üç ile beş tur arasında değişebileceği unutulmamalıdır.

Çoğu Delphi çalışmasındaki sonuç ölçütü, önceki (genellikle iki) turdan sonra fikir birliğini içerir. Yanıtların veya fikir birliğinin istikrarlı olup olmadığı yalnızca iki Delphi turuyla kesin olarak belirlenemez. Maddelerdeki değişiklikler veya kontrollü geri bildirim süreci, panelistlerin görüşlerini değiştirmesine yol açabilir. Ancak, bu değişiklikler her zaman istikrarlı sonuçlar doğurmayabilir. Bu nedenle, sonuçların kararlılığını değerlendirmeden önce tur sayısını sabitlemek, istatistiksel sağlamlık açısından bir uzlaşma niteliği taşır. Modifiye Delphi, keyfi olarak önceden belirlenmiş iki veya üç anket turunu bir kapanış ölçütü olarak kullanır. Ancak, Delphi çalışmalarındaki modifiye terimi tutarsızdır ve genel olarak kabul görmüş bir ölçütten yoksundur. Modifiye Delphi metodolojisinin tek ortak özelliği, yönetim grubunun fikir birliğine ulaşmak için aktif çaba göstermesidir. Bu süreçte yönlendirme grubu, ilgili literatürü sistematik olarak tarar ve açık uçlu bir yaklaşımdan ziyade, ilk Delphi turlarında panelistler arasında doğrudan fikir birliği sağlamaya odaklanır. Grup ayrıca her turdan sonra sonuçları gözden geçirir ve fikir birliğine varılan maddeler sonraki turlarda çıkarılır, ancak kontrollü geri bildirimle rağmen fikir birliğine varılamayan maddeler de çıkarılabilir. Ancak, yönlendirme grubunun bu aktif katılımı, üyelerin görüşleri nedeniyle önyargıya yol açabilir (Nasa, Jain ve Juneja, 2021). Delphi çalışmalarında, katılımcıların süreçten ayrılması önemli bir sorun olabilir ve bunu en aza indirmek için çeşitli stratejiler önerilmiştir.

Bu stratejiler arasında; katılımcıları sürecin bir parçası gibi hissettirmek, konuya olan ilgilerini canlı tutmak, sahiplenme duygusunu artırmak (her turun önceki yanıtlar üzerine inşa edildiğini vurgulamak) ve katılımcılarla güçlü ilişkiler kurarak süreci sürdürmek yer almaktadır (Trevelyan ve Robinson, 2015).

Uzman Seçimi

Delphi yöntemi, araştırmacının panelde yer alacak uzman seçiminde, uzmanların alandaki bilgi ve becerilerine, temsilin değişkenliğine, çalışılan yapıdaki deneyim ve çalışmalarına bağlı olarak dikkate alması gerektiğini vurgular (Nawagi ve ark., 2023). Başarılı bir Delphi çalışmasının en önemli adımlarından biri, belirli bir konunun en az bir yönünde uzmanlaşmış bilgi sahibi kişilerin titizlikle seçilmesidir. Uzman grubunun belirli bir nüfusu temsil etmesi gerekmez; ancak, seçilen katılımcıların ele alınan konuya dair derinlemesine bilgiye ve farklı bakış açılarına sahip olması önemlidir. Panelde yer alan uzmanların güçlü ve zayıf yönleri ile önyargıları, elde edilen sonuçların niteliğini de doğrudan etkileyecektir. Gerekli uzman profili, araştırma sorularına bağlı olarak değişebilir; ancak, başlangıçta panelistlerin dikkatle belirlenmesi, önyargıyı en aza indiren ve toplumun farklı perspektiflerini yansıtan bir çalışma ortaya koymak açısından kritik bir rol oynar (Sablatzky, 2022). Katılımcıları veya uzman panelini seçerken olasılıklı olmayan örnekleme yöntemi tercih edilir. Ancak, "uzman" kavramı tartışmaya açık bir konudur. Uzmanlar bilgili bireyler veya belirli bir konuda bilgi sahibi olanlar olarak tanımlanmıştır. Uzmanların;

- Ele alınan konu hakkında deneyim ve bilgi,
- Katılım için isteklilik ve kapasite,
- Katılım için zaman ve yeterli iletişim becerileri,

gibi özelliklere sahip olması gerekmektedir. Ancak bu özellikler içerisinde bilgi ve deneyimin nasıl tanımlanacağı konusunda tartışmalar bulunmaktadır. Bilgi, mesleki yeterlilik veya tescil yoluyla varsayılabilir de, bu uzmanlığı garanti etmez. Benzer şekilde, uzmanlık da kaç yıl çalışıldığıyla garanti edilemez (Trevelyan ve Robinson, 2015).

Uzman Sayısı

Delphi çalışmalarında araştırmacı tarafından karar verilmesi gereken ve halen devam eden önemli tartışmalardan biri de çalışmaya dahil edilecek uzman sayısı (örneklem büyüklüğü)'dir. Delphi sonuçlarının genellenebilir

olması için yeterli panel büyüklüğüne sahip olunmalı, farklı disiplinlerden üyeler çeşitli şekillerde temsil edilmeli ve coğrafi dağılım gözetilmelidir. Kesin ve katı kurallar olmamakla birlikte Delphi'ye katılacak katılımcı sayısının belirlenmesinde;

- Homojen veya heterojen örneklem: Eğer grup homojen ise, on ila on beş kişilik küçük bir örneklem yeterli olabilir. Ancak grup heterojen, özellikle uluslararası bir çalışma söz konusuysa, daha büyük bir örneklem ihtiyacı duyulacaktır. Heterojen gruplar, veri toplama, fikir birliğine ulaşma, analiz yapma ve sonuçları doğrulama süreçlerini önemli ölçüde zorlaştırabilir. Bununla birlikte, farklı bakış açıları ve geniş bir alternatif yelpazesi sunduğundan, çeşitlilik içeren panellerin daha başarılı sonuçlar verdiği düşünülmektedir.

- Karar kalitesi ve Delphi yönetilebilirliği dengesi: Örneklem büyüdükçe grup hatası azalır ve karar kalitesi artar. Ancak belirli bir sınırın ötesinde, Delphi sürecinin yönetimi ve veri analizi, sağlanan marjinal faydalara kıyasla daha zahmetli hale gelebilir.

- İç ve dış doğrulama: Grup büyüdükçe, sonuçların daha güçlü bir şekilde doğrulandığı kabul edilebilir. Ancak, daha küçük bir örneklemle de çalışılabilir ve elde edilen sonuçlar sonraki araştırmalarla teyit edilebilir,

olarak belirtilen bir dizi faktör göz önünde bulundurulmalıdır (Skulmoski, Hartman ve Krahn, 2007). Panel üye sayısı için kesin bir standart bulunmamaktadır. Yayınlanmış çalışmalarda 10 ile 1000 (tipik olarak 10 ile 100) arasında değişmektedir. Ancak, veri yönetimindeki zorluklar ve lojistik sorunlar (anket turları) nedeniyle, üç haneli rakamlarda örneklem büyüklüğüne sahip bir panel alışılmadık bir durumdur. Genel olarak, 30 ile 50 arasında bir üye sayısı optimal kabul edilmektedir (Nasa, Jain ve Juneja, 2021). Panelist sayısı 4'ten birkaç bine kadar değişebilir. Delphi'ler genellikle 8 ile 20 aralığında olma eğilimindedir. Daha büyük örneklem boyutlarının bulguların istikrarını artıracakı öne sürülmüştür. Örneğin, 10 uzmandan oluşan daha küçük bir panel oldukça istikrarsız olabilir, çünkü 1 kişi yanıtların %10'unu oluşturur ve dolayısıyla çalışmanın sonuçları üzerinde büyük bir etkiye sahiptir. Daha büyük panel boyutlarıyla, çalışmadaki bireysel uzman etkisi daha az olacak ve bulgular daha kararlı olacaktır. Öte yandan, araştırmacılar büyük uzman panellerinin veri toplama ve yönetiminde zorluklara yol açabileceğini bulmuşlardır. Delphi panel büyüklükleri genel olarak zaman ve parasal kısıtlar göz önünde bulundurularak gerçekleştirilmelidir ve ideal olarak katılımcı sayısı 8 ile 23 arasında olmalıdır

(Shang, 2023). Keşifsel faktör analizi için önerilen örneklem-öge oranı, çalışmadaki öge sayısına bağlı olarak örneklem büyüklüğünü belirlemek amacıyla kullanılır. Bu oranın en az 5'e 1 olması gerekmektedir. Örneğin, 30 ögeden (sorudan) oluşan bir araştırma için en az 150 katılımcı gereklidir (Memon ve ark., 2020). Örneklem büyüklüğü, araştırılan konunun niteliğine, sorunun karmaşıklığına, örneklem homojen veya heterojen olmasına ve mevcut kaynaklara bağlı olarak değişir. Homojen bir örneklem için (örneğin, 8-12 veya 10-15 katılımcı) önerilen örneklem büyüklüğü dikkate alındığında, sağlık alanındaki birçok Delphi çalışmasının nispeten küçük bir örneklemle gerçekleştirildiği ve belirli bir konu hakkında uzmanlığa sahip katılımcıları içerdiği görülmektedir. Ancak, Delphi çalışmalarının genellenebilirliğini artırmak için daha büyük örneklem gruplarına sahip olması gerektiği öne sürülebilir. Nitekim bazı araştırmalar, kolayda örnekleme yöntemiyle daha fazla katılımcıyı panele dâhil ederek bu ihtiyacı karşılamaya çalışmaktadır. Bir Delphi çalışması tasarlanırken örneklem özellikleri ve büyüklüğü dikkatle değerlendirilmelidir (Trevelyan ve Robinson, 2015).

Konsensus / Fikir Birliği

Delphi çalışmalarında fikir birliğinin nasıl tanımlanacağı, sorulan sorunun niteliğine ve araştırma sonuçlarına bağlıdır. Fikir birliği, bir uzlaşmanın olup olmadığını belirlemek veya sürecin bir durma noktasına ulaştığını göstermek için kullanılabilir. Farklı Delphi çalışmalarında fikir birliği farklı şekillerde ele alınmış olup, hangi kriterin en uygun olduğu konusunda ortak bir görüş bulunmamaktadır. Fikir birliği genellikle yargıların toplamı, merkezi eğilim ölçütleri ve istikrarın doğrulanması yoluyla belirlenir. Çalışma turları boyunca fikir birliği, bireysel katılımcıların belirli ifadelerle ne ölçüde katıldığını değerlendirir; bu da nihayetinde grup görüşünü ve katılımcılar arasındaki uyumu yansıtır. Yanıt istikrarı, fikir birliğinin baştan sona mevcut olup olmadığını ve turlar arasında gelişip gelişmediğini veya değişip değişmediğini gösterebilir. Delphi çalışmaları ile ilgili yapılan sistematik bir inceleme sonucunda, fikir birliğine varmak için kullanılan beş ana yöntemin olduğu belirlenmiştir. Bunlar arasında en yaygın kullanılan yöntemler, belirlenen bir eşik değerin üzerindeki medyan puanlar ile panel üyeleri arasında yüksek derecede uzlaşma sağlanmasıdır. Örneğin, medyan puanın belirli bir seviyenin üzerinde olması ve en düşük ya da en yüksek dilimdeki toplam puanların belirli bir yüzdesine ulaşması, fikir birliği ölçütleri olarak kabul edilebilir. Rastgele örneklemle seçilen 100 Delphi makalesi üzerinde yapılan bir başka sistematik incelemede ise en yaygın olarak anlaşma yüzdesinin kullanıldığı

görülmüştür. Yanıt istikrarının, fikir birliğini daha güvenilir şekilde yansıttığı ileri sürülse de, aslında bunun fikir birliğinden çok iç güvenilirliği değerlendirdiği de ifade edilmiştir. Güvenirlik ve fikir birliğinin birbirinden farklı olduğu, güvenilirliğin puanlayıcılar arasındaki varyansın oransal tutarlılığını ölçtüğü, fikir birliğinin ise katılımcıların birbirleriyle ne ölçüde hemfikir olduğunu ölçtüğü ileri sürülmüştür. Delphi çalışmalarında fikir birliğini ölçmek için farklı yöntemlerin gözden geçirildiği bir çalışmada, özneliği azaltmak ve sonuçların geçerliliğini sağlamak için istatistiklerin bir kombinasyonunun kullanılması tavsiye edilmiştir. Medyan ve IQR (Interquartile Range) güvenilir ölçütler olarak değerlendirilirken, IQR özellikle fikir birliğini belirlemede objektif ve hassas bir yöntem olarak kabul edilmiştir (Trevelyan ve Robinson, 2015). Tipik olarak, 4 ile 5 puanlık bir Likert ölçeğinde 1 veya daha küçük, 10 puanlık bir ölçekte ise 2 veya daha küçük bir IQR fikir birliği olarak kabul edilebilir. Ölçek puanı yükseldikçe, beklenen IQR de artar. Bu yüzden araştırmacılar, fikir birliğini göstermek için standart sapma yerine IQR'yi kullanmayı tercih etmelidir (Shang, 2023). Fikir birliği için literatürde önerilen ve yaygın olarak kullanılan kavramlar;

