

DENEY DİZAYNININ PRENSİPLERİ¹

Deney dizaynları sahip oldukları özellikleri, yapılandırılmaları ve analizleri bakımından çok çeşitlidir. En basitinden en kompleksine kadar her dizayn belli bir rasyonele dayanır ve belli deneysel durumlarda uygulanır. Böyle olmakla beraber, deneylemenin ve deney dizaynlarının müşterek bazı temel prensipleri var ki bu prensipleri açıkça anlamak gerekir. Bu prensipler soruşturulan problemin formülasyonu, deney dizaynının seçimi, deneyin icrası, verinin analizi ve sonuçların yorumlanmasına dairdir. Burada, spesifik dizaynlara ait spesifik detaylara girmeden, bu prensipleri genel olarak tartışacağız.

Onaycı ve Keşifçi Deneyler

Pek çok deneyin şu manada bir keşifçi doğası vardır. Soruşturmacı belli bir prosesin çıkımına tesir eden faktörleri bulmakla ilgilenir. Mesela, belli bir tipteki pişirme kabının kırılabilirlik derecesine bir kimyasal bileşiğin yoğunluğunun, pişirme süresinin, fırının sıcaklığının, soğutma derecesinin ve basınç miktarının bireysel ve/veya ortak bir etkisinin olup olmadığıyla veya ne derece etkisi olduğuyula ilgileniyor olabilirsin. Burada takip edilecek apaçık prosedür bu faktörlerin “düzeylerini” değiştirmek ve çeşitli düzey kombinasyonlarının performanslarını karşılaştırmaktır. Bununla beraber, bunun tam olarak nasıl yapılacağı pek te öyle apaçık değildir. Deneyi yapmak pek çok kararı vermeyi gerektirir: mesela çeşitli faktörlerin düzeylerinin seçimi ve bunların sayısı ki tüm fizibil kombinasyonların bir alt kümesi mümkünce seçilecektir; deney düzeninin seçimi ki kısmen fiziksel koşullar, kısmen istatistiksel mülahazalarla belirlenir; performans ölçümünün seçimi; ve kastedilen maksada matuf neticelere varmak için en mütenasip istatistiksel analizin seçimi. Sonraki bölümlerde bu tipteki soruları öyle çok detaylıca ele alacağız ve altta yatan prensipleri tartışacağız ki bir soruşturmacı elindeki belirli bir problem için mütenasip kararları verebilsin.

Deneyleme esasen ardışık bir prosestir. Bir deney bir başka deneye götürür çünkü bir prosese nüfuz-u nazar edilir ve yeni sorular sorulur. Bir keşifçi deney, yukarda betimlendiği üzere, bir başka deneyce takip edilebilir ki biz bunu onaycı deney diye adlandırabiliriz. Mesela, keşifçi deneyden bulunan “en iyi” prosedürü yerleşik/müesses bir prosedür veya ürün ile karşılaştırmak ve bu prosesin veya ürünün eskisinden daha “iyi” olduğunu delillendirmek/kanıtlamak (“tesis etmek”) isteyebiliriz. Bu, bizatihi, önceki betimlediğimiz problemden daha iyi-tanımlı ve daha dar bir problemdir. Böyle olunca, bu, farklı dizayn mülahazalarını gerektirir. Mesela, deney tekrarlarının sayısı çok önemli olabilir öyle ki nihai istatistiksel analiz, yani istatistiksel test, belli bir arzulanan güce erişebilsin.

Onaycı deney fikrini farklı bir yönde geliştirmek için, “en iyi” prosedürü bulmuş olmalıyız ve onun sahip olduğu istatistiksel özellikleri, mesela proses kontrolüne matuf maksatlarla, tesis etmek isteyebiliriz. Biliyoruz ki proses koşulları farklılaşabilir ve bu yüzden aritmetik-ortalama performansı ve prosese ilişkin değişkenliği tesis etmek önemlidir. Tatmin edici olmayan değerler için bu, aktüel üretim prosesinin rafine edilmesine yol açabilir.

¹ Bu metin başlıca Hinkelman and Kempthorne (2008) Volum 1 Chapter 2 çevirisine dayanır. Daha hassasiyetle okumak isteyenler metnin orjinaline gidebilir.

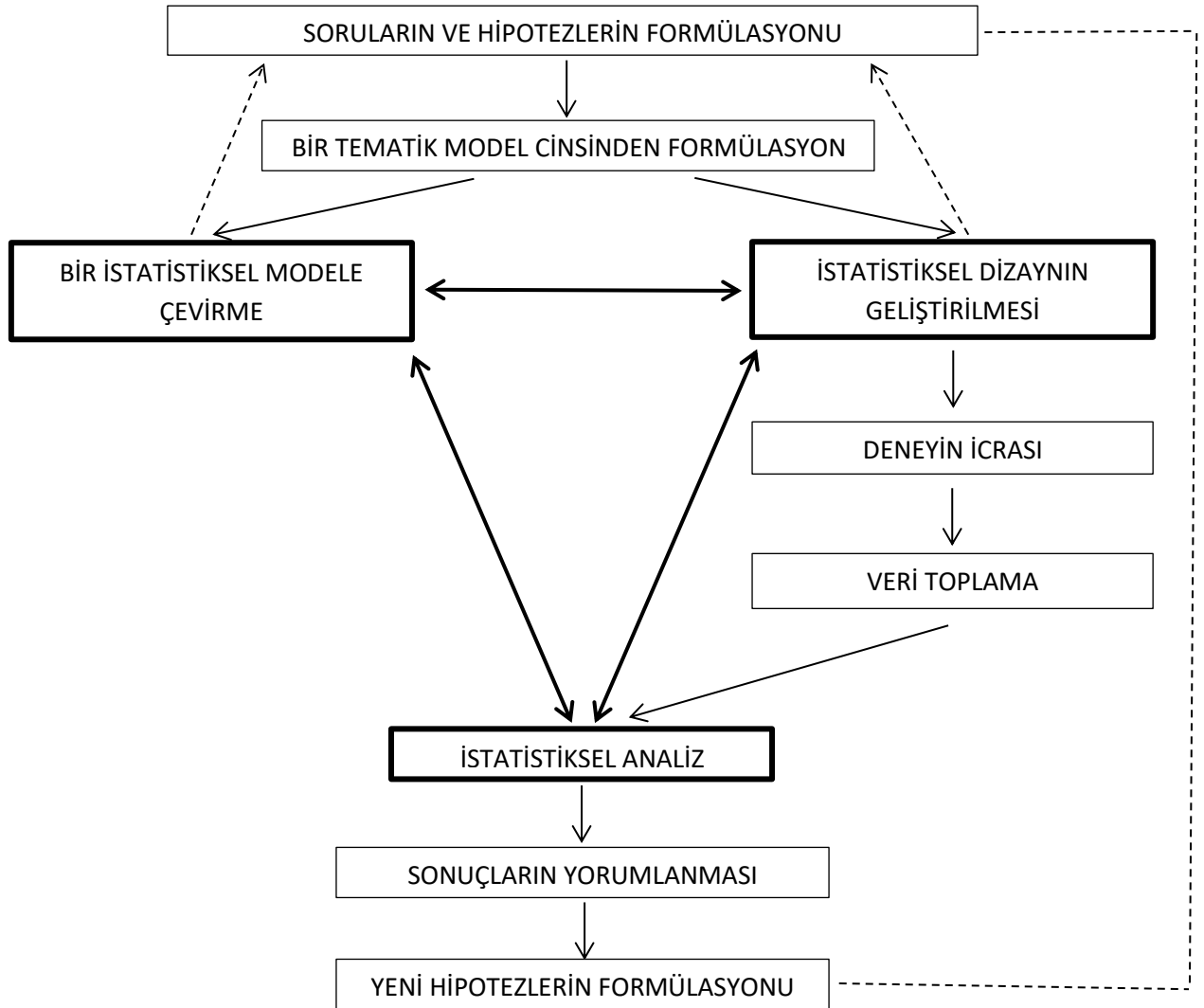
Hinkelman, K. and Kempthorne, O. (2008) Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design, Second Ed., Wiley.

Buraya kadarki tartışma pek açık olmadı ancak bunu kasten yaptım. Çünkü niyetim şimdilik deneylerin türleri hakkında bir fikir sahibini olmanı sağlamak. Kafanda soru işaretlerinin belirmesini bekliyorum. Başka soruşturma alanlarındaki benzer deneyler hakkında düşünmeni ve onları deneylemenin her bir adımına mümkün olduğunca taşınmanı istiyorum ki bu adımları müteakip kısımlarda ana hatlarıyla ortaya koyacağız.

Dizayn Edilmiş Soruşturmaların Adımları

İşin pratiğinde pek çok bilimsel ve endüstriyel soruşturmaların başarısızlık kaderidir. Bunun pek çok ve çeşitli sebepleri vardır fakat en çok karşılaşılan sebep soruşturmanın münasipçe planlanmamış olmasıdır. Çoğu soruşturmacı şunu anlamakta zorlanır ki dikkatle önceden-planlama başarılı bir deney için elzemdir. Bu belli adımları takip etmeyi ve aktüel soruşturmaya başlamadan önce her noktada kararlar vermeyi kapsar.

Bilimsel ve endüstriyel deneylemenin mantıksal adımlarını gösteren bir şematik sunum şu şekilde yapılabilir:



Müteakip kısımlarda her bir adımı yorumlayacağız ve prosesin bütünü içinde bu adımların önemini açıklayacağız (böyle bir yaklaşımın endüstriyel deneyler için alternatif betimlemesi için Coleman and Montgomery, 1993 e bak). Bu adımlar iki kategoriye ayrılabilir: (1) istatistiksel (yani, istatistiksel dizaynın geliştirilmesi, bir istatistiksel modele çevrilmesi, ve istatistiksel analiz ki biz bunları bilimsel deneylemenin mantıksal adımlarını temsil eden şemada koyu hatlarla gösterildiği gibi “istatistiksel üçgen” diye anarız) ve (2) doğası gereği istatistiksel-olmayan. Bu derste her ne kadar deneylemenin saf istatistiksel taraflarına yoğunlaşacak olsak ta, şunu realize etmek önemlidir ki istatistiksel-olmayan adımlar istatistiksel adımlarla yakınca bağlantılıdır ve mevzubahis bilimci/soruşturmacı ile istatistikçi arasında interaksiyonu gerektirirler ve hiç bir deney dizaynı tartışmasında göz ardı edilmemelidirler.

Problemin İfadesi

Soruşturma çoğu zaman basit bir spekülasyon ile başlar: “Bahçemdeki ağaçlardan biri kuruyor; acaba daha fazla suya mı gereksinim duyuyor?” Bunun akabinde sorular gelir, şöyle ki “ağaç ne kadar sulanmalıdır ve ne sıklıkta? veya “giderilmesi gereken başka yetersizlikler var mı ve eğer varsa nasıl?” Daha bilimsel bir düzeyde, bu türden spekülasyonlar ve sorular bir problemin formülasyonuna götürür: “Ağacın şimdiki durumunu iyileştirecek en iyi tedaviyi belirlemek istiyorum” veya daha bu dersin mevzusu dahilinde, “Ağacın tedavisi için alternatif prosedürlerin etkinliklerini karşılaştırmak istiyorum.”

Bu aşık diyebilirsin belki ama her bilimsel soruşturma bir problemin geliştirilmesi ve bir ifadesi ile başlamalıdır. Bu yalnızca sonraki istatistiksel mülahazalar bakımından değil, fakat aynı zamanda problemi gerçekçi olarak ele alınabilir bir probleme sınırlandırmak bakımından da önemlidir. Ne yazık ki çoğu zaman deneyler açıkça formüle edilmiş bir soru, zihinde bir maksat veya amaç olmaksızın başlar ve ne yazık ki çoğu zaman kafaya çok geç dank eder ki böyle bir deneye içinden çıkılamayacak kadar geniş bir bazda girişilmiştir ve yayılmıştır öyle ki deneyin aktüel olarak yürütülmesinde bu çoğu zaman pratik zorluklara yol açar. Böyle olunca, deney yürütülürken deneyi sınırlandırma yoluna gidilir ve bu da tatmin edici olmayan bir deneyle sonuçlanır ki bu deney bilimsel araştırmacının en önemli sorusunu/sorularını cevaplayabilemez. Apaçık, deneyeyle ilişkin diğer mülahazalar da sıralarını beklemektedirler. Bunları, aşağıda, deneysel soruşturmalar için akla uygun stratejilerin belirlenmesinde bir kılavuz olarak ana hatlarıyla vereceğiz.

Tematik Model

Birinci adım, problemin ifadesi, soruların ve bilimsel-araştırmanın hipotezinin formülasyonuna denktir. Yukarıda zikredildiği üzere, böyle hipotezler mümkün olduğunca açık ifade edilmelidir ancak formülasyon her zaman bir istatistiksel hipotezin kesinliği kadar kesin olamayabilir. “Y muamelesini (prosedürünü) X muamelesi (prosedürü) ile karşılaştırmak istiyorum” veya “bir kaç muameleyi karşılaştırmak istiyorum” veya “Bir prosesin belli bir çıkımı üzerinde hangi faktörlerin etkisi olduğunu bulmak istiyorum” gibi ifadeler genelde ziyadesiyle mütenasiptir. Bu şimdi bizi açık bir yoldan bir *tematik modelin* formülasyonuna götürecektir. Model ile kastettiğimiz deneyin çıkımına tesir edebilecek tüm faktörlerin bir listesidir. Böyle faktörlerin kapsamında muamele faktörleri olacaktır ki bunlar soruşturmanın ana ereğidir ve bunların yanında sınıflandırma (veya bloklama) faktörleri olacaktır ki bunlar da altında deneyin icra

edildiği koşullar tarafından belirlenirler. Böyle bir liste soruşturmanın tamamı için yaşamsal öneme sahiptir çünkü: (i) ağırlıklı olarak deney dizaynlarının seçimini belirler ve (ii) hedef popülasyonu tanımlar ki çıkarımlarımızı bu popülasyona göre yaparız.

Bu fikirleri müteakip misallerle göstereceğim.

Misal 1 Farzedelim ki bir kronik kalp yetmezliği rastgeleleştirilmiş klinik denemesi planlıyoruz (Hryniewicz et al., 2003). Bu öğrenme-çalışmasının ereği carvedilol ve metoprololün adrenerjik uyarana karşı bölgesel vazküler yanıtlar üzerine etkisini soruşturmak. Nihai deneme protokolüne karar verirken, muamele faktörüne ilaveten şu faktörlerin de dikkate alınması gerekebilir: deneklerin cinsiyeti; deneklerin tipi, yani normal denekler, New York Kalp Derneği sınıf I, II, III hastalar; önceki veya devam eden muamelenin/tedavinin tipi; deneklerin yaşı.

Misal 2 Gece paniği için kognitif davranışsal terapinin etkisini değerlendirmek ereğiyle yürütülecek bir öğrenme-çalışmasında (Craske et al., 2005), ilave faktörler şunları kapsayabilir cinsiyet, ırk, medeni durum, eğitim-öğretim düzeyi, istihdam durumu.

Misal 3 Orman ağaçlarının tesisinde ve büyümesinde mikorizal kolonizasyonun önemi uzun zamandır bilinmektedir ve mikorizal aşılama yeniden-dikimlerde düzenli olarak kullanılmaktadır (Amaranthus et al., 2005). Bu etkiyi daha ileri düzeyde öğrenmeye-çalışmak ereğiyle, şu muamele faktörü mülhaza edilebilir: ektomikorizal fungus aşılama tipini, aşılama miktarı, aplikasyon tipi. Başka sınıflandırma faktörleri dahil edilebilir: ağaçların türleri, aşılama zamanında fidelerin yaşı, seranın çevresel koşulları, pestisid uygulamasının tipi ve miktarı.

Misal 4 Çiftlikte yetiştirilen geyiklere tuzlu su içirilmesinin etkisini soruşturmak için bir öğrenme-çalışması yürütüldü (Kii and Dryden, 2005). Muamele faktörü tuzluluğa ilaveten şu faktörler tellaki edilebilir: Geyiğin türü, geyiğin cinsiyeti, geyiğin yaşı, geyik sürüsünün lokasyonu, ve diğer çevresel faktörler.

Bu faktörlere ve biyolojik sınırlandırmalarca veya istatistiksel mülhazalarca empoze edilen diğer olanaklı faktörlere (ki ileriki konularda detaylıca açıklanacak) dayalı olarak, uygun bir deney dizaynı, mütenasip bir istatistiksel model ile birlikte seçilmelidir. Bu iki adım eş zamanlı atılmalıdır ve bunlara karar verildikten sonra, soruşturmanın gidişatı/seyri çok güzel belirlenecektir ve tabii ki baz alınacak istatistiksel analiz de. Böyle olunca, bu noktada, eğer gerekirse deneyin herhangi bir gerekçeyle yeniden mülhaza edilmesi mümkün olacaktır (bilimsel deneylemenin mantıksal adımlarını temsil eden şemada kesikli çizgilerle gösterilmiştir). Deneyin niçin böyle bir yeniden-değerlendirmesinin veya yeniden-formülasyonunun gerekli olacağı pek çok sebebe bağlı olabilir: (i) Deney çok büyük ve çok kompleks olmuştur öyle ki mevcut koşullarda yürütmek olanaksızdır, (ii) Kullanılabilecek deney materyalinin empoze ettiği fiziksel sınırlandırmalar, elde edilmek istenen enformasyona kısmen veya tamamen ulaşmayı olanaksız kılabilir, ve (iii) yetersiz deney üniteleri “iyi” enformasyon sağlamaz.

Dizaynın Üç Unsuru

Burada anlatmak istediğim, şu ki deneye başlamadan önce deneyi en başından mantıksal neticesine kadar düşünmek ve zihninde tasarlamak gerekir. Bu bakımdan, deney dizaynının seçimi çok kritiktir. Bir deney dizaynı üç bileşeni havidir: (i) muamele dizaynı, (ii) hata-kontrol

dizaynı, ve (iii) örnekleme ve gözlem dizaynı. *Muamele dizaynı* öğrenme-çalışmasına dahil edilecek muameleleri belirler: hangi muameleler seçilecek ve bunların sayısı kaç olacak? Yani muamelelerin doğası nedir. Çoğu zaman seçilen muamelelerin bir yapısı olur. Yani muameleler çeşitli muamele faktörlerince ve bu faktörlerin düzey kombinasyonlarıncaya belirlenebilir. Biz buna faktöriyel yapı diyoruz. Sonra, şu sorular gelir: Kaç faktör kullanılacak? Her bir faktörün kaçar düzeyi olacak? Bu düzeyler hangi aralıklarda değişecek veya bu düzeyler için olanaklı seçimler nelerdir? Düzeylerin seçimi ve sayısı yalnızca faktörlerin nitel veya nicel olup olmamasına değil aynı zamanda ne çeşit enformasyonun elde edilmek istendiğine ve bunun analizde nasıl yansıtılacağına da bağlıdır. Bu soruların nasıl cevaplanacağına dair spesifik bir kılavuzlama yapmak imkansız. Her deney kendi karakteristikleri ve taleplerine sahiptir. Bununla beraber, sonraki bölümlerde daha spesifik dizaynları (hem muamele dizaynlarını hem de hata-kontrol dizaynlarını) tartışırken genel kılavuzlamalar kendini gösterecektir.

Muamele dizaynına ait hususlar hata-kontrol dizaynına ait hususlarla ile sıkıca irtibatlıdır. *Hata kontrol-dizaynı* dediğimizde, bir deney planında muamelelerin aktüel/fiili düzenlemesini kastediyoruz ki bu düzenlemede muameleler deney ünitelerine (yani deney materyalinin parçalarına) bir kurala göre atanır. Böylesi dizaynlara misal olarak tamamen rastgeleleştirilmiş dizayn, rastgeleleştirilmiş (tam veya eksik) blok dizaynı, Latin karesi dizaynı, ve benzerleri verilebilir. Bir hata-kontrol dizaynının seçimi, deney ünitelerinin elverişliliğine, bu ünitelerin yapısına ve soruşturmacının arzu ettiği kestirimin kesinliğine bağlıdır. Mesela, eğer deney üniteleri bir blok yapısına sahip, yani homojen ünitelerden oluşan kümelerde (bloklarda) gruplandırılabilirler ise rastgeleleştirilmiş blok dizaynının bir formu icap edebilir. Ya da deney üniteleri iki yönde heterojenlik gösteriyorlar (mesela belki bir tarla denemesinde) ise satır-sütun dizaynının bir formu (mesela Latin karesi, Youden karesi gibi) en mütenasip dizayn olabilir. Bloklama prensibi ve bu prensibi uygulayan dizaynların hatayı nasıl kontrol ettiği, sonraki bölümlerde, bir dizaynın niçin diğerine tercih edilmesi gerektiği ile beraber açıklanacak.

Seçtiğin muamele dizaynını buna uygun olarak seçtiğin bir hata-kontrol dizaynının içine yerleştirirsin.