- Fikir birliği yüzdesi (aynı derecelendirmeye sahip % x),
- Merkezi eğilim ölçüsü (örneğin 9 puanlık Likert ölçeğinde medyan ≥ 7),
- Bir aralıktaki oran (örneğin Likert ölçeğinde belirli bir puanın üzerinde puan alan katılımcıların % x'i)
- Yanıtların dağılımı (örneğin 5 puanlık Likert ölçeğinde $IQR = 1$),

olarak belirtilmektedir. Farklı fikir birliği tanımları olsa da, şu anda bunlardan birini diğerine tercih etmek için genel kabul görmüş standartlar veya kılavuzlar mevcut değildir. Fikir birliğinin bu kadar farklı tanımları olsa bile, her tanımın farklı seviyeleri vardır. Örneğin, fikir birliği yüzdesi kullanacak olsa, tüm katılımcılardan %100 fikir birliği elde etmek inanılmaz derecede zor olurdu. Öte yandan, belki %30 ile %50 gibi etkileyici olmayan düşük bir fikir birliği seviyesine ulaşmak oldukça kolay olacak, ancak Delphi sonuçlarını daha az sağlam hale getirecektir. Bu yüzden bir denge sağlanmalıdır ve çözüm, araştırma sorusunun öneminde gizli olabilir. Örneğin, eğer konu hayati bir mesele ise, oldukça yüksek bir fikir birliği düzeyi tercih edilecektir. Araştırma sorusu hassasiyet gerektiren bir soru değilse araştırmacılar fikir birliği için %70 ile %80'i, hassasiyet gerektiren bir soru ise %90 ile %100'ü

hedeflemelidir (Shang, 2023). Hernandez-Leal ve ark., (2022) tarafından fikir birliği yüzdesinin belirlenmesinde,

$$C = \left(1 - \frac{V_n}{V_t}\right) * 100$$

V_n : olumsuz yanıt sayısı,

V_t : toplam yanıt sayısı,

eşitliğinin kullanılması önerilmiştir (Hernandez-Leal ve ark., 2022). Yaygın olarak kullanılan diğer fikir birliği tanımları, ortalama, medyan ve mod gibi merkezi eğilim ölçülerini içerir ve fikir birliği yüzdesi yerine tercih edilmesi gerektiği vurgulanmaktadır. Ancak, Likert türü ölçek verileri geleneksel olarak sıralı ölçek kabul edildiğinden, yalnızca aralık ölçeğindeki veriler için uygun olan ortalama hesaplamasında kullanılamaz. Bu nedenle, fikir birliğini tanımlarken ortalamanın raporlanmasından kaçınılması, bunun yerine medyan ve mod'un tercih edilmesi önerilmektedir. Ayrıca, medyan aykırı değerlerden daha az etkilendiği için, belirli bir konuda güçlü ve farklı görüşlere sahip uzmanların bulunduğu durumlarda daha güvenilir bir ölçüt olarak kabul edilir. Mod da kullanılır, ancak bir konunun bazı uzmanlar tarafından orta derecede önemli görülürken diğerleri tarafından aşırı derecede önemli görülmesi gibi çok mod'lu veri dağılımlarını yakalayamadığı için sorunlu olabilir. Bu nedenle, merkezi eğilimi kullanırken ortancanın tek başına kullanılması, mod'un ise ikincil bir analiz biçimi olarak değerlendirilmesi tavsiye edilir. Standart sapma, yaygın olarak kullanılan bir tanımlayıcı istatistik ve dağılım ölçüsüdür. Ortalama etrafındaki dağılımı gösteren bu istatistik, varyansın karekökü alınarak hesaplanır. Dolayısıyla, dağılım küçüldükçe standart sapma da azalır ve bu durum, fikir birliğinin sağlanma olasılığının arttığını gösterebilir. Ancak, ortalamanın kullanımına yönelik eleştirilerde olduğu gibi, standart sapma da aykırı değerlere karşı duyarlıdır (her ne kadar daha az etkileniyor olsa da) ve sıralı veri sağlayan Likert türü ölçeklerde kullanımı uygun değildir (Shang, 2023). Yanıt istikrarı, sonuçların güvenilirliğini değerlendiren ve turlar arasındaki mutabakat düzeyini gösteren bir ölçüttür. Sonuçların tutarlı ve güvenilir olması önemlidir; hatta grup içinde istikrar sağlanmadıkça fikir birliğinin anlamlı olmayacağı öne sürülmüştür. İstikrarı test etmek amacıyla Ki-kare analizi kullanılmıştır, ancak bu test turlar arasındaki yanıtların istikrarını değil, turların içerdikleri yanıtlardan bağımsız olup olmadığını belirlediği için Delphi çalışmalarında önerilmemektedir. Buna karşın, bireylerin grup içindeki görüşlerinin tutarlılığını gösterdiği için yüksek veya artan Kappa değerleri, istikrarın ölçülmesinde daha uygun bir

yöntem olarak savunulmuştur. Ancak Kappa, nominal ölçekte bir uyum ölçüsüdür ve derecelendirmelerin doğal bir sırası olmadığını varsayar. Yanıtların istikrarını test etmek için sıralı verilerle kullanıma uygun testler arasında, aynı gruptaki bireylerin önce ve sonra durumlarına ilişkin eşleştirilmiş verileriyle çalışan Wilcoxon eşleştirilmiş çiftler işaretli sıralama testi yer alır. Önemli bir değişiklik yoksa yanıtlar istikrarlı kabul edilir. Yanıtın istikrarı, turlar arasında medyan ve IQR verileri sağlanarak, ortalamaların ve standart sapmaların grafiksel sunumu veya sınıf içi korelasyon katsayısı (Intra-class Correlation Coefficient = ICC) ile de rapor edilebilir. Turlar arasında medyan ve IQR değerlerinin belirlenmesi, yorumlanması kolay veriler sağlar ve bu önerilmektedir. Ayrıca Wilcoxon eşleştirilmiş çiftler işaretli sıralama testinin veya ICC'nin kullanılması tavsiye edilmektedir. Delphi yöntemi artık yaygın olarak kullanılan bir yöntemdir ve diğer uzlaşi yöntemlerine göre yüz yüze temas gerektirmeme avantajına sahip olmakla birlikte grup görüşünü yönlendirmeye devam etmektedir. Diğer uzlaşi yöntemlerinden (nominal grup tekniği ve uzlaşi geliştirme konferansı gibi) daha geniş bir grup katılımını kolaylaştırabilir ve böylece katılımcıların coğrafi konumlarından kaynaklanan katılım önyargılarını önleyebilir (Trevelyan ve Robinson, 2015).

Sağlık Araştırmalarında Delphi Yöntemi

Sağlık alanında Delphi, 1960'lı yıllardan itibaren öncelikli araştırma ve finansman alanlarının belirlenmesi, eğitim ve sağlık politikalarına ilişkin gelecekteki eylemlere ilişkin kararların alınması ve klinik davranışların tanımlanması amacıyla yapılan çalışmalarda kullanılmaya başlanmıştır. Tıp araştırmacıları, fikir birliği sağlamak amacıyla genellikle Delphi yöntemini tercih etmektedir. Bu yöntemin doğru uygulanması, bir çalışmanın geçerliliği açısından büyük önem taşır. Ancak, mevcut tıbbi literatür bu konuda yeterince net değildir (Massaroli ve ark., 2017). Delphi kullanılan tıbbi çalışmaların raporları genellikle Delphi çalışmalarının nasıl yürütüldüğüne dair ayrıntılı bilgi içermemektedir. Bu durum, kısmen tıp dergilerindeki kelime sınırlamalarından kaynaklanıyor olabilir. Düşük tekrarlanabilirlik ve sınırlı dış geçerlilik gibi dezavantajlarına rağmen, bu tür bilgiler yeni kullanıcılar için oldukça değerlidir. Çünkü yöntemin yüzeyde basit görünmesi, aslında doğru şekilde uygulanmasının gerektirdiği çaba ve zorluklarla çelişmektedir. Bir Delphi çalışmasını tamamlamak için gereken zaman genellikle küçümsenir ve yanıt oranları çoğu zaman hayal kırıklığı yaratacak kadar düşük olur. Yöntemin hatalı uygulanması, elde edilen sonuçların geçerliliğini

zayıflatmaktadır. Ayrıca, panelistlerin turlara nasıl yanıt verdiği, yüksek yanıt oranlarını destekleyen faktörler ve bu oranların nasıl iyileştirilebileceği konusunda literatürde oldukça sınırlı bilgi bulunmaktadır (Veugelers ve ark., 2020). Sağlık bilimlerinin çeşitli alanlarında, örneğin klinik hasta bakımı, halk sağlığı, sağlık hizmetleri araştırmaları ve tıp eğitimi gibi disiplinlerde, özellikle fikir birliğini sağlamak amacıyla kullanılmaktadır (Donohoe ve Needham, 2009). Kanıta dayalı bilgi, pratik deneyimler ve bazen de günlük yaşam tecrübeleri, yapılandırılmış bir süreç içinde bir araya getirilir. Bu sürecin temel amacı, sağlık sektöründe karar alma ve uygulamaları desteklemek üzere uzlaşmaya dayalı kılavuzlar veya standartlar oluşturmaktır (Niederberger ve Spranger, 2020; Spranger ve ark., 2022).

SONUÇ

Delphi yöntemi, yeni kavramların oluşturulmasında ve gelecekteki araştırmaların yönlendirilmesinde güvenilir bir ölçüm aracı olarak kendini kanıtlamıştır. Bu yöntem, bir konu üzerindeki fikir birliği düzeyini değerlendirmek ve mevcut görüş ayrılıklarını gidermek amacıyla uzmanların görüşlerini sistematik bir şekilde toplar (Vogel ve ark., 2019). Sağlık araştırmalarında Delphi yöntemlerini değerlendirmek için genel kabul görmüş standart kalite parametreleri bulunmamaktadır. Delphi metodolojisinin temel unsurları arasında anonimlik, yineleme, kontrollü geri bildirim ve fikir birliğinin istatistiksel sağlamlığı yer almaktadır. Yayımlanan çalışmalarda çoğunlukla uyarlanmış Delphi yöntemleri kullanılmasına rağmen, uzman panelinin seçimi veya fikir birliği kriterleri gibi detaylar genellikle net bir şekilde açıklanmamaktadır. Bir Delphi çalışmasının kalitesi büyük ölçüde tasarımına ve uygulama sürecinin doğruluğuna bağlıdır. 2017 yılında, bu tür çalışmaların yürütülmesi ve raporlanmasına yönelik CREDES (Guidance on Conducting and REporting DELphi Studies) kılavuzları yayımlanmıştır. O tarihten itibaren binlerce Delphi çalışması yayınlanmış olsa da, CREDES'in kullanımı yaygın bir şekilde benimsenmemiştir. Delphi çalışmalarının standardizasyonu konusunda hâlâ güçlü bir talep bulunmaktadır ve araştırmacılar, yöntemin güvenilirliğini ve kalitesini artırmak amacıyla Delphi metodolojisini incelemeye devam etmektedir. Ancak, pratik tasarım tercihlerinin nasıl şekillendiği ve nihai sonuçları üzerindeki etkileri konusunda literatürde hâlâ sınırlı bilgi bulunmaktadır (Veugelers ve ark., 2020).

Kaynaklar

Dergi Makalesi

- Barrett, D., Heale, R., (2020). What are Delphi studies? *Evidence Based Nursing*, 23(3), 68-69. <https://doi.org/10.1136/ebnurs-2020-103303>
- Bettcher, D.W., Sapiirie, S., Goon, E.H., (1998). Essential public health functions: results of the international Delphi study. *World Health Statistics Quarterly*, 51(1), 44-54. <https://iris.who.int/handle/10665/55726>
- Donohoe, H.M., Needham, R.D., (2009), Moving best practice forward: Delphi characteristics, advantages, potential problems, and solutions. *International Journal of Tourism Research*, 11(5), 415-437. <https://doi.org/10.1002/jtr.709>
- Hernandez-Leal, M.J., Codern-Bove, N., Perez-Lacasta, M.J., et al., (2022), Development of support material for health professionals who are implementing shared decision making in breast cancer screening: validation using the Delphi technique. *BMJ Open*, 12: e052566. doi:10.1136/bmjopen-2021-052566
- Hsu, C., Sandford, B.A., (2007), The Delphi technique: making sense of consensus. *Practical Assessment Research, and Evaluation*, 12(1), 10. <https://doi.org/10.7275/pdz9-th90>
- Landeta, J., (2006), Current validity of the Delphi method in social sciences. *Technological Forecasting & Social Change*, 73(5), 467-482. <https://doi.org/10.1016/j.techfore.2005.09.002>
- Massaroli, A., Martini, J.G., Lino, M.M., Spenassato, D., Massaroli, R., (2017), The Delphi method as a methodological framework for research in nursing. *Texto & Contexto Enfermagem*, 26(4), e1110017. <http://dx.doi.org/10.1590/0104-07072017001110017>
- Memon, M.A., Ting, H., Cheah, J.H., Thurasamy, R., Chuah, F., Cham, T.H., (2020), Sample size for survey research: review and recommendations. *Journal of Applied Structural Equation Modeling*, 4(2), i-xx. [https://doi.org/10.47263/JASEM.4\(2\)01](https://doi.org/10.47263/JASEM.4(2)01)
- Mukherjee, N., Hüge, J., Sutherland, W.J., McNeill, J., Van Opstal, M., Dahdouh-Guebas, F., Koedam, N., (2015), The Delphi technique in ecology and biological conservation: applications and guidelines. *Methods in Ecology and Evolution*, 6(9), 1097-1109. <https://doi.org/10.1111/2041-210X.12387>
- Naisola-Ruiter, V., (2022), The Delphi technique: a tutorial. *Research in Hospitality Management*, 12(1), 91-97. <https://doi.org/10.1080/22243534.2022.2080942>
- Nasa, P., Jain, R., Juneja, D., (2021), Delphi methodology in healthcare research: How to decide its appropriateness. *World Journal of Methodology*, 11(4), 116-129. <https://dx.doi.org/10.5662/wjm.v11.i4.116>
- Nawagi, F., Munabi, I.G., Vyt, A., Kiguli, S., Rabin, T., Waggie, F., Mubuuke, A.G., (2023), Using the modified Delphi technique to develop a framework for interprofessional education during international electives in health professions training institutions in Sub-Saharan Africa. *Frontiers in Medicine*, 10, 1225475. doi: 10.3389/fmed.2023.1225475
- Niederberger, M., Spranger, J., (2020), Delphi technique in health sciences: a map. *Frontiers in Public Health*, 8, 457. <https://doi.org/10.3389/fpubh.2020.00457>
- Sablitzky, T., (2022), Methods Moment: the Delphi method. *Hypothesis*, 34(1), <https://doi.org/10.18060/26224>
- Shang, Z., (2023), Use of Delphi in health sciences research: A narrative review. *Medicine*, 102(7), e32829. <http://dx.doi.org/10.1097/MD.00000000000032829>
- Shi, X., He, X., Liu, Q., Feng, L., Li, Y., Zhang, X., Cheng, Z., Zhang, C., Gao, Y., (2024), Conducting and reporting the Delphi method in traditional Chinese medicine syndrome diagnosis research: A cross-sectional analysis. *Heliyon*, 10(3), e25162. <https://doi.org/10.1016/j.heliyon.2024.e25162>
- Skulmoski, G.J., Hartman, F.T., Krahn, J., (2007), The Delphi method for graduate research. *Journal of Information Technology Education: Research*, 6, 1-21. <https://doi.org/10.28945/199>
- Spranger, J., Homberg, A., Sonnberger, M., Niederberger M., (2022), Reporting guidelines for Delphi techniques in health sciences: A methodological review. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 172, 1-11. <https://doi.org/10.1016/j.zefq.2022.04.025>