Deney dizaynının üçüncü bileşeni *örnekleme ve gözlem dizaynı*dır. Bu, gözlemlerin hangi düzeyde yapılacağını ve ne çeşit gözlemlerin yapılacağını belirler. Daha net bir ifadeyle, bu, gözlem ünitelerinin deney üniteleriyle aynı olup olmayacağını veya deney ünitelerinden alt-örnekleme yapıp yapılmayacağını bize söyler. Ve dahi, bu, tek-değişkeli gözlemlerin mi yoksa çok-değişkeli gözlemlerin mi yapılacağını belirtir.

Daha evvel zikredildiği ve bilimsel deneylemenin mantıksal adımlarını temsil eden şemada belirtildiği üzere, deney dizaynının geliştirilmesi ve mütenasip bir istatistiksel modelin formülasyonu sıkıca bağlantılıdır öyle ki muamele dizaynının, hata-kontrol dizaynının ve örnekleme ve gözlem dizaynının yapıları, esasen istatistiksel modelin kompleksliğini belirler. Bu ders bağlamında, istatistiksel modeller doğrusal modeller veya doğrusal olmayan modellerin doğrusallaştırılmış formları olacak; daha spesifik olarak söylersem sınıflandırma ve regresyon modelleri. Bu ders, mutlak deneylerden ziyade karşılaştırmalı deneyleri başlıca konu edindiği için kullanılacak modeller yukarıda tartıştığımız deney dizaynının üç bileşeni ile alakalı etkileri biraraya getiren sınıflandırma modelleri olacak. Çeşitli dizaynları tartışırken bu daha iyi anlaşılacak.

Uygun bir deney dizaynı seçtikten sonra, şimdi deneyin kendisini icra edebiliriz. Şunu burada ifade etmek gerekir ki bütün deney prosesi içerisinde bu safha doğası itibarıyla istatistiksel değilmiş gibi görünse bile bu safhanın da seçilmiş bulunan dizaynın istatistiksel gerektirmelerine uygun olarak yürütülmesi elzemdir. Bu neyi kapsar? Mesela, deney ünitelerine

muamelelerin münasip rastgeleştirilmesini ve muamelelerin uygulanmasının münasip tekrarlanmasını kapsar. Bu sebeple, istatistikçinin ve soruşturmacının birlikte çalışıp en azından deneyin mümkün olduğunca tüm detaylarını ele alan bir protokolü yazmaları önemlidir. Bu protokol aynı zamanda aktüel/fiili veri toplama ve ölçüm prosesi hakkındaki detayları da içermelidir. Mesela veri ne zaman toplanacak ve ölçüm skalaları/ölçekleri ne olacak.

Yanıtın Modellenmesi

Aktüel/fiili deneyde her muamele faktörü farklı “düzeylerce” temsil edilir. Yani, farklı formlar veya farklı miktarlar; mesela aşılamanın farklı tipleri, aşılamanın farklı miktarları gibi. Tematik modelde yer alacak faktörlerin düzeyleri soruşturmacının elde etmeyi arzuladığı enfomasyona ve mevcut koşulların olanaklılığına göre, soruşturmacı tarafından belirlenir. Bir faktör yalnızca bir düzey ile sınırlandırılabilir gibi birden çok düzeyi de dahil edilebilir. Mesela, yalnızca kızıl geyik veya hastalığın farklı şiddetle seyrettiği gruplardaki hastalar. Birden çok düzeyi olan faktörlerin verinin analizine dahil edilmesi gerekecektir. Bu maksatla, yanıt verisinin uygun bir modelini ortaya koymak önemlidir.

Bu fikri genel terimlerle formalize etmek için (ki sonraki bölümlerde daha spesifikleştireceğiz) şöyle yazarız:

$$\text{Yanıt} = f(\text{Açıklayıcı değişkenler}) + \text{Hata}$$

ki burada f bir meçhul fonksiyonu temsil ederken, açıklayıcı değişkenler muamele ve hata-kontrol dizaynlarında yeralan muamele ve bloklama faktörlerini gösterir, sırasıyla.

İstatistiksel çıkarıma matuf maksatlarla hedef popülasyonu tanımlamak için tematik modelin tanımladığı faktörler de bloklama faktörleri arasındadır. Cox (1984) bu faktörleri *içkin* faktörler diye adlandırır. Biz de bu terminolojiyi takip ederek bloklama faktörlerini *içkin* faktörler ve *spesifik-olmayan* faktörler olarak ikiye ayıracağız. Spesifik-olmayan faktörler hata-kontrol dizaynının gerektirmeleri tarafından belirlenir. Yani, deney materyalinin heterojenliğini daha fazla azaltmaya matuf mülahazalarla. Eğer biz muamele faktörleri kümesini $\mathcal{X} = \{x_1, x_2, \dots, x_t\}$ ile, içkin faktörler kümesini gösterirsek $\mathcal{Z} = \{z_1, z_2, \dots, z_q\}$ ile ve spesifik-olmayan faktörler kümesini $\mathcal{U} = \{u_1, u_2, \dots, u_s\}$ ile gösterirsek, yukarıda yazmış olduğumuz modeli daha kesinlikle şöyle yazabiliriz:

$$y = \{x_1, x_2, \dots, x_t; z_1, z_2, \dots, z_q; u_1, u_2, \dots, u_s\} + e.$$

Şimdi bu terminolojiyi şu misallerle aydınlatmaya çalışacağım.

Misal 5 Pearce’in (1935) rapor etmiş olduğu bir deneyi mülahaza ediyoruz. Bu deney armut ağaçlarının farklı budama rejimlerini karşılaştırıyor. Farklı tipteki ve miktardaki budama kombinasyonları bir kaç ağaç sırasının her birisindeki bireysel ağaçlara atanmış. Her sıradaki ağaçlar oldukça uniform, fakat sıralar arasında, çevre koşullarından dolayı bir takım farklılıklar mevcut. Çıkarıma matuf maksatlarla birden çok armut ağacı çeşidi deneye dahil edilmiş. Bu kurulumda iki muamele faktörü var: $x_1 = \text{budama tipi}$ ve $x_2 = \text{budama miktarı}$, bir içkin faktör var: $z_1 = \text{armut çeşidi}$, bir spesifik-olmayan faktör var: $u_1 = \text{ağaç sırası}$. Nihai deney dizaynı rastgeleştirilmiş tam blok dizaynı içerisinde bir faktöriyel muamele dizaynıdır.

Misal 6 Şimdi başka bir illüstrasyon olarak hizmet etsin diye bir klinik denemesini betimleyeceğim. Farzedelim ki insan vücudunda belli bir tipteki deri kaşıntısının giderilmesine

yönelik farklı muamelelerin efektifliğini soruşturmak istiyoruz. Muamele faktörleri şunlardır: x_1 = losyonun konsantrasyonu ve x_2 = uygulama sıklığı. Faktörün farklı düzeylerinin her bir kombinasyonu bir medikal muameleyi tanımlar ve denemeye dahil edilen her bir hastanın her bir kolu farklı bir muameleye maruz bırakılır. Deneme kaşıntının şiddetine göre tasnif edilmiş bay ve bayan hastaları kapsar. Böylece cinsiyet ve kaşıntının şiddeti sınıfları içkin birer faktör olarak sırasıyla z_1 ve z_2 ile temsil edilir. Hasta ise bir spesifik-olmayan faktör olarak u_1 ile temsil edilir. Nihai deney dizaynı bir faktöriyel muamele dizaynı ve hata-kontrol dizaynı olarak tam-olmayan blok dizaynının bir formundan ibarettir.

Yanıtın Seçimi

Şimdiye kadarki tartışmamızda "yanıt" terimini genel manada kullandık. Pek çok durumda yanıtın veya yanıt değişkeninin ne olacağı fiilen bellidir. Eğer mesela farklı üretim proseslerinin belli bir tip plastik tüpün mukavemeti üzerine etkisini belirlemek istiyorsak, bu taktirde apaçık yanıt, tüpü tahrip edecek kuvvet olarak psi, kg/m^2 cinsinden ölçülecektir. Başka bir misal olarak, eğer biz farklı kirleticilerin belli bir bitki üzerindeki etkilerini değerlendirmek istiyorsak, neyin ölçüleceği öyle açık olamayabilir. Denemenin sonunda veya büyüme periyodunun sonunda bitkilerin büyümesini, veya büyüme sezonunun sonunda bitkilerin verimini, veya denemenin sonunda bitkilerin yapraklarındaki hasarın derecesini ölçebiliriz.

Niyetimiz burada şunu işaret etmek ki bilimsel-araştırmacı yanıt değişkeninin seçimini dikatle mülâhaza etmelidir. Yanıt değişkeni deneyin beklenen çıkımı bağlamında en manalı olmalıdır ve deneyin çıkımı ile açıkça en ilgili olanı olmalıdır. Başka bir deyişle, yanıt değişkeni öyle seçilmelidir ki deneyden yapılacak çıkarımlar ve deneyin sonuçları net olarak ifade edilebilsin ve başkalarıyla paylaşılabilir. Bu bağlamda, bir sürekli yanıt değişkeni bir ikili veya dereceli değişkene tercih edilir çünkü daha çok enformasyon içerir. Yine aynı sebeplerle, bir nesnel yani ölçülebilir değişken bir öznel değişkene tercih edilir. Tüm bunlar, elbette, nihayetinde bir seçim yapıyor olabilmemize bağlıdır.

Hala başka mülâhazar gerekli. Mesela, bir bitki bireyinin verimini mi yoksa bir deste veya parselin verimini mi ölçeceğiz? Ya da, bir bireyin nabzını belli bir zaman noktasında mı yoksa belirtilmiş bir periyot içerisindeki birden çok zaman noktalarında mı ölçeceğiz. Yine, bunu belirleyecek olan deneyde kullanılan muamelelere dair yapmak istediğimiz çıkarımın tipidir.

Analizin Prensipleri

Veriyi topladıktan sonra bu toplanan veriyi deneyin dizaynı ve deney için uygun görülen model ile uyarlı bir istatistiksel analize tabii tutacağız. Bu analizleri daha sonraki bölümlerde detaylarıyla ele alacağız. Bu noktada sadece müşterek temel prensipleri anıyoruz.

Dizayn edilmiş deneylerden elde edilen veriyi analiz etmenin bir majör hedefi değişkenliğin/varyasyonun olanaklı kaynaklarını nicelemek ve önemini değerlendirmektir. Bunu başarmak için altta yatan doğrusal modele ilişkin varyans analizinden (ANOVA: Analysis of variance) yararlanır ki tek-değişkeli ya da çok-değişkeli bir formda olabilir. ANOVA daha sonra detaylıca tartışılacak fakat şimdilik kısa bir çerçevesini çizceğim.