- Stitt-Gohdes, W.L., Crews, T.B., (2004), The Delphi technique: a research strategy for career and technical education. *Journal of Career and Technical Education*, 20(2), 55-67. <https://doi.org/10.21061/jcte.v20i2.636>
- Trevelyan, E.G., Robinson, N., (2015), Delphi methodology in health research: how to do it?. *European Journal of Integrative Medicine*, 7(4), 423-428. <http://dx.doi.org/10.1016/j.eujim.2015.07.002>
- Varndell, W., Fry, M., Elliott, D., (2021), Applying real-time Delphi methods: development of a pain management survey in emergency nursing. *BMC Nursing*, 20, 149. <https://doi.org/10.1186/s12912-021-00661-9>
- Veugelers, R., Gaakeer, M.I., Patka, P., Huijsman, R., (2020), Improving design choices in Delphi studies in medicine: the case of an exemplary physician multi-round panel study with 100% response. *BMC Medical Research Methodology*, 20(1), 156. <https://doi.org/10.1186/s12874-020-01029-4>
- Vogel, C., Zwolinsky, S., Griffiths, C., Hobbs, M., Henderson, E., Wilkins, E., (2019), A Delphi study to build consensus on the definition and use of big data in obesity research. *International Journal of Obesity*, 43(12), 2573-2586. <https://doi.org/10.1038/s41366-018-313-9>

İSTATİSTİKSEL PROSES KONTROL GRAFİKLERİ

Özlem Toluk**, İlker Ercan***

Giriş

İstatistiksel proses kontrol (İPK), bir sistemin iyileştirilmesi amacı ile uygulanan bir dizi yöntemler olduğu gibi sağlık hizmeti yönetiminde ve tedavi süreçlerinin geliştirilmesinde de kullanılan çok yönlü bir araçtır. İPK yöntemi ile hastaya yapılan müdahalenin başarılı, etkisiz veya yan etki oluşturup oluşturmadığı, cerrahi süreçteki değişimin sabit veya sürecin anlamlı bir değişim sinyali verip vermediği izlenebilmektedir. Biyomedikal değişkenler, kardiyovasküler sistem ölçümleri, testlerin ertelendiği ya da zamanında raporlanmadığı vakaların oranı, diyabet, hipertansiyon, kalp cerrahisi başarısı, kas ağrıları, astım hastalarının tedavisi, aylık ilaç uygulama hataları sayısı gibi birçok alanda kullanılabilmesinden dolayı sağlık alanında önemli bir yere sahiptir (Thor ve ark., 2007).

İPK bir sürecin verileri için geliştirilen, sürecin anlaşılmasını uygun hale getiren istatistiksel analiz yöntemidir. İPK süreç performansının gözlenmesi ile alakalı olduğundan sadece üretim süreci için değil birçok alanda kullanılabilir. Bundan dolayı performansın ölçülebildiği her alanda İPK ilgili süreci analiz edebilir (Stapenhurst, 2005). İstatistiksel proses kontrolü zaman serisi analiz yöntemlerini grafik ile birleştirerek, sürecin yönetilebilmesi ve geliştirilebilmesi için istatistiksel yöntemleri kullanmaktadır. Bir ürünün istenen nitelikte olması için uygulanan bütün koşullar bir süreci oluşturur. İPK ile ilgili en önemli yaklaşım kontrol grafikleridir. Kontrol grafiği, süreç karakteristiğinin belirlenen zaman süresince istatistiksel yöntemlerle sınırları belirlenen bir çizimdir. Prosesi gözlemek için kullanıldığında, süreç üzerinde uygun eylem tipinin belirlenmesine imkân sağlamaktadır (Benneyan ve ark., 2003; Stapenhurst, 2005).

** Dr., Bezmialem Vakıf Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, İstanbul, Türkiye
erozlem@outlook.com.tr, ORCID: orcid.org/0000-0001-6495-0839

*** Prof. Dr., Bursa Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Bursa, Türkiye iercan@msn.com,
ORCID: orcid.org/0000-0002-2382-290X

Süreç Parametrelerinin Nokta Tahmini

Bir rasgele değişken olasılık dağılımıyla tanımlanmakta ve dağılım parametreleriyle ifade edilmektedir. Parametreler genelde bilinmediğinden örneklem verisi ile tahmin edilmekte, nokta tahmini ise bilinmeyen parametre için tek bir sayısal değeri ifade etmektedir. Normal dağılıma sahip bir anakütleden çekilen örneklem ortalaması $\bar{x}$ ve varyansı s^2 , anakütle ortalaması μ ve anakütle varyansı σ^2 'nin nokta tahmincileridir. Nokta tahmincisi yansız, tutarlı ve etkin olmalıdır (Çömlekçi, 1989).

Örneklem standart sapması s anakütle standart sapması σ 'nın yansız bir tahmin edicisi değildir. Tablo 1' de belirtilen faktörler kullanılarak düzeltme yapılmakta, böylece standart sapmanın yansız tahmincisi eşitlik 1 ile elde edilebilmektedir (Montgomery, 2009).

$$\hat{\sigma} = \frac{s}{c_4} \quad (1)$$

Tablo 1. Kontrol grafikleri için kullanılan düzeltme katsayıları

n	Ortalama grafiği için					Standart sapma grafiği için					Aralık grafiği için						
	Kontrol sınırları çarpanları			Merkez çizgi çarpanları		Kontrol limiti çarpanları				Merkez çizgi çarpanları		Kontrol limiti çarpanları					
	A	A ₂	A ₃	c ₄	1/c ₄	B ₃	B ₄	B ₅	B ₆	d ₂	1/d ₂	d ₃	D ₁	D ₂	D ₃	D	
2	2,12	1,88	2,66	0,78	1,25	0	3,27	0	2,61	1,13	0,89	0,85	0	3,69	0	3,	2,5
3	1,73	1,02	1,95	0,87	1,13	0	2,57	0	2,28	1,69	0,59	0,89	0	4,36	0	2,	2,5
4	1,50	0,73	1,63	0,92	1,08	0	2,27	0	2,09	2,06	0,49	0,88	0	4,70	0	2,	2,2
5	1,34	0,57	1,43	0,94	1,06	0	2,09	0	1,96	2,33	0,43	0,86	0	4,92	0	2,	2,1
6	1,23	0,48	1,29	0,95	1,05	0,03	1,97	0,03	1,87	2,53	0,39	0,85	0	5,08	0	2,	2,0
7	1,13	0,42	1,18	0,96	1,04	0,12	1,88	0,11	1,81	2,70	0,37	0,83	0,20	5,20	0,08	1,9	1,9
8	1,06	0,37	1,10	0,97	1,04	0,18	1,82	0,18	1,75	2,85	0,35	0,82	0,39	5,30	0,14	1,8	1,8
9	1,00	0,34	1,03	0,97	1,03	0,24	1,76	0,23	1,71	2,97	0,34	0,81	0,55	5,40	0,18	1,8	1,8
10	0,95	0,31	0,98	0,97	1,03	0,28	1,72	0,28	1,67	3,08	0,32	0,80	0,69	5,47	0,22	1,7	1,7
11	0,91	0,29	0,93	0,98	1,02	0,32	1,68	0,31	1,64	3,17	0,31	0,79	0,81	5,53	0,26	1,7	1,7
12	0,87	0,27	0,89	0,98	1,02	0,35	1,65	0,35	1,61	3,26	0,31	0,78	0,92	5,59	0,28	1,7	1,7
13	0,83	0,25	0,85	0,98	1,02	0,38	1,62	0,37	1,59	3,34	0,30	0,77	1,03	5,65	0,31	1,6	1,6
14	0,80	0,24	0,82	0,98	1,02	0,41	1,59	0,40	1,56	3,41	0,29	0,76	1,12	5,69	0,33	1,6	1,6
15	0,78	0,22	0,79	0,98	1,02	0,43	1,57	0,42	1,54	3,47	0,29	0,76	1,21	5,74	0,35	1,6	1,6
16	0,75	0,21	0,76	0,98	1,02	0,45	1,55	0,44	1,53	3,53	0,28	0,75	1,28	5,78	0,36	1,6	1,6
17	0,73	0,20	0,74	0,99	1,02	0,47	1,53	0,46	1,51	3,59	0,28	0,74	1,36	5,82	0,38	1,6	1,6
18	0,71	0,19	0,72	0,99	1,01	0,48	1,52	0,48	1,50	3,64	0,27	0,74	1,43	5,85	0,39	1,6	1,6
19	0,69	0,19	0,70	0,99	1,01	0,50	1,50	0,49	1,48	3,69	0,27	0,73	1,49	5,89	0,40	1,6	1,6
20	0,67	0,18	0,68	0,99	1,01	0,51	1,49	0,50	1,47	3,73	0,27	0,73	1,55	5,92	0,41	1,5	1,5
21	0,66	0,17	0,52	0,99	1,01	0,52	1,48	0,52	1,46	3,78	0,26	0,72	1,61	5,95	0,43	1,5	1,5
22	0,64	0,17	0,53	0,99	1,01	0,53	1,47	0,53	1,45	3,82	0,26	0,72	1,66	5,98	0,43	1,5	1,5
23	0,63	0,16	0,54	0,99	1,01	0,54	1,46	0,54	1,44	3,86	0,26	0,72	1,71	6,01	0,44	1,5	1,5
24	0,61	0,63	0,55	0,99	1,01	0,55	1,45	0,55	1,43	3,89	0,26	0,71	1,76	6,03	0,45	1,5	1,5
25	0,60	0,62	0,56	0,99	1,01	0,56	1,44	0,56	1,42	3,93	0,25	0,71	1,80	6,06	0,46	1,5	1,5

n: örneklemdeki gözlem sayısı

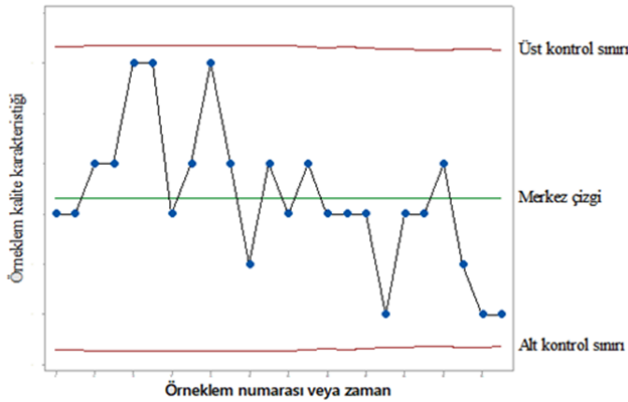
Kontrol grafiklerinin hesaplamalarında kullanılacak olan eşitlikler eşitlik 2'de verilmiştir (Montgomery, 2009):

$$\begin{aligned}
 A &= \frac{3}{\sqrt{n}} & A_2 &= \frac{3}{d_2\sqrt{n}} & A_3 &= \frac{3}{c_4\sqrt{n}} & c_4 &\cong \frac{4(n-1)}{4n-3} \\
 B_3 &= 1 - \frac{3}{c_4\sqrt{2(n-1)}} & B_4 &= 1 + \frac{3}{c_4\sqrt{2(n-1)}} \\
 B_5 &= c_4 - \frac{3}{\sqrt{2(n-1)}} & B_6 &= c_4 + \frac{3}{\sqrt{2(n-1)}}
 \end{aligned} \tag{2}$$

Alt ve Üst Kontrol Sınırları

Süreçin gözlemlenmesi için uygulanan istatistiksel proses kontrolün başlıca yöntemlerinden biri kontrol grafiğidir. Değişkenliği azaltmak açısından kontrol grafikleri düzenli olarak uygulanmalıdır (Montgomery, 2009).

Süreçteki rasgele değişkenin normal ya da normale yakın dağıldığı varsayılmaktadır. Laboratuvar ortamında veya hastalıkların izlenmesinde süreç ne kadar iyi tasarlanmış ve korunmuş olsa da belirli miktarda doğal kabul edilen değişkenlik oluşmaktadır. Bu değişkenlik oldukça küçük kümülatif etkiye sahip olduğundan, süreç bu değişkenlikle devam ederse, istatistiksel olarak kontrol sınırları içinde kalmaktadır. Örneklem numarası veya zamana göre örneklemden hesaplanan ya da ölçülen süreç değişkeninin grafiksel olarak gösterimi kontrol grafiği ile izlenmektedir. Klasik bir kontrol grafiği Şekil 1’de gösterilmektedir. Merkez çizgi (MÇ) süreç değişkeninin ortalama değerini göstermektedir (Montgomery, 2009). Grafikteki diğer çizgiler ise alt kontrol sınırı (AKS) ile üst kontrol sınırını (ÜKS).



Şekil 1: Klasik bir kontrol grafiği

Shewhart tarafından önerilen kontrol grafiklerinin limitlerinin 3σ olarak uygulanmasından dolayı, alt ve üst kontrol limitleri standart sapmanın 3 katı olarak kabul görmektedir (Birgören, 2015; Şenol 2012). Alt kontrol sınırı merkez çizginin $-Z\alpha/2\sigma$ aşağısında ve üst kontrol sınırı merkez çizginin $+Z\alpha/2\sigma$ üstünde çizilmektedir. Süreç kontrol altında ise örneklem noktalarının

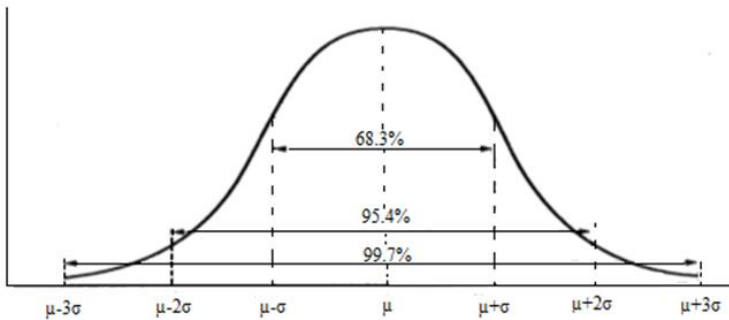
neredeyse tamamı alt ve üst sınırlar arasında kalmaktadır (Montgomery, 2009). Eğer bir nokta kontrol sınırlarının dışında ise süreç kontrol altında olmamaktadır. Bu durumun araştırılması ve düzeltilmesi gerekmektedir.

Şekil 2.2’te normal dağılım eğrisinde belirli sınırlar arasında kalan alanların olasılıkları gösterilmektedir. Değişkenlerin $\pm 2\sigma$ alanında bulunma olasılığı %95,45; $\pm 3\sigma$ alanında bulunma olasılığı %99,73’tür. İstatistiksel proses kontrol yönteminde n hacimli örneklemdeki değişkenlik rasgelelikten kaynaklı ise süreç değişkeni 0,9975 ile 0,9545 aralığında bulunmaktadır. Süreç normal dağılım gösteriyor ise çıktılarının 0,9973’ü $\mu \pm Z\alpha/2\sigma = \mu \pm 3\sigma$ sınırlarının içinde bulunmakta ve 0,0027’si bu sınırların dışında kalmaktadır (Şenol, 2012).

$\mu \pm Z\alpha/2\sigma = \mu \pm 3\sigma$ olarak belirtilen bu sınırlara kontrol sınırları (KS) denilmektedir (Şenol, 2012). Kontrol sınırlarının genel formülü eşitlik 3’te belirtilmektedir.

$$\mu: \text{Merkez çizgi} \quad (3)$$

$$\text{Kontrol Sınırları} : \mu \pm Z\alpha/2\sigma$$



Şekil 2: Normal dağılım

Bütün noktalar sistemli, düzenli ya da rasgele olmayan davranış sergiliyorlar ise kontrol limitleri arasında bulunsalar dahi bu durum sürecin kontrol dışı olduğunu gösteriyor olabilir. Süreç kontrol altında olduğunda noktalar rasgele bir desen oluşturmaktadır (Montgomery, 2009).

Kontrol grafikleri ve hipotez testleri arasında yakın bir ilişki bulunmaktadır. Şekil 1’de gösterilen grafikte dikey eksenin örneklem ortalaması $\bar{x}$ ’ yı belirttiği varsayılırsa; $\bar{x}$ noktalarının şu anki değerlerinin kontrol sınırları arasında bulunmasından dolayı süreç ortalamasının μ_0 ’a eşit olduğu ve sürecin kontrol altında olduğu düşünülmektedir. $\bar{x}$ her hangi bir kontrol sınırının dışında ise süreç ortalamasının kontrol dışı $\mu_1 \neq \mu_0$ olduğu anlamına gelmektedir (Montgomery, 2009).

H_0 : süreç kontrol sınırları içinde;

H_1 : süreç kontrol sınırları dışında (Montgomery, 2009).

$$H_0: \mu_1 = \mu_0 \quad (4)$$

$$H_1: \mu_1 \neq \mu_0$$

Kontrol grafikleri ile hipotez testleri arasında bakış açısı yönünden farklılıklar bulunmaktadır. İstatistiksel hipotezleri test ederken genellikle varsayımların geçerliliği kontrol edilmektedir. Kontrol grafiklerinde ise istatistiksel kontrol durumundan sapmayı belirlemek için kullanılmaktadır (Montgomery, 2009).

Kontrol Grafikleri

Proses kontrol grafikleri nicel ve nitel bütün kalite karakteristiklerine uygulanabilmektedir. Shewhart kontrol grafikleri ($\bar{X}$ -S, $\bar{X}$ -R, X-MR, P, U, C, NP), birikimli toplam kontrol grafiği (cumulative sum control charts-CUSUM- CUSCC) ve üstel ağırlıklı hareketli ortalama (exponentially weighted moving average - EWMA) kontrol grafiği gibi birçok kontrol grafiği çeşidi bulunmaktadır. Süreç verileri incelenerek uygun kalite özelliği belirlenmeli ve bu özellik için en uygun kontrol grafiği seçilmelidir (Tablo 2) (Şenol, 2012).

Ölçülebilen süreç karakteristikleri için örneklem ortalaması ($\bar{X}$), değişim aralığı (R) ve standart sapma (S) kontrol grafikleri kullanılmaktadır. $\bar{X}$, R ve S grafikleri değişkenliğin gerçek değerini gösteren ölçülebilen değişkenler olarak adlandırılan ölçümlere sahip olmalıdır. Süre, uzunluk gibi değişkenler ölçülebilen değişkenlere örnek olarak verilebilir. Ölçülebilen kalite karakteristiklerinin kontrol grafikleri, sürecin ortalaması ve yayılmasındaki değişkenliği, bu değişimin yönü ve büyüklüğünü göstermektedir. Kontrol grafikleri aynı zamanda sürecin istatistiksel olarak kontrol altında olup olmadığını izlemek için kullanılmaktadır. Ortalaması μ ve standart sapması σ olan normal dağılım gösteren bir anakütleden alınan n birimlik örneklem ile kontrol sınırları hesaplanmaktadır (Western Electric Company, Inc. 1958). $\bar{X}$, R ve S birer rasgele değişken olduğundan yani her örneklem için farklı değer alacağından, olasılık dağılımlarından bahsedilebilmektedir. Örneklem hacmi $n \leq 10$ ise $\bar{X}$ ve R grafikleri, $n > 10$ ise $\bar{X}$ ve S kontrol grafikleri kullanılmaktadır. $\bar{X}$, R ve S grafikleri için alt grup sayısı $k \geq 25$ olmalıdır. $\bar{X}$ ve R ya da $\bar{X}$ ve S grafikleri beraber kullanılmaktadır. $\bar{X}$ grafiği belirlenen ortalamaya veya verilerin ortalamasına göre süreci izlemekte, sürecin değişkenliğini ise R veya S grafikleri göstermektedir. Böylece örneklem istenilen ortalamaya ve değişkenliğe sahip olduklarında sürecin istatistiksel olarak kontrol altında olduğu söylenebilmektedir (Şenol, 2012).

Ölçülemeyen kalite karakteristikleri için kusurlu oranı (P), kusurlu sayısı (NP), kusur sayısı (C) ve birim başına düşen kusur sayısı (U) kontrol grafikleri kullanılmakta, NP ve C kontrol şemalarının çizilebilmesi için sabit örneklem hacmi gerekirken, P ve U kontrol şemaları için hem sabit hem değişken örneklem hacmi uygulanabilmektedir (Karkalousos ve Evangelopoulos, 2015).

Nicel ölçümler için örneklem ya da alt grup hacmi (n) 2- 12 arasında, nitel ölçümler için 10- 50 arasında olmaktadır. Örneklem sayısı (k) ya da alt grup için en az 20- 25 arasında olmalıdır (Şenol, 2012).

Tablo 2. Shewhart kontrol grafiklerinin yaygın kullanım türleri ve örnekler (Benneyan, 2008; Montgomery, 2009)

Grafik Türleri	Olasılık Dağılımı	Kullanım Yeri
$\bar{X}$ ve S grafikleri	Normal Dağılım	Normal dağılım gösteren sürekli ölçümlerde
np grafikleri	Binom Dağılımı	Süreç tarafından üretilen ikili vakaların toplam sayısı (örneklem hacimleri aynıdır)
p grafikleri	Binom dağılımı	Süreç tarafından üretilen ikili vakaların oranı (örneklem hacimleri değişken olabilir)
c grafikleri	Poisson dağılımı	Her bir hasta veya örneklem birimi için bir veya birden fazla olabilen toplam vaka sayısı (örneklem hacimleri aynıdır)
u grafikleri	Poisson dağılımı	Her bir hasta veya örneklem birimi için bir veya birden fazla olabilen vaka hızı (örneklem hacimleri değişken olabilir)

X- Ortalama ($\bar{X}$) Kontrol Grafiği

$\bar{X}$ kontrol grafiği ile sürecin merkezi eğilimi incelenerek süreç ortalamasında sapma olup olmadığına bakılmaktadır (Birgören, 2015, Şenol, 2012). X süreç özelliğinin normal dağılıma uyduğu varsayımına göre ortalaması μ ve standart sapması σ olmaktadır. n birimlik alt grup olarak adlandırılan örneklem alınıldığında $X = (x_1, x_2, x_3, \dots, x_n)$ ortalaması eşitlik 5'te hesaplanmaktadır (Birgören, 2015):

$$\bar{X} = \frac{\sum x_i}{n} \quad (5)$$

Örneklem $\bar{X}$ veya μ ortalamalı ve $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ standart sapmalı normal dağılım göstermektedir. Bu sebeple $\bar{X}$ 'nin güven sınırları $\mu \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ olmaktadır. $\bar{X}$ kontrol grafiğinin kontrol sınırları (KS) ve merkez çizgi (MÇ) sınırı eşitlik 6 ile hesaplanmaktadır (Şenol, 2012):

$$MÇ = \mu \quad (6)$$

$$KS = \mu \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \mu \pm Z_{\alpha/2} \sigma_{\bar{x}}$$

Kontrol grafiklerinde Tip-I hata $\alpha = 0,0027$ olarak alındığından sürecin %99,73'ünün kontrol altında olduğu varsayılmaktadır (Şenol, 2012). Genellikle anakütle ortalaması μ ve standart sapması σ bilinmediğinden süreçten alınan en az 20-25 alt gruba sahip referans örneklerle μ ve σ tahmin edilmektedir (Birgören, 2015).

Hipotezler eşitlik 4'te gösterildiği gibi kurulmakta ve H_0 hipotezi $\alpha = 0,0027$ anlamlılık seviyesinde test edilmektedir. μ ve σ bilindiğinde $\bar{X}$ kontrol grafiğinin sınırları eşitlik 7 ile belirlenmektedir (Şenol, 2012):

$$M\check{C} = \mu \quad (7)$$

$$KS = \mu \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Tablo 1'deki çarpanlar dikkate alındığında $\bar{X}$ kontrol grafiği sınırları eşitlik 8 ile hesaplanmaktadır (Şenol, 2012):

$$M\check{C} = \mu \quad (8)$$

$$KS = \mu \pm A\sigma$$

$\bar{X}$ kontrol grafiğinde μ ve σ bilinmediğinde örneklem ortalamalarının ortalaması $\bar{\bar{X}}$, μ 'nün tahmin edicisi olmaktadır (Şenol, 2012).

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

$$\bar{\bar{X}} = \frac{1}{k} \sum_{j=1}^k \bar{x}_j \quad (9)$$

k: örneklem sayısı

n: örneklem büyüklüğü

j=1,...,k

$\bar{x}_j$: j'inci örneklemin ortalaması

σ 'nın tahmin edicisi R yardımıyla bulunur (Şenol, 2012).

$$R = x_{\max} - x_{\min} \quad (10)$$

$R_1, R_2, \dots, R_k$ olmak üzere k tane değişim aralıklarının ortalaması eşitlik 11 ile hesaplanmaktadır (Şenol, 2012):

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_k}{k} \quad (11)$$

Örneklem ölçümüne göre belirlenen $Z_{\alpha/2} \sigma$ sınırlarının hesaplanmasında kullanılan çarpanlar Tablo 1’de verilmektedir.

Örneklem normal dağılıyorsa R ve standart sapma oranlanarak, örneklem büyüklüğü ile ilişkili, $w = \frac{R}{\sigma}$ göreceli etkinlik olarak adlandırılan rasgele bir değişken tanımlanmaktadır. σ ’nın yansız tahmincisi olarak $n \leq 10$ ise R, $n > 10$ ise s standart sapma kullanılmaktadır (Şenol, 2012).

$$n \leq 10 \text{ ise } \frac{\bar{R}}{d_2} \quad E\left(\frac{\bar{R}}{d_2}\right) = \sigma \quad (12)$$

$$n > 10 \text{ ise } \frac{\bar{s}}{c_4} \quad E\left(\frac{\bar{s}}{c_4}\right) = \sigma \quad (13)$$

Eşitlik 12 için σ ’nın yansız tahmincisi $\hat{\sigma} = \frac{\bar{R}}{d_2}$ ile hesaplanmaktadır (Şenol, 2012).