Bir deneye ait gözlemler, y diyelim, verildiğinde varyans analizinin genel fikri "total değişkenliği" (ya da total kareler toplamını, $KT(\text{Total}) = \sum(y - \bar{y})^2$) altta yatan doğrusal

modelin belirttiği şekliyle bileşen parçalarına bölmektir. Böyle bir model gözlemlerin yapısını yansıtır ki bu yapı muamele dizaynı, hata-kontrol dizaynı ve örnekleme dizaynı tarafından belirlenmiştir. Her dizayn birden çok etki kümesi (/parametreler) ile temsil edilir. Bu etkiler bir veya daha çok hata terimine ilaveten açıklayıcı değişkenlerin ana etkileri ve aralarındaki çeşitli interaksyonlar formunda modelin daha bariz bir ifadesini sağlar. Farzedelim ki böyle kümelerin q adedi birarada olsun. Buna göre, en küçük kareler metodu kullanılarak $KT(\text{Total})$ şöyle (yegane olması gerekmez) parçalanır

$$KT(\text{Total}) = KT(1) + KT(2) + \dots + KT(q)$$

ki burada $KT(i)$ etkilerin i . kümesine ($i = 1, 2, \dots, q$) ait kareler toplamını temsil eder ki bu etkilere hamledilebilecek varyasyonu bir manada açıklar. Tartışmamızı muamelelerin ve deney hatasının kareler toplamları üzerinden devam ettireceğiz.

İster hipotez testlerinde (veya anlamlılık testleri) ister güven aralığı kestiriminde olsun, ANOVA istatistiksel çıkarım yapmak için bize gerekli olan temel enformasyonu sağlar. Bir kareler toplamı kendisine ait serbestlik derecesine, v_i , bölüdüğü zaman kareler aritmetik-ortalaması elde edilir, $KAO(i) = KT(i)/v_i$. Hipotez testlerinin icra edilmesi ve hata varyanslarının kestirimi için beklenen kareler aritmetik-ortalaması $E[KAO(i)]$ teşkil edilir.

Tüm hipotez (ya da anlamlılık) testleri ve parametrik fonksiyonların kestirimi deneyin hedefleri ile uyarlıca yapılır. Böylece istatistiksel sonuçların yorumlanması soruşturmacının formüle etmiş olduğu orijinal hipotezler cinsinden yapılmış olacaktır.

Doğrusal Model

Etkilerin Üç Tipi

Şimdiye kadarki tartışmamızda tematik model ve deney dizaynı ile ilişkili olarak defaaten doğrusal modelin önemine işaret ettik. Doğrusal modeller sonraki bölümlerde tüm dizaynlar için verilecek veya türetilecek fakat bu moellerin formu hakkında bazı müşterek argümanları burada vermek yararlı olur. Genel fikir, genellikle y ile temsil edilen gözlemleri y ye katkıda bulunan “etkiler” cinsinden ifade etmektir. Bu etkiler ya da bileşenler temel olarak üç kategori oluşturur: (i) muamele etkileri, (ii) dizayn etkileri ve (iii) hata etkileri.

Muamele etkileri müdahale prosedürü veya muamele dizaynının bir yansımasıdır. \mathcal{X} te listelenen muamele faktörleri tek bir muamelenin mi yoksa birden çok muamele faktörlerinin kombinasyonlarının mı kullanıldığını belirtir. Tematik mülahazalarla beraber bu, hangi etkilerin, yani ana ve interaksyon etkilerinden hangilerinin, doğrusal modele dahil edileceğini belirler.

Dizayn etkileri \mathcal{Z} ye ve \mathcal{U} ya dahil edilen açıklayıcı değişkenler tarafından belirlenir. Bu etkilere biz aynı zamanda hata-kontrol dizaynının parçası olarak blok etkileri de deriz.

Saf muamele ve dizayn etkilerine ilaveten muamele \times dizayn etkilerini de gerektiği zaman modele dahil edebiliriz. Bu \mathcal{X} teki muamele faktörleri ile \mathcal{Z} deki için faktörler arasındaki olanaklı interaksyonlardan kaynaklanır.

Nihayetinde, hata etkileri, ya da kısaca hatalar, farklı tiplerde rastgele değişim/varyasyon gösterirler. Böylesi değişim/varyasyon, aktüel/fiili deneyleme ve veri toplamanın bazı hususları

ile olduğu kadar deney ve gözlem üniteleri ile alakalıdır. Yine, bu hususlar sonraki bölümlerde detaylıca tartışılacaktır.

Şimdi yukarıda tartıştığımız bazı nosyonları bir misalle aydınlatalım.

Misal 7 Rosen ve arkadaşlarının (2005) yapmış olduğu bir öğrenme-çalışmasının ereği nitrojen ve sülfür ile gübrelemenin lahanada glucosinolate konsantrasyonlarını etkileyip etkilemeyeceğini belirlemektir. Mamele faktörleri $x_1 = \text{nitrojen} = N$ (iki oranda), $x_2 = \text{sülfür} = S$ (iki oranda), $x_3 = \text{çeşitler}$ (iki tip: yeşil lahanaya ve kırmızı lahanaya) idi.

Buna göre modele dahil edilen muamele faktörleri şunlardır: N oranı, S oranı, N oranı \times S oranı interaksyonu, çeşit, çeşit \times N oranı interaksyonu, çeşit \times S oranı interaksyonu, çeşit \times N oranı \times S oranı interaksyonu. İlaveten, muamele \times dizayn interaksyonu etkileri N oranı \times yıl interaksyonunu, S oranı \times yıl interaksyonunu, çeşit \times yıl interaksyonunu kapsar.

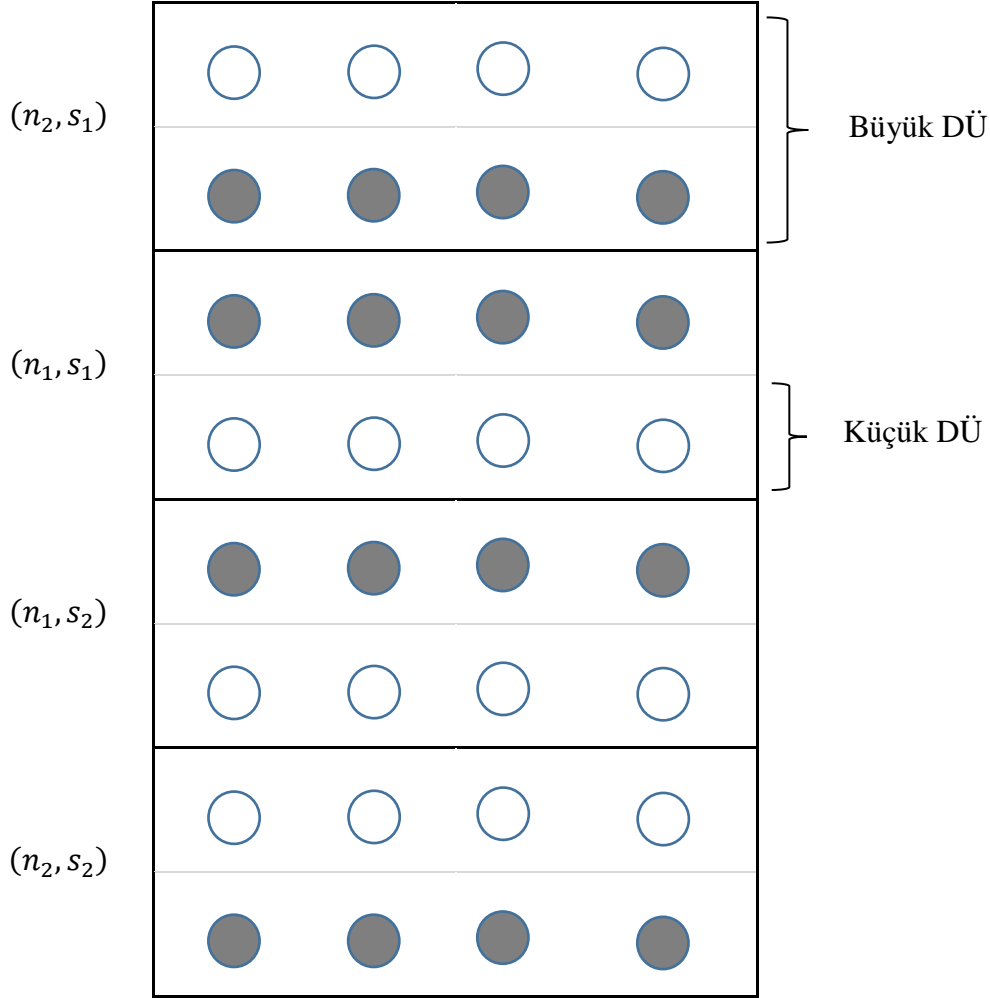
Deney ve Gözlem Üniteleri

Hata etkilerinin ya da hata bileşenlerinin mahiyetini ve kullanımını anlamak için muamelelerin uygulandığı ve üzerinde gözlemlerin ya da ölçümlerin yapıldığı (farklıkları mümkün) üniteler arasındaki ayrımı anlamak esastır. Bu üniteleri deney üniteleri ve gözlem yada örnekleme üniteleri diye adlandırıyoruz, sırasıyla. *Deney ünitesi* deney materyalinin parçasıdır ki bu parçaya bir muamele atanır ve uygulanır. Mesela farklı hastalara farklı ilaçların verildiği bir klinik denemesinde her hasta bir deney ünitesidir. Eğer, diğer taraftan, her hastaya farklı kollarından farklı merhemler uygulanıyor ise bu taktirde her kol bir deney ünitesini oluşturur. *Deney hatası* bir deney ünitesine aittir. Bu hata şu gerçeği yansıtır ki deney üniteleri birbirinin aynı değildir, yani tam olarak tekrarlanamaz. Deney hatasına bilimsel-araştırmacı da katkıda bulunur, muameleyi tam olarak tekrarlayamaz ise. Şöyle ki belli bir maddeden protokolde belirtildiği üzere 15 ppm yerine 14 ppm veya 16 ppm vb. uygulanabilir. Esasen hassasiyeti artırdıkça hiç bir tekrarlama diğerinin aynı olamaz. Deney hatasının bu bileşenini biz *muamele hatası* diye adlandırıyoruz.