$E\left(\frac{\bar{R}}{d_2}\right) = \sigma$ ya da $E\left(\frac{\bar{R}}{\sigma}\right) = d_2$ olmak üzere, $\bar{X}$ kontrol sınırları aşağıdaki gibi hesaplanmaktadır (Şenol, 2012):

$$M\bar{C} = \bar{X} \quad (14)$$

$$KS = \mu \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \cong \bar{X} \pm Z_{\alpha/2} \frac{\bar{R}/d_2}{\sqrt{n}} = \bar{X} \pm Z_{\alpha/2} A_2 \bar{R}$$

Eşitlik 13 için σ ’nın yansız tahmincisi $\hat{\sigma} = \frac{\bar{s}}{c_4}$ ile hesaplanmaktadır. $\bar{X}$ kontrol sınırları aşağıdaki eşitlik ile oluşturulmaktadır (Şenol, 2012):

$$M\bar{C} = \bar{X} \quad (15)$$

$$KS = \mu \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \cong \bar{X} \pm Z_{\alpha/2} \frac{\bar{s}/c_4}{\sqrt{n}} = \bar{X} \pm Z_{\alpha/2} A_3 \bar{s}$$

Alt örneklem grubunun standart sapması ($i=1, 2, 3, 4, \dots, k$) aşağıdaki gibi hesaplanmaktadır:

$$s_i = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}} \quad (16)$$

$$\bar{s} = \frac{\sum s_i}{k} \quad (17)$$

Değişim aralığı (R) ve Standart Sapma (S) Kontrol Grafikleri

Rasgele değişken olan R ki-kare dağılımına ve S normal dağılıma sahiptir. R ve S kontrol grafikleri süreç standart sapmasındaki sapmayı belirlemek için kullanılmaktadır. σ_0 proses standart sapması $\alpha=0,0027$ anlamlılık düzeyinde aşağıdaki gibi kurulan hipotezler ile test edilmektedir (Şenol, 2012):

$$H_0: \sigma = \sigma_0 \quad (18)$$

$$H_1: \sigma \neq \sigma_0$$

Değişim Aralığı (R) Kontrol Grafiği

R kontrol grafiği için kontrol sınırlarının genel eşitliği (Şenol, 2012):

$$\mu_R \pm Z_{\alpha/2} \sigma_R \quad (19)$$

şeklinde. μ_R 'ın tahmin edicisi $\bar{R}$ aşağıdaki eşitlik ile hesaplanmaktadır.

$$\mu_R = \bar{R} = \frac{1}{k} \sum_{j=1}^k R_j \quad (20)$$

σ_R , değişim aralığı örnekleme dağılımının standart sapmasıdır. Süreç değişkeninin normal dağılıma uyduğu varsayımı ile göreceli değişim aralığı $w = \frac{R}{\sigma}$ dağılımından hesaplanmaktadır. w 'nın standart sapması n 'in bilinen faktörlerinden d_3 alındığında $R = w\sigma$ olduğundan değişim aralığının standart sapması eşitlik 2.19 ile hesaplanmaktadır (Şenol, 2012).

$$\sigma_R = d_3 \sigma \quad (21)$$

Anakütle parametreleri bilindiğinde, R kontrol grafiğinin parametreleri için değişim aralığının dağılım ortalaması d_2 ve standart sapması d_3 olarak alınarak R'nin standart sapması eşitlik 21 ile hesaplanmaktadır. Kontrol sınırları için hesaplamalar aşağıdadır (Şenol, 2012):

$$M\check{C} = d_2 \sigma \quad (22)$$

$$KS = d_2 \sigma \pm Z_{\alpha/2} d_3 \sigma$$

Başka bir ifadeyle $d_2 \sigma + Z_{\alpha/2} d_3 \sigma$ ya da $\sigma(d_2 \pm Z_{\alpha/2} d_3)$ olarak düzenlendiğinde, Tablo 1'de verilen $D_1 = d_2 - Z_{\alpha/2} d_3$ ve $D_2 = d_2 + Z_{\alpha/2} d_3$ sabitlerine dönüştürülmektedir. Kontrol sınırları yeniden düzenlendiğinde alt (AKS) ve üst kontrol sınırları (ÜKS) aşağıdaki eşitlik ile hesaplanmaktadır (Şenol, 2012).

$$\check{U}KS = d_2 \sigma + Z_{\alpha/2} d_3 \sigma = D_2 \sigma \quad (23)$$

$$AKS = d_2\sigma - Z_{\alpha/2}d_3\sigma = D_1\sigma$$

Anakütle parametreleri bilinmediğinde σ 'nın tahmincisi $\frac{\bar{R}}{d_2}$ ve σ_R 'nin tahmincisi $\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2}$ olmaktadır. Kontrol sınırları eşitlik 24 ile hesaplanmaktadır (Şenol, 2012).

$$M\check{C} = \hat{\mu}_R = \bar{R} \quad (24)$$

$$\check{ÜKS} = \hat{\mu}_R + Z_{\alpha/2}\hat{\sigma}_R = \bar{R} + Z_{\alpha/2} \frac{d_3}{d_2} \bar{R} = D_4 \bar{R}$$

$$AKS = \hat{\mu}_R - Z_{\alpha/2}\hat{\sigma}_R = \bar{R} - Z_{\alpha/2} \frac{d_3}{d_2} \bar{R} = D_3 \bar{R}$$

$D_4 = 1 + 3 \frac{d_3}{d_2}$ ve $D_3 = 1 - 3 \frac{d_3}{d_2}$ değerleri örneklem büyüklüğüne bağlı olarak değişmektedir (Şenol, 2012). $n \leq 6$ olduğunda D_3 eksi değer almakta; fakat R 'nin pozitif bir değer olmasından dolayı $D_3 = 0$ alınır. Çarpanlar Tablo 1'de gösterilmiştir (Şenol, 2012).

Standart Sapma (S) Kontrol Grafiği

R parametresi örneklem standart sapmasına göre kolay hesaplanabildiği için daha çok tercih edilmektedir (Birgören, 2015). Anakütlenin ortalaması μ ve standart sapması σ bilindiğinde n birimlik $x_1, x_2, x_3, \dots, x_n$ şeklinde alınan örneklem standart sapması $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ eşitliği ile hesaplanmaktadır (Şenol, 2012).

$E(s^2) = \sigma^2$; fakat $E(s) \neq \sigma$ 'dır (Montgomery, 2001). Buradaki sorun $E(s) = c_4\sigma$ bir katsayı ile düzeltilmektedir. (Birgören, 2015). Anakütle normal dağılım gösteriyorsa σ 'nın yansız tahmincisi $c_4\sigma$ 'dır. c_4 değeri Tablo 1'de belirtilmiştir.

S kontrol grafiğinin kontrol sınırları (Şenol, 2012):

$$M\check{C} = c_4\sigma$$

$$KS = c_4\sigma \pm Z_{\alpha/2}\sigma\sqrt{1 - c_4^2} \quad (25)$$

Farklı örneklem büyüklükleri için $B_5 = c_4 - 3\sqrt{1 - c_4^2}$ ve $B_6 = c_4 + 3\sqrt{1 - c_4^2}$ eşitliklerindeki B_5 ve B_6 sabitleri Tablo 1'de belirtilmiştir (Şenol, 2012).

Anakütle ortalaması μ ve standart sapması σ bilinmediğinde n birimlik k örneklem alındığında (Şenol, 2012);

s_i : i'inci örneklemin standart sapmasıdır, $\bar{s} = \frac{1}{k} \sum_{i=1}^k s_i$

σ 'nın yansız tahmincisi $\hat{\sigma} = \bar{s}/c_4$ 'dür. Kontrol sınırları $\bar{s} \pm Z_{\alpha/2} \hat{\sigma}_{\bar{s}}$ olmaktadır. S kontrol grafiğinin sınırları aşağıdaki gibidir (Şenol, 2012):

$$M\check{C} = \bar{s} \quad (26)$$

$$\check{U}KS = \bar{s} + Z_{\alpha/2} \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2} = B_4 \bar{s}$$

$$AKS = \bar{s} - Z_{\alpha/2} \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2} = B_3 \bar{s}$$

$$B_3 = 1 - \frac{Z_{\alpha/2}}{c_4} \sqrt{1 - c_4^2} \quad \text{ve} \quad B_4 = 1 + \frac{Z_{\alpha/2}}{c_4} \sqrt{1 - c_4^2} \quad \text{eşitlikleriyle}$$

hesaplanan B_3 ve B_4 sabitleri Tablo 1'de verilmiştir (Şenol, 2012).

$\bar{X}$, R ve S kontrol grafiklerinde prosesin kontrol dışı olduğu durumlar aşağıdaki gibi sıralanabilir:

- i) σ değişmemiş, μ değişmiş ise
- ii) μ değişmemiş, σ değişmişse
- iii) μ ve σ her ikisi de değişmiş ise.

Nicel ölçümler için merkez çizgi ile alt ve üst kontrol sınırları Tablo 3'te özetlenmiştir.

Tablo 3: Nicel ölçümler için Kontrol Sınırları

Kontrol Grafiği	$\bar{X}$ -Ortalama	R-Açıklık	S-Standart Sapma
μ ve σ biliniyorsa	$\check{U}KS = \mu + A\sigma$ $M\check{C} = \bar{X} = \mu$ $AKS = \mu - A\sigma$	$\check{U}KS = D_2\sigma$ $M\check{C} = \bar{R} = d_2\sigma$ $AKS = D_1\sigma$	$\check{U}KS = B_6\sigma$ $M\check{C} = \bar{s} = c_4\sigma$ $AKS = B_5\sigma$
μ ve σ bilinmiyorsa $\bar{X}$ ve $\bar{R}$	$\check{U}KS = \bar{X} + A_2\bar{R}$ $M\check{C} = \bar{X}$ $AKS = \bar{X} - A_2\bar{R}$	$\check{U}KS = D_4\bar{R}$ $M\check{C} = \bar{X} = \mu$ $AKS = D_3\bar{R}$	
μ ve σ bilinmiyorsa $\bar{X}$ ve $\bar{s}$	$\check{U}KS = \bar{X} + A_3\bar{s}$ $M\check{C} = \bar{X}$ $AKS = \bar{X} - A_3\bar{s}$		$\check{U}KS = B_4\bar{s}$ $M\check{C} = \bar{s}$ $AKS = B_3\bar{s}$

X ve Hareketli Aralık (Moving Range- MR) Kontrol Grafikleri

Tekrarlanması mümkün olmayan durumlarda belirtilen koşulun sağlanması için tek bir ölçüm yeterli olabilmektedir. Sıcaklık, basınç, kaza raporları, medikal testlerin kayıtları, kimyasal analiz sonuçları gibi durumlar bu vakaları ifade edebilmektedir. Bir dizi tekli ölçümler ardışık olarak kaydedilerek aralarındaki fark alınmaktadır. İkinci ve birinci değer arasındaki

fark daha sonra üçüncü ve ikinci arasındaki fark şeklinde devam ederek hareketli aralık (moving range- MR) değeri türetilmektedir. MR değerinin ortalamasını ($\overline{MR}$) ifade etmektedir (Western Electric Company, Inc., 1958).

$$MR_{i+1} = |x_{i+1} - x_i|, \quad i=1,2,\dots,k \quad (27)$$

X-MR grafiğinin işlemleri (Western Electric Company, Inc., 1958):

- i) Mümkün ise süreçten 20 veya daha fazla, mümkün değil ise en az 10 adet tekli ölçümler alınarak kaydedilir.
- ii) Ölçüm sayısının 1 eksiği kadar MR değeri oluşmaktadır.
- iii) Ölçümlerin ortalaması ($\bar{X}$) alınarak merkez çizgi belirlenmektedir.
- iv) MR değerlerinin ortalaması alınır.
- v) Hesaplanan $\overline{MR}$ değeri 2,66 sabit değerle çarpılarak kontrol sınırlarının genişliği bulunur.
- vi) Kontrol sınırları $\bar{X} \pm 2,66\overline{MR}$ ile hesaplanmaktadır.

$n = 2$ gözlemler için hareketli aralığın hesaplanması gerekiyor ise Tablo 1'de belirtilen $d_2 = 1,128$ sabiti kullanılır (Montgomery, 2009).

$\hat{\mu} = \bar{X}$ ile $\hat{\sigma} = \overline{MR} / 1,128$ ile tahmin edilmektedir (Birgören, 2015). $\bar{X}$ grafiği için kontrol sınırları (Montgomery, 2009):

$$MÇ = \bar{X} \quad (28)$$

$$KS = \bar{X} \pm 3 \frac{\overline{MR}}{d_2} = \bar{X} \pm 2,66\overline{MR}$$

MR grafiğinin kontrol sınırları (Montgomery, 2009):

$$MÇ = \overline{MR} \quad (29)$$

$$\ddot{ÜKS} = D_4 \overline{MR} = 3,267\overline{MR}$$

$$AKS = D_3 = 0$$

P Kontrol Grafiği

Bir süreçte iki duruma sahip bir oluşumun oranını açıklamak için Shewhart P grafiği kullanılmaktadır (Mohammed ve ark., 2013). Örneklem hacminin

büyük olduğu nadir vakalarda da Shewhart P kontrol grafiği ile proses incelenebilmektedir (Stapenhurst, 2005).

Süreç kusurlu oranı p bilinmediğinde, sabit ya da farklı örneklem hacmine sahip gözlenen verilerden hesaplanmalıdır. Genel kural örneklem sayısının en az 20- 25 olmasıdır (Montgomery, 2009).

$$p_i = \frac{D_i}{n_i} \quad i = 1, 2, \dots, k \quad (30)$$

D_i : i. örneklemdeki kusurlu sayısı

n_i : i. örneklem hacmi

p_i : i. örneklemdeki kusurlu oranı

Ortalama kusurlu oranı eşitliği aşağıda belirtilmektedir (Montgomery, 2009).

$$\bar{p} = \frac{\sum_{i=1}^k D_i}{\sum_{i=1}^k n_i} \quad (34)$$

Shewhart P grafiğinin kontrol sınırları (Montgomery, 2009):

$$M\check{C} = \bar{p} \quad (32)$$

$$KS = \bar{p} \pm Z_{\alpha/2} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_i}}$$

eşitliği ile hesaplanmaktadır.

$\pm Z_{\alpha/2} \sigma$ sınırları normal dağılım varsayımına uymaktadır. n büyükse ya da $np > 5$ ise binom dağılımı normal dağılıma yaklaşmaktadır. Binom dağılımının normal dağılıma yeterince yaklaşmamasından dolayı bazen alt kontrol sınırları negatif değerler alabilmekte; fakat kusurlu oranının negatif bir değer alması anlamsız olduğundan AKS sıfır olarak gösterilmektedir (Birgören, 2015).

NP Kontrol Grafiği

NP grafiğinde örneklem hacmi sabittir. Kusurlu oranı yerine kusurlu sayısı alınmaktadır (Stapenhurst, 2005). NP grafikleri P grafiği ile benzer uygulamaya sahip olmasına rağmen kullanımı kolay olduğundan daha çok tercih edilmektedir (Montgomery, 2009). Kontrol sınırları aşağıdaki eşitlik ile hesaplanmaktadır:

$$M\check{C} = np \quad (33)$$

$$KS = np \pm Z_{\alpha/2} \sqrt{np(1-p)}$$

C Kontrol Grafiği

C kontrol grafiği sabit örneklem büyüklüğüne sahip durumlarda kusur sayısını izlemek için kullanılmaktadır. Birçok durumda süreç tek bir örneklemden oluşmaktadır. Poisson dağılımına uyan C kontrol grafiği için kontrol sınırları aşağıdaki eşitlikte verilmiştir (Jones ve Govindaraju, 2000). $c \geq 5$ olduğunda poisson dağılımı, normal dağılıma yaklaşmaktadır (Birgören, 2015).

D_i : i 'inci olaydaki kusur sayısı ya da i 'inci birimdeki kusur sayısı

k : örneklem sayısı

$$\bar{c} = \frac{\sum_{i=1}^k D_i}{k} \quad i = 1, 2, \dots, k$$

$$M\check{C} = \bar{c} \quad (34)$$

$$KS = \bar{c} \pm Z_{\alpha/2} \sqrt{\bar{c}}$$

U Kontrol Grafiği

U kontrol grafiği poisson dağılımına uymaktadır. Farklı örneklem büyüklüklerine sahip her bir alt örneklem grubunun belli bir zaman dilimindeki kusur sayılarını izlemek için kullanılmaktadır. Alt grup büyüklükleri birbirinden farklı olabileceği için kontrol sınırları da buna göre değişkenlik göstermektedir (Benneyan, 1998).

D_i : i 'inci nitel özelliğin kusur sayısı

k : örneklem sayısı

n : alt grup örneklem hacmi

$$\bar{u} = \frac{\sum_{i=1}^k D_i}{\sum_{i=1}^k n_i} \quad i = 1, 2, \dots, k$$

$$M\check{C} = \bar{c} \quad (35)$$

$$KS = \bar{u} \pm Z_{\alpha/2} \sqrt{\frac{\bar{u}}{n_i}}$$

Nitel Özellikler için Geliştirilen Kontrol Grafikleri

İPK'nın birincil amacı olağan değişim ile olağandışı değişimi belirlemektir. Doğal sebeplerden kaynaklı değişimi düzeltmek için süreçte bir aksiyon alınmalıdır. Olağandışı değişimde ise ilk önce buna sebep olan etkenin araştırılması gerekmektedir. Nitel ya da sayılabilen ölçümler onbinler, yüzbinler ve daha fazlası gibi çok büyük örneklem hacmine sahip olabilmektedir. Böyle durumlarda Shewhart kontrol grafiklerinin kontrol limitleri dar olduğundan birçok nokta kontrol dışı görünmektedir. Bu durum doğal ve doğal olmayan değişkenlik ile ilgili yanlış izlenim vermekte ve hatalı eylem alınmasına neden olabilmektedir. Sağlık alanında büyük veri setlerinin artan kullanımı geleneksel nitel ölçümlerin kontrol grafiklerinin, alt grup içi değişimi dikkate almasından kaynaklanan problemler ortaya çıkmaktadır. Bu durumun çözülebilmesi için alt grup içi ve alt gruplar arası değişimin dikkate alınması gerekmektedir. Laney'in nitel ölçümlü grafiklerle ilgili yaklaşımı bu sorunu çözmektedir (Mohammed ve ark., 2013).

Geleneksel İPK grafikleri P ve U, büyük örneklem hacmine uygulandığında dar kontrol sınırları üreterek birçok noktayı kontrol dışı göstermektedir. Bu durum aşırı yayılım (overdispersion) olarak bilinmekte, ayrıca çok az değişim görülmesi de az yayılıma (underdispersion) yol açmaktadır. Aşırı yayılım görülmesi az yayılıma göre daha yaygındır. Büyük örneklem büyüklüklerinden kaynaklı nitel ölçümlü grafiklerdeki bu problem İPK'nın endüstri uygulamalarında dikkatle incelenirken, sağlık alanında çok az ilgi görmüştür (Mohammed ve ark., 2013).

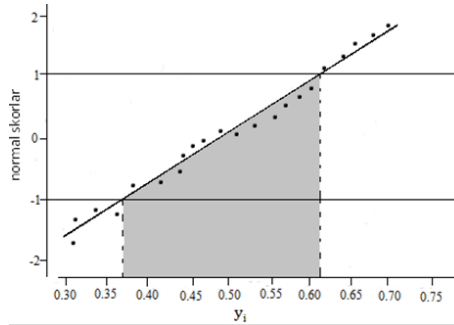
Süreç ölçüm verileri normal, binom veya poisson gibi bir olasılık dağılımı izliyor ise, $\bar{X}$, P ve C gibi geleneksel kontrol grafikleri (Shewhart kontrol grafikleri) sürecin istatistiksel olarak kontrol sınırları içinde olduğu varsayımına dayanmaktadır. İPK'nın bu tanımı bazı süreçlerde doğal veya kaçınılmaz herhangi bir rasgele olmayan değişime izin vermemektedir. Süreçteki rasgele olmayan değişim için otokorelasyon örneği verilebilir. Birçok makalede standart kontrol grafiği prosedürlerinin alt gruplar arası ekstra değişimin varlığı durumunda nasıl etkili olabileceği incelenmiştir. Box-Cox dönüşümünün nitel veriler üzerinde etkili olmadığı görülmekte, verilerin normal dağılıma dönüştürülmesi aşamasında uygunsuzluk sebepleri gizlenebilmektedir. Aşırı yayılım ve az yayılım durumunu düzeltmek için verilerin bazı yöntemlerle dönüştürülmesi yerine uygun grafik yöntem uygulanabilmektedir (Jones ve Govindaraju, 2000).

Geleneksel (Shewhart) P kontrol grafiği alt örneklemdeki uygunsuzluğun birbirlerinden bağımsız olduğunu varsaymaktadır. Sabit olmayan, değişken nitelikte olan olasılıklar aşırı yayılıma sebep olduğundan bu durumu test etmek için, binom dağılımındaki beklenen değişim ile verilerdeki gerçek değişimi karşılaştırmak gerekmektedir. d_i ; n_i alt grubunda i 'inci nitel özelliğin kusur sayısı olmak üzere veriler aşağıdaki eşitlikle ortalaması $\sin^{-1}\sqrt{p}$ ve varyansı $1/4n$ olan normal dağılıma uymaktadır (Jones ve Govindaraju, 2000);

$$y_i = \sin^{-1}\sqrt{(d_i + 3/8)/(n_i + 0,75)} \quad (36)$$

Dönüştürülmüş y_i değerleri ile normal dağılım grafiği çizilir. Şekil 2.4'te gösterildiği gibi Y ekseninin +1 ve -1 noktalarından yatay çizgiler çizilir. Bu çizgilerle normallik doğrusunun kesişim noktalarının y_i eksenindeki değerleri arasındaki fark alınarak gözlenen 2σ mesafesi hesaplanır. Şekil 3'te verilen örnek grafikte, gözlenen 2σ mesafesi $0,62 - 0,37 = 0,25$ 'dir (Jones ve Govindaraju, 2000).

$$\text{Beklenen } 2\sigma \text{ mesafesi} = 1/\sqrt{n} \quad (37)$$



Şekil 3: Kusurlu birimlerin dönüştürülmüş değerlerine karşı normal skorların grafiği (y_i)

Simülasyon çalışmalarının sonuçlarına dayanarak kabul görmüş kural olarak “beklenen 2σ mesafesi”, “gözlenen 2σ mesafesinin” yaklaşık 1,5 katı olmaktadır (Jones ve Govindaraju, 2000).

Az yayılım ya da aşırı yayılım görülmediği takdirde Shewhart (standart) P grafiği kullanılmalıdır. Aşırı ve az yayılım aşağıdaki ölçülere göre değerlendirilmektedir (Jones ve Govindaraju, 2000).

$$\frac{\text{Gözlenen } 2\sigma \text{ mesafesi}}{\text{Beklenen } 2\sigma \text{ mesafesi}} > 1,5 \text{ ise aşırı yayılım vardır.}$$

$$\frac{\text{Gözlenen } 2\sigma \text{ mesafesi}}{\text{Beklenen } 2\sigma \text{ mesafesi}} < 0,7 \text{ az yayılım vardır.}$$

$$0,7 > \frac{\text{Gözlenen } 2\sigma \text{ mesafesi}}{\text{Beklenen } 2\sigma \text{ mesafesi}} > 1,5 \text{ ise Shewhart P grafiği kullanılmalıdır.}$$

Bu tür bir durumla karşılaşmamak için İPK üretim sürecine uygulanırken genelde örneklem büyüklükleri küçük alınmaktadır (Laney, 2002). Büyük örneklem genişliklerinde sıklıkla görülen aşırı yayılım ve az yayılım durumu için David B. Laney bir çözüm geliştirmiştir.

Klasik kontrol sınırları tamamıyla örneklemdeki değişime bağlı olduğundan örneklem hacmi çok büyük olduğunda, limitler grafiğin merkez çizgisi yönünde sıkışmaktadır. Veri noktaları tahmin edilen olasılık dağılımını aşan bir değişim gösteriyorsa Jones ve Govindaraju'nun (Jones ve Govindaraju, 2000) belirttiği gibi aşırı yayılım söz konusudur (Laney, 2002).

Noktaların büyük çoğunluğu kontrol sınırları dışında ise özel sebeplerden bahsedilemez. Büyük örneklem bu duruma sebep olmayıp aksine ortaya çıkarmaktadırlar. Daha büyük örneklem boyutları tahminlerimizin örneklem hassasiyetini arttırmakta, böylece binom veya poisson varsayımının geçersizliğini daha net ortaya çıkarmaktadır (Laney, 2002).

Laney P grafiği grup içi ya da gruplar arası değişim ayırımı yapmadan verideki tüm değişimi ele almaktadır (Laney, 2002). Standart P grafiği için:

n_i : i 'inci alt grubun örneklem büyüklüğü ($i = 1, 2, \dots, k$)

d_i : ilgilenilen nitel özelliğin kusurlu sayısı

$$p_i = d_i/n_i \quad \bar{p} = \sum d_i / \sum n_i \quad \sigma_{p_i} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}} \quad (38)$$

$$KS = \bar{p} \pm Z_{\alpha/2} \sigma_{p_i} \quad M\check{C} = \bar{p}$$

Alt grup örneklem büyüklükleri farklı olduğunda kontrol sınırları da farklı olmaktadır (Laney, 2002). σ_{p_i} sabit $\bar{p}$ 'ya bağlı olarak grup içi değişimi göstermektedir (Mohammed ve ark., 2013).

Laney P grafiği için p değerleri z skorlarına dönüştürülerek ve bu noktalar ile z -grafiği çizilmektedir. z skorlarının ortalaması sıfır olduğundan grafiğin merkez çizgisi sıfır ve z 'nin standart sapmasının 1 olduğu varsayılmaktadır.

Böylece kontrol limitleri $\pm 3\sigma$ dan dolayı $+3$ ve -3 şeklinde olmaktadır (Laney, 2002).

$$z_i = \frac{p_i - \bar{p}}{\sigma_{p_i}} \quad (39)$$

$$M\check{C} = 0 \quad KS = \pm 3$$

Her bir birimin varyansı sabit olmayacağı için z-grafigi dönüşümü geliştirilerek, standart sapmasının 1 olduğu varsayımı kullanılmayıp gerçek değeri hesaplanmaktadır. Aşağıdaki eşitliklerle z'nin tekil grafigi çizilmektedir. Örneklem sayısı $n=2$ olduğundan Tablo 1'de verilen 1,128 yansız sabiti (d_2) kullanılmaktadır (Laney, 2002). z skorlarının hareketli açıklığı ile elde edilen σ_z gruplar arası değişimi belirtmektedir (Mohammed ve ark., 2013).

$$R_i = |z_i - z_{i-1}| \quad (i = 1, 2, \dots, k)$$

$$\bar{R} = \frac{1}{k-1} \sum_{i=2}^k R_i \quad \sigma_z = \frac{\bar{R}}{1,128} \quad (40)$$

$$M\check{C} = 0 \quad KS = \pm Z_{\alpha/2} \sigma_z$$

Son olarak z-dönüşümünü çözüme kavuşturmak için elde edilen bu sonuçlar p-düzlemine tekrar yerleştirilerek aşağıdaki eşitliklere ulaşılmaktadır (Laney, 2002):

$$z_i = \frac{p_i - \bar{p}}{\sigma_{p_i}} \quad sd(p_i) = \sigma_{p_i} \sigma_z \quad (41)$$

$$M\check{C} = \bar{p} \quad KS = \bar{p} \pm Z_{\alpha/2} \sigma_{p_i} \sigma_z$$

Burada σ_z 'nin gerçek değeri görülmektedir (Laney, 2002).