Burada halihazırda önemini vurguluyoruz ki belli bir deney için deney ünitelerini daima açıkça tanımlamak gerekir. Birden çok muamele faktörü mevcut olduğunda, bu farklı muamele faktörlerini farklı tiplerdeki deney ünitelerine bağımsız olarak ve birbirinden ayrı olarak uygulamak gerekebilir. Bunun bir vargısı olarak o zaman farklı deney hataları da olacaktır ve anlamlılık testleri için bu çok önem arzedecektir. Bu bölümümüz parsel tipi dizaynların spesiyal karakteristiğidir.

Misal 7 deki öğrenme-çalışmasının deneysel kurulumunu kullanarak yukarıda bahsettiğim noktayı aydınlatacağım.

Misal 7 (devam) Burada deney ünitelerinin iki tipine sahibiz. Daha spesifik olarak söylersek, “büyük” deney ünitelerimiz var ki biz bunlara N ve S nin ikişer oranlarının kombinasyonlarını uyguluyoruz, ve her geniş deney ünitesinin içerisinde iki “küçük” deney ünitesi var ki biz bunların her birinde bir tip lahanayı büyütüyoruz. Eğer N nin iki oranını n_1, n_2 ile, S nin iki oranını s_1, s_2 ile ve bunların bir kombinasyonunu (n_i, s_j) ($i, j = 1, 2$) ile gösterirsek, buna göre bir yılda bir tekrarlama için deney konfigürasyonunu esasen aşağıdaki gibi şematize edebiliriz. Ki bu şemada boş daireler yeşil lahanayı ve dolu daireler kırmızı lahanayı temsil etmektedir.



Deney ünitesi ve gözlem ünitesi arasında ayırım yapmak önemlidir. *Gözlem* (ya da *örnekleme*) ünitesi o ünedir ki üzerinde gözlemler, yani ölçümler yapılır. Pek çok durumda deney ünitesi ve gözlem ünitesi özdeştir fakat bazı durumlarda değildir. Mesela, bir eğitim-öğretim öğrenme-çalışmasında bir sınıf, yani öğrencilerin bir topluluğu, deney ünitesidir eğer ki bu sınıf muamele olarak belirli bir öğretim metoduna bütünüyle maruz bırakılıyor ise. Mamafih, gözlemler bireysel olarak öğrenciler üzerinde test puanları formunda yapılacaktır. Bu durumda öğrenciler gözlem üniteleridir. Gözlem ünitesine ait bir *gözlem* (ya da *örnekleme*) hatası vardır ki bu bize ölçme hatasını ve eğer deney başka öğrenciler üzerinde tekrarlanmış ise aynı zamanda örnekleme hatasını da diğer şeyler arasında yansıtır.

Bu noktayı daha derinlemesine anlamak için yine Misal 7 de betimlenen deneyi mütalaa edeceğiz.

Misal 7 (devam) Bu kurulumda veri elde etmeye matuf farklı senaryolar düşünebiliriz:

- i. Her sıradan (küçük DÜ) rastgele bir lahana kellesi seçip onun üzerinde mütenasip kimyasal analizi yapabiliriz. Bu durumda, en azından istatistiksel bir bakış açısından, deney ünitesi ve gözlem ünitesi arasında bir ayırım yok. Mamafih, bu durum bize şu noktayı gösterir ki gözlem hatasının parçası olarak örnekleme hatası mevcut olacaktır.

- ii. Her sıradan rastgele iki lahana kellesi hasat edebiliriz, doğrayıp bunları karıştırarak kuruttuktan sonra ileriki analize hazırlarız. İstatistiksel analiz bağlamında bu durum i. madde ile aynıdır.
- iii. Her sıradan iki (veya daha çok sayıda) lahana kellesini rastgele hasat edebiliriz ve her kellede glucosinolate ekstraksiyonu yapılır. Bu durumda deney üniteleri ve gözlem üniteleri farklıdır. Sıralar deney üniteleridir ve bireysel kabak kelleleri gözlem üniteleridir. Bu *alt-örnekleme* diye adlandırdığımız şeyi ifade eder.

Bir Model Çerçevesi

Yukarıda “Yanıtın Modellenmesi” alt-başlıklı bahisteki iki denklem ile bir deneyden elde edilecek gözlemler için bir modelin çok genel bir formunu vermişim. Hemen yukarıdaki tartışmada ise o denklemlerdeki f fonksiyonunun doğrusal bir fonksiyon olduğunu telkin ettim. Ünite-muamele eklenirliği nosyonuna dayanan bir formal türetmeyi sonraki bölümlerde vereceğim. Fakat şimdi burada şu fikri aktarmak istiyorum. Müdahaleyi takip eden yanıt (ki bir tarımsal pratiği/uygulamayı müteakiben verim/ürün diye adlandırılır) bir ünite etkisi artı bir muamele etkisi artı hata etkilerinden (mesela ünite veya deney hatası, gözlem veya ölçme ve örnekleme hatası) eklemeli olarak meydana gelir. Ünite etkileri ve deney hatası etkileri hata-kontrol dizaynının, sırasıyla, sistematik ve rastgele tesirlerini modellemek için kullanılır. Bu yüzden biz ünite etkilerine aynı zamanda dizayn etkileri diyeceğiz. Bunlar denklemdeki içkin ve spesifik-olmayan faktörlerin etkilerini kapsamaktadır.

Şimdi tüm söylediklerimi şematize edersem,

$$\begin{aligned} \text{Gözlem} = & \text{Dizayn etkisi} + \text{Muamele etkisi} \\ & [+ \text{Dizayn} \times \text{Muamele interaksyonu etkisi}] \\ & + \text{Deney hatası} \\ & + \text{Gözlem hatası.} \end{aligned}$$

Bu denkleme dizayn \times muamele interaksyonu etkisini ekledik ki bu etki belli deneysel durumlarda ilgilendiğimiz bir bileşen olarak zuhur edebilir.

Adımları Misaller ile Aydınlatalım: Öğrenme-çalışması 1

Sorular ve Hipotezler

Şimdi “Dizayn Edilmiş Soruşturmanın Adımları” başlıklı kısımda betimlediğim adımları ve “Doğrusal Model” başlıklı kısımda bir misal üzerinden anahatlarını verdiğim bir doğrusal modelin formülasyonunu misallerle aydınlatacağım. Varsayalım ki bir soruşturmacı hava kirleticilerinin Loblolly çamı (*Pinus taeda*) fidelerine etkilerini öğrenmeye-çalışmak istiyor. Kullanılacak kirleticiler, tek başlarına ve kombinasyonları olarak, ardışık 28 günde 6 saat/gün ve 0.00, 0.05, 0.01 ppm düzeylerinde üniform yaştaki fidelere uygulanmak üzere ozon (O_3) ve nitrojen dioksittir (NO_2). Soruşturmacı şununla ilgileniyor: (i) kirleticilerin zararlı etkilerini karşılaştırmak ve (ii) O_3 ve NO_2 nin potansiyel sinerjik etkilerini değerlendirmek. Bu etkilere aynı zamanda muhtemelen ağaçların genetik yapılarının (yani hava kirleticilerine karşı ne

derece hassas veya dayanıklı olduklarının) da tesiri vardır. Daha formel olarak, bilimsel-araştırma hipotezlerini şöyle ifade edebiliriz:

- i. O₃ ve NO₂ kirleticilerine uzun süreli maruziyet çam fideleri üzerinde büyüme, beneklenme ve klorotik leke semptomları bakımından zararlı etkilere sahiptir.
- ii. Zararın miktarı kirliliğin düzeyi ile birlikte artar.
- iii. Kirleticilerin bir kombinasyonu sinerjik etkiler ortaya çıkarır.
- iv. Zararın miktarı aynı zamanda fidelerin geldikleri ailelerin (genetik olarak belirlenen) duyarlılık derecesine de bağlıdır (yani bir tipteki aile nispeten dayanıklı iken başka bir tipteki aile nispeten hassas olabilir)

Bu taktirde gözlemleri ve ölçümleri belirleyecek ve onlara tesir edecek olan muameleler yani kirletici (kombinasyonlarının) tipleri, kirlilik düzeyi, genetik arkaplan ve büyüme için başlangıç boyu olacaktır. Sıcaklık, ışık miktarı ve nemlilik gibi diğer faktörler uyarlanacak deney dizaynına bağlı olarak hesaba katılabilir. Fideler kirletenlere kirlilik odalarında maruz bırakılacaktır.

Deneyimiz ve Bir Model

Burdan devamla çok kesin/tam bir dil ve terminoloji kullanmak yerine okuyucuya şimdiye kadar mevzubahis ettiğim çeşitli kavramları hissettirmeye ve farkettermeye çalışacağım.

Hata-kontrol dizaynı kirlilik odalarının elverişliliğine ve düzenlenmesine bağlıdır. Farzedelim ki bilimsel-araştırmacının kullanımı için 18 oda elverişlidir. Bu odalar bir laboratuarda ısı, ışık ve nem gibi üniform çevre koşulları altında dağılmaktadır. Buna göre bir olanaklı düzenleme her muameleyi yani olanaklı dokuz kirlilik kombinasyonunun her birini iki kirlilik odasına rastgele atamak olurdu. Her odada iki aileye mensup belirtilen eşit sayıdaki, diyelim 2s fide, aynı protokol uyarınca, atanmış olan kirleticiye maruz bırakılacaktır. Buna düzenleme I diyelim.

Bu deneysel kuruluma ilişkin bir doğrusal model şöyle olabilir:

$$y_{ijkl} = \mu + P_i + C_{ij} + F_k + (PF)_{ik} + \varepsilon_{ijk} + \eta_{ijkl}$$

ki bu modelde yer alan sembollerin temsil ettiği terimler şunlardır:

y_{ijkl} : i inci kirleticiye atanmış j inci odadaki k ncı aile tipinin l inci fidesi için bir gözlem

μ : genel aritmetik-ortalama

P_i : i inci kirleticinin etkisi, ($i = 1,2, \dots, 9$)

C_{ij} : i inci kirleticiye atanmış j inci odanın etkisi, ($j = 1,2$)

F_k : k ncı aile tipinin etkisi, ($k = 1,2$)

$(PF)_{ik}$: i inci kirletici ile k ncı aile tipi arasındaki interaksiyonun (eklenmezlik) etkisi

ε_{ijk} : deney hatası

η_{ijkl} : gözlem (örnekleme) hatası

Şunu ifade etmek gerekir ki burada deney hatası iki bileşenden mürekkeptir: birinci bileşen (C_{ij}) kirleticilerin farklı odalara (=DÜ) uygulanmasıyla ortaya çıkar; diğer bileşen (ε_{ijk}) her oda-aile kombinasyonu ile bağlantılı olarak ortaya çıkar yani mesela dayanıklı olduğu bilinen aileler bile özdeş değildir yani tamamen aynıyı yeniden üretebilir değildir (dayanıklılıklarının kendi içinde bir değişkenliği vardır). Muamelelerin yani kirleticilerin

gerçekte O_3 ve NO_2 olmak üzere iki faktörün aşağıda gösterilen düzey kombinasyonları olduğu hakikatini kullanarak modelimizi daha da açabiliriz:

		NO ₂		
		0.00	0.05	0.10
O ₃	0.00	P_1	P_2	P_3
	0.05	P_4	P_5	P_6
	0.10	P_7	P_8	P_9

Bu yapı, ki buna aynı zamanda faktöriyel yapı diyoruz, bizim her P_i etkisini O_3 ve NO_2 ana etkilerine ve $O_3 \times NO_2$ interaksiyon etkisine parçalamamızı sağlar. Bu da bize dokuz kirleticiyi biribiriyile karşılaştırmanın yolunu açar ki mesela eğer O_3 ve NO_2 nin beraber uygulanmasının bir sinerji yaratıp yaratmayacağı sorusunu böylece cevaplayabiliriz.