Poisson dağılımına sahip bir süreç için $\sigma_{p_i} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}$ yerine $\sigma_{u_i} = \sqrt{\frac{\bar{u}}{n_i}}$ kullanılarak Laney U grafigi çizilmektedir (Laney, 2002).

CUSUM ve EWMA Kontrol Grafikleri

Süreç parametrelerindeki büyük değişimlerin belirlenmesinde Shewhart kontrol grafikleri yeterli olmasına rağmen küçük değişimleri belirlemek konusunda yetersiz kalmaktadır. Bu nedenle 1954 yılında Page tarafından Birikimli Toplam Kontrol Grafigi (Cumulative Sum Control Charts - CUSUM - CUSCC) geliştirilmiştir. CUSUM grafikleri geçmiş örneklemelerden gelen bilgileri en son alınan örneklemelerdeki bilgilerle birleştirilerek istatistiksel olarak grafige dönüştürmektedir. Süreç karakteristiğinin örneklem değerinin, hedeflenen değerden zaman içindeki sapmalarının kümülatif toplamı ile grafik çizilmektedir (Koshti, 2011). Ortalama için kayma büyüklüğü $1,5\sigma$ ile 2σ arasında ya da daha fazla ise Shewhart kontrol grafigi daha etkilidir

(Montgomery, 2009). Bazı bilimsel makalelerde ortalamadaki küçük değişimlerin belirlenmesinde $\bar{X}$ grafiği yerine CUSUM kontrol grafiğinin kullanılmasının daha etkili sonuçlar vereceği belirtilmektedirler (Koshti, 2011).

Klasik kontrol grafiklerine alternatif olarak 1959 yılında Roberts tarafından Üstel Ağırlıklı Hareketli Ortalama (Exponentially Weighted Moving Average- EWMA) kontrol grafiği geliştirilmiştir. CUSUM ile EWMA kontrol grafiklerinin birbirlerine yakın etkinliğe sahiptir (Şenol, 2012).

CUSUM Kontrol Grafikleri

Örneklem değerlerinin hedef değerden sapmalarının kümülatif toplamaları alınarak çizilen CUSUM kontrol grafiği, örneklem değerleri dizisinin tüm bilgilerini kapsamaktadır. Örneğin örneklem hacmi $n \geq 1$ olan, $\bar{x}_j$ j. örneklem ortalamasıdır. Süreç ortalaması için hedef değer μ_0 ise, i. örnekleme karşılık CUSUM kontrol grafiğinin çizilmesi aşağıdaki eşitlik ile yapılmaktadır (Montgomery, 2009);

$$C_i = \sum_{j=1}^i (\bar{x}_j - \mu_0) \quad (42)$$

C_i , i. örneklemin de dahil olduğu kümülatif toplam miktarı.

Birçok örneklemden gelen bilgileri birleştirerek süreçteki küçük değişimlerin saptanmasında CUSUM grafikleri Shewhart grafiklerine göre daha etkili olmaktadır. Özellikle örneklem büyüklüğü $n=1$ olduğu durumlarda da etkilidir (Montgomery, 2009).

Page tarafından 1954 yılında tasarlanan kümülatif toplam grafikleri (CUSUM), Tabular (Algoritmik) ve V-Maskesi olarak iki şekilde gösterilmektedir (Montgomery, 2009).

Tabular (Algoritmik) CUSUM

CUSUM grafikleri hem bireysel gözlemler hem de alt grup ortalamaları için oluşturulabilmektedir. x_i (süreçteki x_i gözlem değerleri) süreç kontrol altında olduğunda μ_0 ortalamalı σ standart sapmalı normal dağılıma sahiptir. μ_0 , x süreç karakteristiği için hedef değer olarak kabul edilmektedir. Süreçte hedef değerden sapmalar olduğunda CUSUM uyarı vermekte bu durumda ayarlama yapılan değişken üzerinde bazı değişiklikler yapılarak süreç tekrar kontrol altına alınabilmektedir (Montgomery, 2009).

Tabular (algoritmik) CUSUM μ_0 hedef değerden sapmaların birikimiyle oluşturulmaktadır. μ_0 'ın üstündeki sapmaların birikimi C^+ ile, μ_0 'ın altındaki

sapmaların birikimi C^- ile gösterilmektedir. C^+ ve C^- 'nin hesaplaması aşağıdaki eşitlik ile belirtilmiştir (Montgomery, 2009)

$$\begin{aligned} C_i^+ &= \max[0, x_i - (\mu_0 + K) + C_{i-1}^+] \\ C_i^- &= \max[0, (\mu_0 - K) - x_i + C_{i-1}^-] \end{aligned} \quad (43)$$

Başlangıç değeri: $C_0^+ = C_0^- = 0$ alınmaktadır (Montgomery, 2009).

K referans değer olarak adlandırılmaktadır. Genellikle μ_1 kontrol dışı değerlerin ortalaması ile μ_0 hedef değer ortası olarak seçilmektedir. Sapmalar birimlerin standart sapması ile ifade edilecek olursa $\mu_1 = \mu_0 + \delta\sigma$ (veya $\delta = |\mu_1 - \mu_0|/\sigma$), böylece K değeri değişim büyüklüğünün yarısı olmaktadır (Montgomery, 2009).

$$K = \frac{|\mu_1 - \mu_0|}{2} = \frac{\delta}{2} \sigma \quad (44)$$

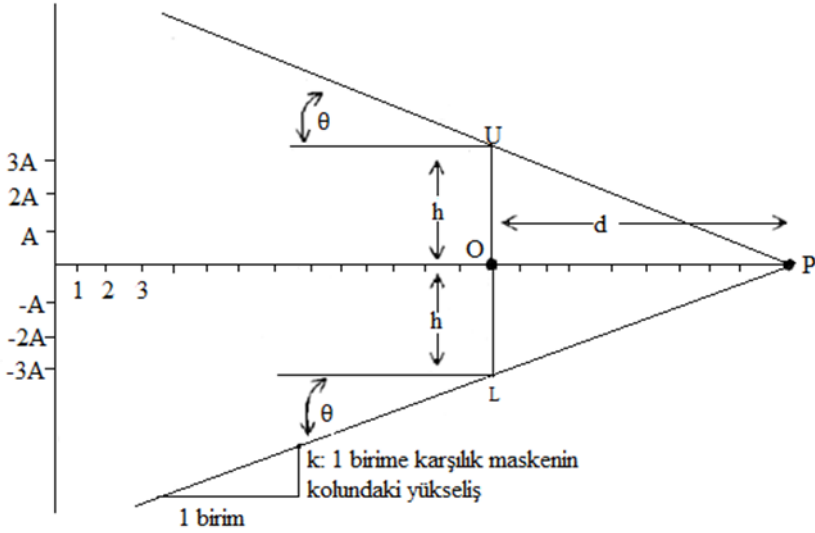
C_i^+ ve C_i^- istatistiklerinden herhangi birisi H karar aralığını aşarsa sürecin kontrol dışında olduğu düşünülmektedir. $H = 5\sigma$ kabul edilmektedir (Montgomery, 2009).

V-Maskesi Kontrol Grafiği

V-maskesi kontrol grafiği 1959 yılında Barnard tarafından önerilmiştir (Nist, 2012). Şekil 2.4'te gösterilen V-Maskesi kontrol grafiği ile hedef değerden olan sapmalar hem artış hem azalış yönünde aynı anda denetlenmektedir (Şenol, 2012).

$y_i = \frac{x_i - \mu_0}{\sigma}$ standardize gözlemler olmak üzere V-maskesi istatistiği aşağıdaki eşitlik ile hesaplanmaktadır (Montgomery, 2009);

$$C_i = \sum_{j=1}^i y_j = y_i + C_{i-1} \quad (45)$$



Şekil 4. Klasik bir V-Maskesi kontrol grafiği (Oktay ve Özçomak, 2001)

Şekil 4'te gösterilen tipik V-maskesi grafiğinde OP doğrusu yatay eksene paralel ve O noktası son C_i değerini göstermektedir. Önceki tüm kümülatif toplamlar $C_1, C_2, \dots, C_i$ V maskesinin kolları arasında ise süreç kontrol sınırları arasındadır. Süreç de bir değişiklik yok ise noktalar yatay çizgiye paraleldir (Şenol, 2012). V-maskesi dizayn parametreleri olarak adlandırılan k mesafesine (alt sınırın eğimi) ve h mesafesine (yükselme mesafesi) bağlıdır. Ayrıca d ve tepe açısı ($\theta = (\text{tepe açısı})/2$) da dizayn parametreleri olarak kullanılmaktadır. Dizayn parametreleri k ile h ve d ile θ aynı V-maskesini çizmelidir. V-Maskesinin performansına Şekil 2.4'te gösterilen d ve θ açısı ile karar verilir. V-maskesi ve Tabular CUSUM grafikleri eşitlik 2.44 sağlandığında aynı sonuçları vermektedir (Montgomery, 2009). A değeri, yatay eksen üzerindeki ardışık noktalar arasındaki mesafeye karşılık gelen dikey eksen üzerindeki birim mesafesidir (Montgomery, 2009; Oktay ve Özçomak, 2001).

$$k = A * \tan\theta$$

$$h = A * d * \tan(\theta) = d * k \quad (46)$$

$\bar{X}$ – CUSUM kontrol grafiği ile ortalamadan sapmaların tek veya çift yönlü değişimini test etmek için kurulan hipotezler aşağıdaki gibidir (Şenol, 2012);

$$\begin{array}{lll} H_0: \mu = \mu_0 & H_0: \mu = \mu_0 & H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 & H_1: \mu > \mu_0 & H_1: \mu < \mu_0 \end{array} \quad (47)$$

Grafik oluşturulurken takip edilmesi gereken sıra aşağıda belirtilmiştir (Şenol, 2012).

i. İlk önce örneklem ortalamalarının standart sapması tahmin edilir.

$\sigma = \frac{\bar{R}}{d_2}$ 'dir. (d_2 için bkzn. Tablo 1)

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \sigma_{\bar{x}} = \left(\frac{R}{d^2} \right) \sqrt{n} \quad (48)$$

ii. Ortalamada meydana gelmesi beklenen artı veya eksi yöndeki en küçük sapma miktarı D'dir.

$$\delta = \frac{D}{\sigma_{\bar{x}}} \quad (49)$$

iii. α anlamlılık seviyesi belirlenir.

iv. Birikimli toplamlar hesaplanır.

v. Y ölçeklendirme faktörü $\sigma_{\bar{x}} < y < 2\sigma_{\bar{x}}$ arasında ve genellikle $2\sigma_{\bar{x}}$ 'ya yakın seçilir.

vi. Tablo 4 ile δ yardımı ile d uzunluğu bulunur.

vii. δ hesaplanarak, Tablo 4'ten θ açısı bulunur.

viii. d ve θ açısı ile V-maskesi çizilir.

Tablo 4: $\bar{X}$ - CUSUM V-Maskesi için δ , α , d ve θ açısı değerleri (Şenol, 2012)

δ	θ	d-değerleri						
		$\alpha = 0,05$	$\alpha = 0,025$	$\alpha = 0,01$	$\alpha = 0,005$	$\alpha = 0,00135$	$\alpha = 0,001$	$\alpha = 0,000$
0,2	5 71	149,8	184,40	230,30	264,90	330,40	345,40	380,00
0,4	11 30	37,4	46,10	57,60	66,20	82,60	86,30	95,00
0,6	16 69	16,6	20,50	25,60	29,40	36,70	38,40	42,20
0,8	21 80	9,4	11,50	14,4	16,5	20,6	21,6	23,8
1	26 56	6	7,38	9,21	10,6	13,2	13,8	15,2
1,2	30 96	4,16	5,12	6,4	7,3	9,18	9,59	10,6
1,4	35 00	3,06	3,76	4,7	5,41	6,74	7,06	7,76
1,6	38 66	2,34	2,88	3,6	4,14	5,16	5,4	5,94
1,8	41 98	1,85	2,28	2,84	3,27	4,08	4,26	4,69
2	45 00	1,5	1,84	2,3	2,65	3,3	3,45	3,8
2,2	47 72	1,24	1,52	1,9	2,19	2,73	2,85	3,14
2,4	50 19	1,04	1,28	1,6	1,84	2,29	2,4	2,64
2,6	52 43	0,89	1,09	1,36	1,57	1,95	2,04	2,25
2,8	54 46	0,76	0,94	1,17	1,35	1,69	1,76	1,94
3	56 30	0,67	0,82	1,02	1,18	1,47	1,54	1,69

CUSUM kontrol grafiğinde maskenin oluşturulması da birikimli toplam değerlerinin düşey eksende yerleştirilmesi de kolay olmamaktadır. Her örneklem ortalamasından hedef ortalama çıkartılarak standart sapmaya

bölünmekte, standartlaştırılmış olan bu değerlerin birikimli toplamı alınmaktadır. Yatay ve düşey eksenlerde ölçeklendirme yapılarak, θ açısı hesaplanmaktadır. Ölçeklendirmeler aşağıdaki yaklaşımlardan birisi kullanılarak yapılmaktadır (Şenol, 2012).