“Yanıtın Modellenmesi” başlıklı alt-kesimde yapmış olduğumuz tartışmanın terminolojisine uygun olarak, bu son yazdığımız modelde şunu not ediyorum ki O_3 ve NO_2 muamele faktörlerini temsil ediyor, aile tipi bir içkin faktörü temsil ediyor ve kirlilik odası bir spesifik olmayan faktörü temsil ediyor.

Analiz

Bir istatistiksel analizin çerçevesi (veya ana hatları) aşağıda verildiği üzere bir varyans analizi tablosunda izhar edilebilir.

Varyasyon Kaynağı	Serbestlik derecesi (s.d.)	Test edilebilecek Bilimsel-araştırma Hipotezi
Kirleticiler	8	Kirleticiler arasındaki farklar (bak i.)
O_3	2	O_3 ün düzeyleri arasındaki farklar ki NO_2 üzerinde ortalananmıştır (bak ii.)
NO_2	2	NO_2 nin düzeyleri arasındaki farklar ki O_3 üzerinde ortalananmıştır (bak ii.)
$O_3 \times NO_2$	4	O_3 ve NO_2 arasındaki sinerji (bak iii.)
Odalar (Hata 1)	9	
Aileler	1	
Kirleticiler \times Aileler	8	Kirleticiler ve aileler arasındaki interaksiyon (bak iv.)
Hata 2	9	
Gözlem Hatası	$36(s-1)$	
Total	$36s-1$	

Bu tablo bize hangi hipotezlerin test edilebileceğini ve bunların soruşturmacının hedefleriyle uyumunu, yine çok kesinlikli olmayan terimlerle, gösteriyor. Daha spesifik hipotezler de elbette müteakip prosedürler kullanılarak test edilebilir. Burada benim esasen anlatmak istediğim şu ki deney öyle bir şekilde düzenlenmelidir ki deneyin başlangıcında ortaya koyulan sorulara bu deney cevap verebilsin. Bu elbette deneyinin ereklerine varmasını garanti etmez. Fiziksel koşullar ve mali koşullar, doğrusunu söylemek gerekirse, deneyin hedefleri ile tutarlı olduğu

sürece başka bir eylem planı yönünde etkili olabilirler. Deneyin icrasında, muamelelerin yani kirleticilerin kirlilik odalarına rastgele atanmalarına ve bir oda içindeki fidelerin rastgele yerleştirilmelerine veya deney süresince bir tür rotasyona tabi tutulmalarına dikkat gösterilmelidir. Diğer “çevresel” koşulları kontrol altına alan bir müesses protokol takip edilmelidir. Değerlendirme prosedürlerine dikkat edilmelidir. Mesela, yapraktaki semptomlar her bir iğne üzerinde mi ölçülmelidir veya değerlendirmelidir ve nasıl, veya her fide bir bütün olarak mı değerlendirilmelidir? Boydaki büyüme nasıl ölçülmelidir, nereden nereye ve hangi zaman periyodunda? Mütenasip veriyi topladıktan sonra, bu veri yukarıda çerçevesini çizdiğimiz modele göre analiz edilecek. Elimizde gerçek veri olmaksızın burada soyut neticelere varmak zor, fakat daha önce söylediklerimizin ışığında bu deneyin sonuçlarının nasıl yorumlanacağı açık olmalıdır. Sinerji sorusu doğrudan cevaplanabilir. Sonuç itibarıyla, yeni soruların neşet edebileceği ki bu soruların da ardışık deneylemenin bir parçası olarak bir yeni soruşturmaya yol açacağını düşünmek zor olmasa gerek. Bir deneyi dizayn etmede canalcı nokta soruşturmacının sorularının istatistiksel analiz bağlamında cevaplanabilirliğinden emin olmaktır. Bu şu manaya gelir ki biz varyans analizi tablosunda bilimsel-araştırma hipotezlerine tekabül eden istatistiksel hipotezleri test edebilmeliyiz.

Alternatif Deney Kurulumu

Yukarıda tartıştığım misal üzerinden nelerin yanlış gidebileceğini göstermek için, şimdi şu alternatif deney düzenini mülâhaza edelim ve buna düzenleme II diyelim. Her kirletici kombinasyonunu iki odaya atıyoruz odalardan birinde bir aileden $2s$ fide ve diğerinde diğer aileden $2s$ fide olacak. Bir başka deyişle, bunun manası kirleticilerin her kombinasyonu ve her aile rastgele bir odaya atanacaktır (bu da demektir ki “aile” şimdi bir muamele faktörüdür). Şunu anlamalısın ki bu düzenleme düzenleme I den esasen farklıdır. Böyle olunca farklı bir doğrusal model kullanacağız, veriyi analiz etmek için. Bu yeni modeli şu formda yazabiliriz:

$$y_{ikl} = \mu + P_i + F_k + (PF)_{ik} + \varepsilon_{ik}^* + \eta_{ikl}$$

ki burada tüm terimler bir önceki modelde tanımlandığı gibidir, yine ε_{ik}^* ve η_{ikl} ($l = 1, 2, \dots, 2s$) deney hatasını ve gözlem hatasını temsil etmektedir. Bu modele ait varyans analizi tablosu şöyledir.

Varyasyon Kaynağı	Serbestlik derecesi (s.d.)
Kirleticiler	8
O ₃	2
NO ₂	2
O ₃ × NO ₂	4
Aileler	1
Kirleticiler × Aileler	8
O ₃ × Aileler	2
NO ₂ × Aileler	2
O ₃ × NO ₂ × Aileler	4
Deney Hatası	0
Gözlem Hatası	18(2s-1)
Total	36s-1

Başlıca sonuç, burada, şu ki deney hatası için sıfır serbestlik derecesi (s.d.) var (bu formel olarak total s.d. 36s-1 den kaynaklanıyor fakat aynı zamanda her muamele kombinasyonunun yalnızca bir odaya uygulanmış olması gerçeğinden de kaynaklanıyor). Bunun manası şu ki kirleticiler ve aileler arasındaki interaksiyonun bütünüyle veya kısmen ihmal edilebilir olduğunu varsaymadıkça hiç bir hipotezi test edemeyiz. Bu varsayım, mamafih, iv. bilimsel-araştırma hipotezi ışığında gerçekçi değildir. Bu yüzden bu düzenlemenin bir değeri yoktur ve artık kullanılmamalıdır. Bu düzenlemeyi yapmış olmamın tek sebebi öyleyse belirli bir düzenlemenin soruşturmacının ortaya koymuş olduğu sorulara cevap verebilecek kapasitede bir istatistiksel modeli ve böylece bir istatistiksel analizi sağlayıp sağlayamayacağını kontrol etmenin önemini vurgulamaktır.

Deney Dizaynının Üç Prensibi

Bir deneyin adımlarını betimlerken, istatistiksel hususların önemini vurgulamıştık. Bu istatistiksel hususları sayfa xx teki figüre atıfla “istatistiksel üçgen” diye adlandırmıştık: (1) bir deney dizaynının seçimi, yani muamele dizaynı ve hata-kontrol dizaynı, (2) bir mütenasip doğrusal modelin formülasyonu ve (3) seçilen deney dizaynı ve ona ait istatistiksel modele dayanan istatistiksel analizin çerçevesi. Analizin geçerliliğini temin etmek ve duyarlılığını artırmak için her deneyin olmazsa olmazı üç temel prensibi göz önünde tutmalıyız.

Birincisi *tekrarlama* prensibi. Her muamele (veya muamelelerin bazıları) birden çok deney ünitesine uygulanmalıdır. Benzer deney üniteleri arasında sistematik farklılıkların olmaması durumunda, tekrarlamalar bize deney hatasını (rastgele hatayı) kestirebilme olanağı sağlar ki muamelerimizin arasındaki farkların büyüklüğüne bu deney hatası ile kıyaslayarak karar veririz. (Tekrarlamasız deneyler ancak belli durumlarda belli varsayımlar altında işe yarar.)

Deney hatasına ait kestirimimizin geçerliliğini temin etmek için ikinci prensibe dayanırız ki bu *rastgeleleştirme* prensibidir. Bu prensip varyansın yansız bir kestirimini sağlamakla beraber aynı zamanda muamele farklarının da bir yansız kestirimini yani başka türlü kontrol altına alınamayan varyanstan kaynaklanan sistematik farklılıklardan arındırılmış kestirimleri bize sağlar. Rastgeleleştirme prensibini sonraki bölümlerde detaylıca ele alacağım ve nasıl icra edileceğini, mütenasip doğrusal modelleri formüle edebilmeyi bize nasıl sağladığını ve istatistiksel analiz üzerindeki etkisini anlatacağım.

Mütenasip bir hata-kontrol dizaynı seçmenin başlıca ereklerinden biri aslında deney hatasını indirmektir. Çoğu durumda üçüncü prensibin kullanılması ile bu başarılıdır. Bu prensip *lokal kontrol* veya *bloklama* prensibidir. Temel fikir deney ünitelerinin total kümesini kendi içlerinde mümkün olduğunca homojen alt-kümelere (bloklara) parçalamaktır. Böyle yapmakla deney üniteleri arasındaki farkların sistematik değişimine/varyansına katkıda bulunan muzır faktörlerin etkileri elemine edilebilir. Bu, o halde, bizi daha duyarlı bir analize götürecektir çünkü deney hatası her bir blok içerisinde değerlendirilecek ve sonra tüm deney için karılacaktır. Bloklama (içkin ve/veya spesifik-olmayan faktörlere göre) çeşitli yollardan ve deneyin çeşitli safhalarında vuku bulabilir ve deney koşulları ve deneyde arzulanan duyarlılığın gerektirimleri bunu dikta eder. Mesela Latin karesi ve bölünmüş-parsel dizaynları çok daha komplike bloklama yapılarıdır ki daha sonraki bölümlerde konu edeceğiz. Yaptığım tartışmanın bariz manası şu ki ne kadar çok bloklama yapılırsa, deney de o kadar çok duyarlı olur. Ancak bu belli bir noktaya kadar doğrudur ve bloklama faktörlerine ilişkin sistematik değişkenliğin miktarına bağlıdır. Yani bloklamanın yararlılığı mevcut deney durumunun ve deneycinin o konudaki bilgi düzeyinin bir fonksiyonudur. Aynı zamanda, diğer tarafından bakınca,

bloklamanın fazlası hem deneyin icrası bakımından ve hem de verinin analizi bakımından deneyleri daha kompleks bir hale getirecektir.

İstatistiksel Üçgen: Öğrenme-çalışması 2

“Dizayn Edilmiş Soruşturmanın Adımları” adlı bölümde dizayn edilmiş soruşturmalar için esansiyel olan adımların detaylıca bir çerçevesini çizmişim. Bunları sayfa xx te yer alan figürde şematize etmişim. Deney dizaynının seçimi, deney dizaynına ilişkin istatistiksel model ve yapılacak istatistiksel analiz arasında sıkı bir ilişkiye dikkat çekmişim. Sayfa xx teki figüre atfen bunu “İstatistiksel üçgen” diye adlandırıyoruz. Bilimsel deneylemenin tüm uğraşında bunun merkezi bir rolü vardır ve bunu daha iyi anlayabilmeniz ve taktir edebilmemiz için biraz daha burada tartışmayı sürdüreceğim. Bu tartışmayı çok basit bir misal (Öğrenme-çalışması 2) üzerinden yapacağım. Gözlemler için farklı deneysel durumları yansıtan modellerin, sadece deneyimsel argümanları kullanarak, nasıl förmüle edilebileceğini göstermeyi amaçlıyorum, ana fikir olarak. Şunu baştan vurgulamalıyım ki aşağıda betimlenen durumlar için tüm modeller – muamele etkisi, deney hatası ve gözlem hatası formundaki – aynı bileşenlere sahip olsalar dahi varyans analizinde (ANOVA) iki hata bileşeninin ve bunlara ait kareler aritmetik-ortalamalarının rolleri büyük ölçüde ve başlıca alta-yatan deney planına bağlıdır.

Problemin İfadesi

Farzedelim ki bir soruşturmacı kirleticilerin çam fideleri üzerindeki etkilerini öğrenmeye-çalışmak ve karşılaştırmak istiyor. Kontrol olarak karbon filtreden geçirilmiş havaya (P_1) ilaveten soruşturmacımız şu kirleticileri deneyine dahil ediyor: ozon (P_2), sülfür dioksit (P_3) ve nitrojen dioksit (P_4). Bu bir keşifçi/keşifsel deney ve soruşturmacımız her kirletici için dört fide, yani toplamda 16 fide kullanabilecek. Fidelerin aynı yaşta ve üniform boyda olduklarını, makul bir fümigasyon protokolü tespit edildiğini ve münasipçe uygulanacağını varsayıyoruz. Burada şu soruları soruyoruz: bu deney için bazı alternatif dizaynlar hangileridir? Bu alternatif dizaynlara tekabül eden doğrusal modeller nelerdir? Bu deneyler nasıl analiz edilebilir? Ve en önemlisi bu deneyler soruşturmacının sorularına ne derece cevap veriyor?

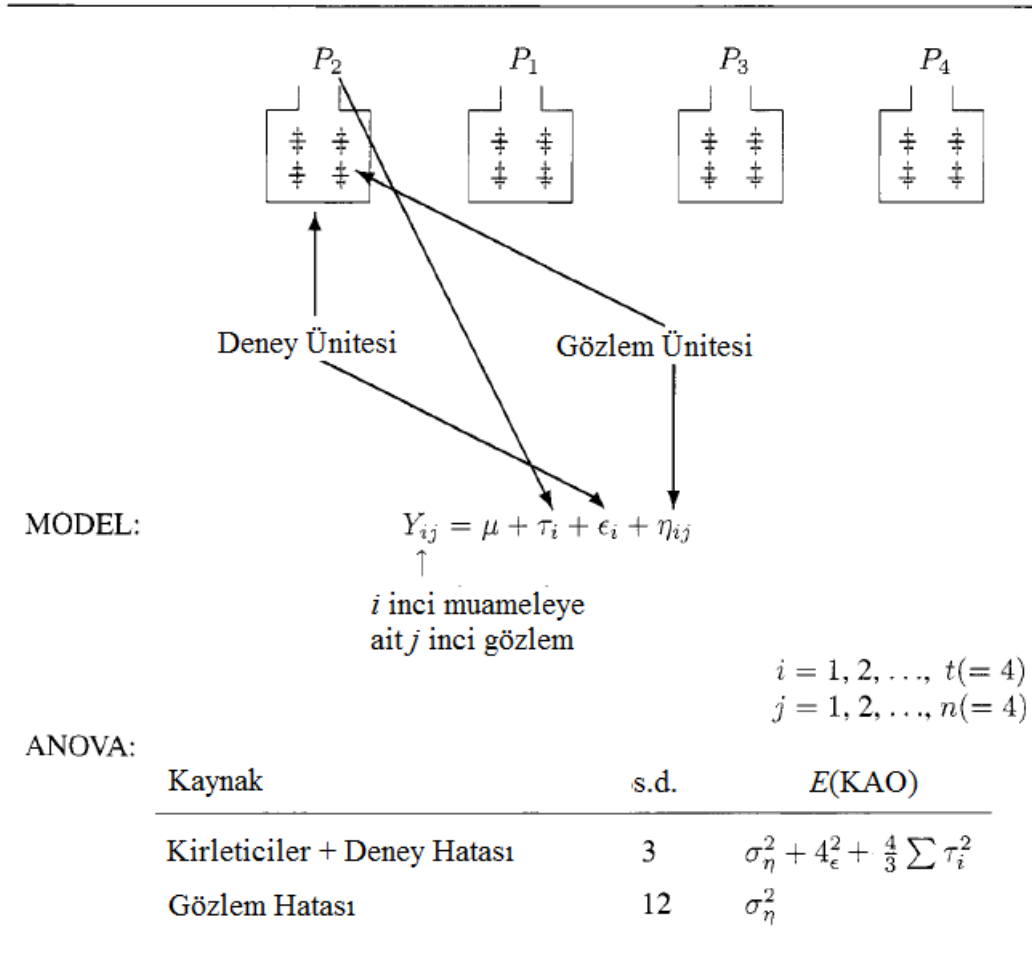
Dört Deney Durumu

Aşağıda dört olanaklı (iyi olması gerekmiyor) deney planını ana hatlarıyla şematize edeceğim ve mütenasip doğrusal modelleri ve ilişkin varyans analizi tablosunu ana hatlarıyla vereceğim. Elbette bu deneyi düzenlemenin başka yolları da vardır fakat burada vereceğim dört durumu kullanarak bu düzenlemelerin ve ilişkin modellerin ve yapılacak analizlerin arasındaki farklılıkları vurgulayacağım.

Deney I

















Birinci deney durumumuzda dört kirlilik odası kullanılıyor, her odada dört fide var. İçerisine dörder fide yerleştirilmiş odalara kirleticiler rastgele atanıyor. Bir odaya belirli bir kirletici uygulandığı için bu oda (veya alternatif olarak içerisindeki dört fide) deney ünitesini (DÜ) teşkil ediyor, her fide gözlem (veya örnekleme) ünitesini (GÖ) teşkil ediyor. Böyle olunca, muamele etkisi ve deney hatası birbirine “karışmaktadır” veya birbirinden ayrılmazlardır;

öyle ki bunu aşağıdaki model eşitliğinde muamele etkisi (τ) ve deney hatasının (ϵ) aynı alt indise sahip olmasıyla görebiliyoruz. Bu bizi elbette ki total kareler toplamının, $KT(\text{Total})$, yalnızca iki bileşene, $KT(\text{Kirleticiler} + \text{Deney Hatası})$ ve $KT(\text{Gözlem Hatası})$ olmak üzere parçalanmasına götürüyor. Bunların beklenen kareler aritmetik-ortalamaları, $E(\text{KAO})$, açıkça gösteriyor ki muamele etkilerine dair hipotezleri test etmek için meşru/uyarlı bir hata terimine sahip değiliz. Yani muamele etkilerinin birbiriyle özdeş (ve sıfıra eşit) olduğunu iddia eden sıfır hipotezi altında, iki KAO aynı beklenen değere sahip değildir. Bu bakış açısına göre bu deney yeterlikli değildir: Bu deney soruşturmacının sorularına cevap verebilemez.



Deney II

Bir manada Deney I ekstrem bir durumu temsil ediyordu, diğer bir ekstrem durum ise şimdikiidir. Her fide ayrı bir kirlilik odasına koyuluyor, her kirletickiye rastgele dört oda atanıyor. Böylece DÜ ve GÜ özdeş olacağından dolayı iki hata tipi birbirinden ayıramayacaktır. Bunu aşağıdaki model eşitliğinde görebiliyoruz. Mamafih, her iki hata muamele etkisinden ayrılabilir ve bu sayede muamelelere ait hipotez testlerini yapmak mümkündür.