i) Tip-I ölçeklendirmelerde yukarıda bahsedilen haliyle hiçbir değişiklik yapılmadan örneklem ölçümü k (alt grup sayısı) k -ekseninde bir birime karşılık düşey ekseninde $S_1, S_2, \dots S_i$ birikimli toplamı yer almaktadır (Şenol, 2012). Eşitlik 2.48 ile kontrol grafiğinin standartlaştırılmış değerleri, eşitlik 2.51 ile $S_1, S_2, \dots S_i$ birikimli toplamlar hesaplanmaktadır:

$$y_i = \frac{x_i - \mu_0}{\sigma} \quad (50)$$

$$S_i = \sum_{j=1}^i y_j = y_i + S_{i-1} \quad (i = 1, 2, \dots k) \quad (51)$$

Süreç ortalamasıyla bireysel gözlemlerin değerleri (ya da örneklem ortalaması) arasındaki sapmanın birikimli toplamı aşağıdaki eşitlik 52 ile hesaplanmaktadır (Şenol, 2012):

$$S_j = \sum_{j=1}^k (X_j - \bar{X}) \quad \text{ya da} \quad S_j = \sum_{j=1}^k (\bar{X}_j - \mu_0) \quad (52)$$

Örneklem ortalaması kullanıldığında, örneklem hacmi $n \geq 1$ için, j . örneklemin ortalaması $\bar{X}_j$, hedef ortalama μ_0 ve k örneklem sayısı ile

$$S_j = \sum_{j=1}^k (\bar{X}_j - \mu_0) = (\bar{X}_k - \mu_0) + \sum_{j=1}^{k-1} (\bar{X}_j - \mu_0) = (\bar{X}_j - \mu_0) + S_{j-1} \quad (53)$$

birikimli toplam değerleri hesaplanmaktadır (Şenol, 2012). S_j 'nin k . örnekleme kadar olan birikimli toplamı belirtilmektedir. $n=1$ ise $\bar{X}_j$ yerine X_j değeri kullanılır.

ii) Tip-II ölçeklendirmelerde k ekseninde bir birim alınırken buna karşın düşey ekseninde y birim alınmaktadır. Böylece d uzunluğunda değişme olmazken ($d_1 = d$), θ açısı $1/y$ kadar değişmektedir (Şenol, 2012).

$$\theta_1 = \tan^{-1} \left(\frac{\delta}{2y} \right) \quad (54)$$

iii) Tip-III ölçeklendirmelerde k ekseninde bir birim alınırken buna karşın dikey ekseninde y birim alınmaktadır (Şenol, 2012).

$$S_j = \sum_{j=1}^k (\bar{X}_j - \mu_0) \quad D = \mu_1 - \mu_0 = \delta \sigma_{\bar{x}} \\ d^* = -\frac{2}{\delta^2} \ln \alpha = d \quad (55)$$

$$\theta^* = \tan^{-1} \left(\frac{D}{2y} \right)$$

α 'nın değişmesi durumunda θ açısı değişmez; fakat d uzunluğu δ ve σ değerine bağlı olduğundan dolayı değişmektedir (Şenol, 2012).

V-Maskesi, d mesafesi ve θ açısı ile eşitlik 2.54 yardımıyla çizilmektedir:

$$d = E(\alpha) \left(\frac{\sigma}{D\sqrt{n}} \right)^2 = \frac{E(\alpha)}{\delta^2} \quad (56)$$

$$\tan\theta = \frac{D}{2y} = \frac{\delta\sigma/\sqrt{n}}{2y}$$

Tip-I hata olasılığı α 'nın $E(\alpha)$ ve 3σ limitleri için değerler Tablo 5'te verilmektedir. Ayrıca Tablo 4 kullanılarak d ve θ değerleri ile V-Maskesi oluşturulmaktadır (Şenol, 2012).

Tablo 5: $E(\alpha)$ ve 3σ limitleri için değerleri						
Tek yönlü test	$\frac{1}{2}\alpha$	0,00135	0,005	0,01	0,025	0,05
Çift yönlü test	α	0,0027	0,01	0,02	0,05	0,10
Beklenen değer	$E(\alpha)$	13,125	10,597	9,210	7,378	5,991

CUSUM kontrol grafikleri için ortalama gözlem sayısı (OGS)

$$OGS = \frac{d}{\delta \cot\theta - 1} \quad (57)$$

eşitliği ile hesaplanmaktadır. $OGS < 5$ ise sonuçlar güvenilir değil demektir (Şenol, 2012).

EWMA Kontrol Grafikleri

EWMA kontrol grafikleri de CUSUM kontrol grafikleri gibi küçük ve sürekli sapmaların belirlenmesinde diğer kontrol grafiklerine göre daha etkilidir. Geçmiş ve şu anki gözlemlerin ağırlıklı ortalaması ile hesaplamalar yapılmaktadır. EWMA normallik varsayımına duyarlı olmasından dolayı bireysel gözlemler için uygundur (Montgomery, 2009). EWMA kontrol grafiğinde aniden meydana gelen büyük bir değişimin hemen gözükmemesi bir dezavantajdır. EWMA kontrol grafiğinde değişimden önceki ve değişimden sonraki verilerin ortalamasının kullanılmasından dolayı aniden meydana gelen büyük bir değişim gözükmemektedir (Doğan ve Doğan, 2019). Örneklem hacmi $n > 1$ alınmalıdır. λ sabit bir değerdir. $0 < \lambda \leq 1$ aralığında bulunur. EWMA aşağıdaki eşitlik ile ifade edilmektedir (Montgomery, 2009):

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1} \quad (58)$$

Eşitlik 2.56'de i gözlem numarasını ya da zaman indeksini belirtmektedir. Başlangıç değeri z_0 çoğunlukla (ilk örneklem $i=1$) süreç hedef değeri olarak belirlenmekte, μ_0 ise kontrol grafiğinde hedeflenen değer olmaktadır (Doğan, 2019). $z_0 = \mu_0$ 'dır. Bazı süreçlerde giriş verisinin ortalaması EWMA'nın başlangıç değeri olarak kullanılmakta ve $z_0 = \bar{x}$ olmaktadır (Montgomery, 2009).

z_i 'nin önceki bütün örneklem ortalamalarının ağırlıklı ortalaması olduğunu göstermek için eşitlik 2.56'de z_{i-1} yerine

$$z_i = \lambda x_i + (1 - \lambda)[\lambda x_{i-1} + (1 - \lambda)z_{i-2}] \quad (59)$$

yerleştirilir. z_{i-j} için yerine koyulmaya devam edildiğinde $j = 2, 3, \dots, i$ için;

$$z_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i z_0 \quad (60)$$

eşitliği elde edilir. Ağırlıklar $\lambda(1 - \lambda)^j$ geometrik olarak azalmaktadır. Ağırlıklı toplam;

$$\lambda \sum_{j=0}^{i-1} (1 - \lambda)^j = \lambda \left[\frac{1 - (1 - \lambda)^i}{1 - (1 - \lambda)} \right] = 1 - (1 - \lambda)^i \quad (61)$$

eşitliği ile hesaplanmaktadır. $\lambda = 0,2$ ise, bu ağırlık mevcut örneklem ortalaması olarak kullanılır. Ayrıca önceki ortalamalara atanan ağırlıklar sırasıyla 0,16; 0,128; 0,1024 ve benzeridir (Montgomery, 2009).

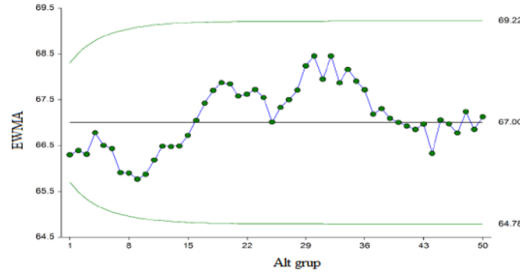
x_i gözlem değerleri σ^2 varyanslı bağımsız rasgele değişkenler ise, z_i 'nin varyansı eşitlik 62'de belirtilmektedir (Şenol, 2012);

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}] \quad (62)$$

Örneklem numarası i 'ye (veya zaman) karşı z_i 'ler işaretlenerek Şekil 5'te örnek olarak gösterilen EWMA grafiği oluşturulmaktadır. Merkez çizgi ve kontrol sınırları aşağıdaki eşitlik ile hesaplanmaktadır (Montgomery, 2009);

$$M\bar{C} = \mu_0 \quad (63)$$

$$KS = \mu_0 \pm L\sigma \sqrt{\frac{\lambda}{(2 - \lambda)}} [1 - (1 - \lambda)^{2i}]$$



Şekil 5. EWMA kontrol grafiği

Eşitlik 63'te görülen L faktörü kontrol limitlerinin genişliğini göstermektedir. $[1 - (1 - \lambda)^{2i}]$ ifadesi arttıkça 1'e yaklaşmaktadır. EWMA kontrol grafiği birçok zaman periyodu için çalıştırıldığında kontrol sınırları kararlı duruma yaklaşmaktadır. EWMA kontrol grafiği için kontrol sınır değerleri aşağıdaki eşitlikte belirtilmektedir (Montgomery, 2009);

$$\begin{aligned} M\check{C} &= \mu_0 \\ KS &= \mu_0 \pm L\sigma\sqrt{\frac{\lambda}{2-\lambda}} \end{aligned} \quad (64)$$

i değerleri küçük olduğunda eşitlik 2.61 kullanıldığında EWMA kontrol grafiğinin performansı oldukça artacaktır (Montgomery, 2009).

Yapılan çalışmalar sonucu genel olarak λ değerleri $0,05 \leq \lambda \leq 0,25$ aralığında alınır ise EWMA kontrol grafiğinin etkinliği yüksek olacaktır. Çoğunlukla $\lambda=0,05$; $\lambda=0,10$; $\lambda=0,20$ değerleri seçilmektedir. λ büyük olduğunda etkili sonuçlar vermekte ve genel kabul görmüş kural olarak $L=3$ (üç-sigma sınırları) kullanılmaktadır. $\lambda \leq 0,1$ gibi küçük bir değer ise L değerini 2,6 veya 2,8 aralığında alarak kontrol sınırlarını daraltmak avantajlı olacaktır (Montgomery, 2009).

SONUÇ

İPK sürecin zaman içindeki değişiminin kontrol sınırları dahilinde olup olmadığını istatistiksel yöntemler kullanarak izlemektedir. Tedavi veya girişim sonrasında gelişen komplikasyonları, enfeksiyonların seyri, ilaç yan etkileri gibi durumların zaman içinde değişimi incelenerek gerekli önlemlerin alınmasına yardımcı olmaktadır. Hastalıkların izlenmesi sonucunda uygulanan hastalık kontrol prosedürlerinin etkinliği incelenebilmektedir. Sanayi alanında yaygın olarak kullanılan İPK, sağlıkla ilgili süreçlerde de prosesin analiz edilmesi ve iyileştirilmesinde kullanılan önemli bir yöntem olmaktadır.

Kaynaklar

Kitap

- Çömlekçi, N. (1989). *Temel istatistik ilke ve teknikleri*. 1. Basım, Bilim Teknik Yayınevi, İstanbul, s: 1-505.
- Birgören B. (2015) *İstatistiksel kalite kontrolü*. 1. Basım, Nobel Akademik Yayıncılık, Ankara, s: 1-198.
- Montgomery, D. J. (2001). *Supplemental Text Material to Support Introduction to Statistical Quality Control* 4th Edition, John Wiley & Sons Inc., New York, pp:1-54.
- Montgomery, D. J. (2009). *Introduction to statistical quality control*. Sixth Edition, John Wiley & Sons Inc., Jefferson city, pp: 1-754.
- National Institute of Standards and Technology (NIST)/SEMATECH (2012) *e-handbook of statistical methods*, <http://www.itl.nist.gov/div898/handbook/>
- Stapenhurst T. (2005) *Mastering statistical process control*. First Published, Elsevier Butterworth-Heinemann, Oxford, pp: 1-415.
- Şenol Ş. (2012) *İstatistiksel kalite kontrol*. 1. Basım, Nobel Akademik Yayıncılık, Ankara, s: 1-439.
- Western Electric Company, Inc. (1958) *Statistical quality control handbook*, Second Edition, The Mack Printed Company, Easton, Pennsylvania, pp: 1-328.

Dergi Makalesi

- Benneyan J.C. (1998) Statistical quality control methods in infection control and hospital epidemiology part I: introduction and basic theory, *Infect Control Hosp. Epidemiol.* 1998; 19: 194-214.
- Benneyan J.C., Lloyd R.C., Plsek P.E. (2003) Statistical process control as a tool for research and healthcare improvement, *Qual Saf Health Care* 12: 458-464.
- Benneyan J.C. (2008) The design, selection, and performance of statistical control charts for healthcare process improvement, *International Journal of Six Sigma and Competitive Advantage* 4(4):209-239.
- Doğan İ., Doğan N. (2019) EWMA Kontrol Çizelgeleri ve Sağlık Alanında Kullanımına Genel Bir Bakış, *Türkiye Klinikleri J Biostat*, 11(1): 72-82.
- Jones G., Govindaraju K. (2000) A graphical method for checking attribute control chart assumptions, *Quality Engineering* 13(1): 19-26.
- Karkaloulos P. L., Evangelopoulos A., (2015) The History of Statistical Quality Control in Clinical Chemistry and Haematology (1950 – 2010). *International Journal of Biomedical Laboratory Science (IJBLS)*, Vo1. 4, No. 1:1-11.
- Koshti V.V. (2011) Cumulative sum control chart, *International Journal of Physics and Mathematical Sciences* 1: 28-32.
- Laney D.B. (2002) Improved control charts for attributes, *Quality Engineering* 14(4): 531-537.
- Mohammed M.A., Panesar J.S., Laney D. B. et al. (2013), Statistical process control charts for attribute data involving very large sample sizes: a review of problems and solutions, *BMJ Qual Saf* 22: 362-368.
- Oktay E., Özçomak M.S. (2001) CUSUM kontrol grafikleri ve bir uygulama, *İktisadi ve İdari Bilimler Dergisi* 15: 1-2.
- Thor J., Lundberg J., Ask J. et al (2007) Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care* 16(5): 387-399.