P_1	P_3	P_2	P_2	P_3	P_3	P_1	P_4
							
P_2	P_1	P_3	P_4	P_2	P_4	P_4	P_1
							

Deney Ünitesi
 = Gözlem Ünitesi

MODEL:
$$Y_{ij} = \mu + \tau_i + \underbrace{\epsilon_{ij} + \eta_{ij}}_{e_{ij}}$$

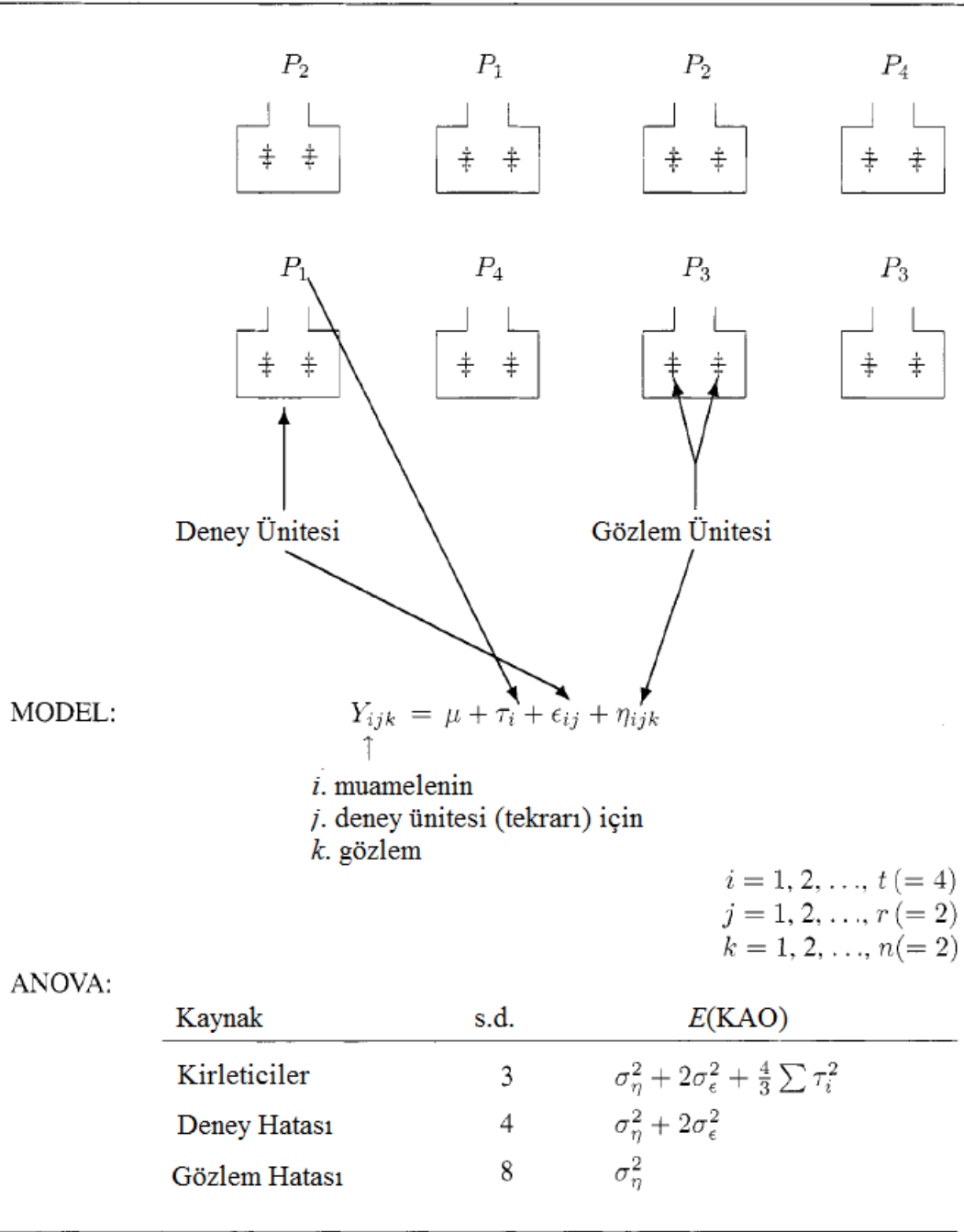
$i = 1, 2, \dots, t (= 4)$
 $j = 1, 2, \dots, n (= 4)$

ANOVA:

Kaynak	s.d.	$E(KAO)$
Kirleticiler	3	$\sigma_\eta^2 + \sigma_\epsilon^2 + \frac{4}{3} \sum \tau_i^2$
Hata (Deney + Gözlem)	12	$\underbrace{\sigma_\eta^2 + \sigma_\epsilon^2}_{\sigma_e^2}$

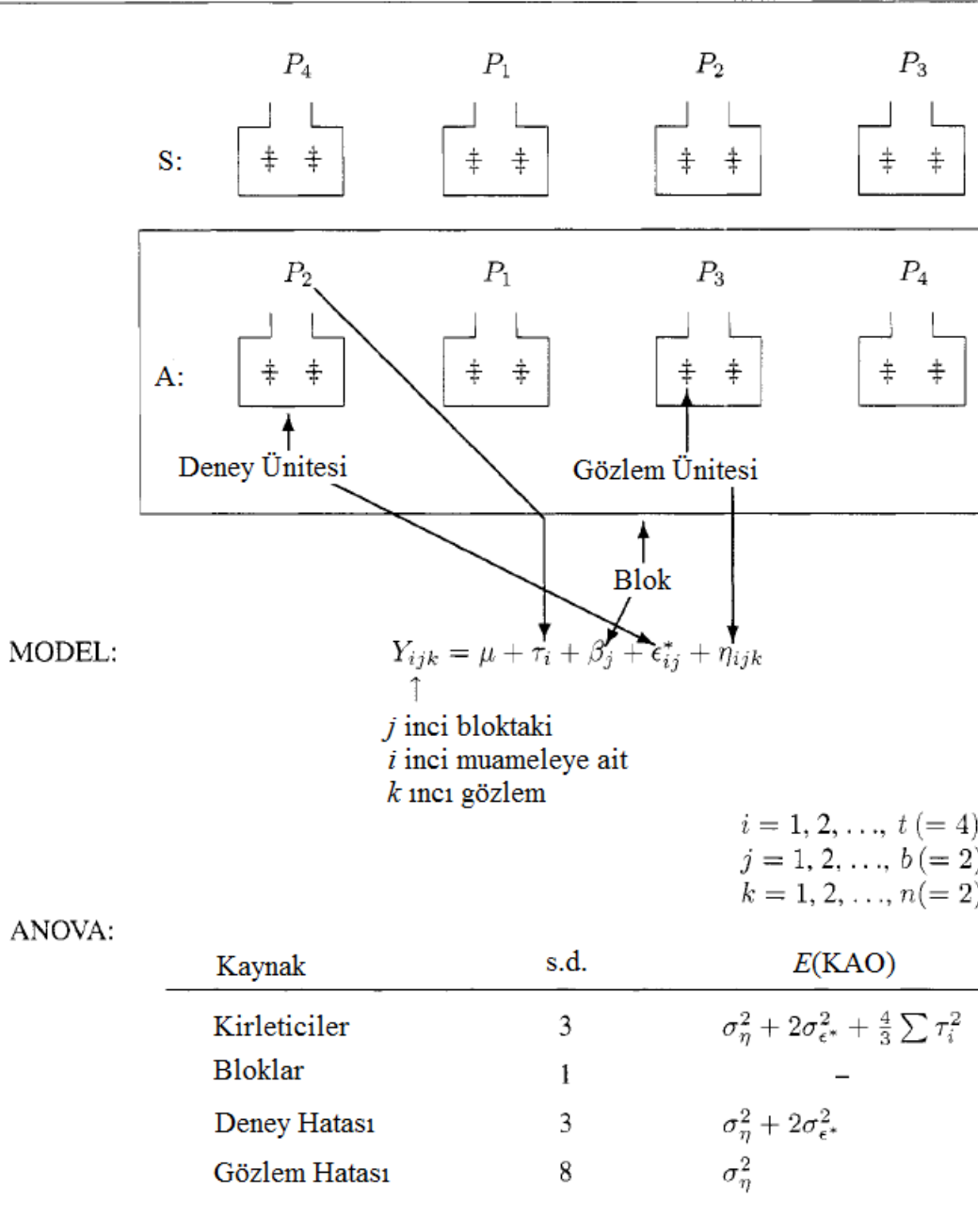
Deney III

Bu deney durumunda her kirleticinin uygulandığı iki oda var ve her odada iki fide var. Aynı kirleticisi ile muamele edilen odaların (DÜ) arasındaki değişim/varyasyon deney hatasının bir “ölçüsüdür”, oysa bir oda içindeki fideler (GÜ) arasındaki değişim/varyasyon gözlem veya örnekleme hatasının bir “ölçüsüdür”. Her iki tip hata birbirinden ayrılabilir gibi aynı zamanda kirleticisi etkilerinden de ayrılabilir. Bu aşağıdaki model eşitliğinde ve varyans analizi tablosunda formel olarak ifade edilmektedir.



Deney IV

Sonuncusu olarak bu deney durumunda Deney III durumunun bir değişimini sunuyoruz. Kirlilik protokolü her kirleticisi ile dörder kirlilik odasında bir kere sabah (S) ve bir kere de akşam (A) icra ediliyor. Bitkilerin günlük ritimlerinden dolayı sabah ve akşam fideler arasında sistematik farkların olması beklenir. Bu sistematik farklar, dört odalık iki kümeyi birer blok olarak addetmek suretiyle “elemine” edilebilir. Dahası, bu düzenleme bizi deney hatasının indirgenmesine (aşağıdaki tabloda σ_{ϵ}^2 yerine σ_{ϵ}^{2*} yazılarak belirtilmiş olarak) götürecektir.



Deney I, Deney II ve Deney III durumları tamamen rastgeleştirilmiş dizaynın farklı versiyonlarıdır ve Deney IV durumu bir rastgeleştirilmiş tam blok dizaynını temsil ediyor. Bu farklı argümanların nasıl farklı modellere ve böylece farklı analizlere yol açtığına dikkatini çekerim. Bu tartışma aynı zamanda daha önce ifade etmiş olduğum şu mevzunun da anlaşılmasına yardım etmiş olmalı. Geçerli istatistiksel çıkarımların yapılabilmesini sağlamak istiyorsak eğer, analizi deney dizaynı ile birlikte mülhaza etmek zorundayız. Burada hangi düzenlemenin en iyisi olduğunu tartışmayacağım. Ancak I nolu düzenlemenin kullanılmaması gerektiğini söyleyeceğim. II, III ve IV nolu düzenlemelerin kullanılmasında ise tamamıyla uygulamaya ait mülhazalar ve bizim burada bahsetmediğimiz başka koşullar belirleyici olabilir.

Deneyin Planlanması: Üzerinde Düşünülecek Şeyler

Önceki kısımlarda dizayn prosesinin muhtelif hususlarını tartıştım. Bu hususların bir deneyin planlanmasına özenli ve sistematik bir tarzda yaklaşmak ile nasıl bağlantılı olduğunu ve bu yüzden niçin önemli olduğunu görmüş olduk. Bu noktanın önemini vurgulamak maksadıyla deneyleme adımlarının önemli hususiyetlerini aşağıda özetleyeceğim.

1. Ereğin (veya ereklelerin) ifadesi:

Soruşturulacak problemin en azından bir genel formülasyonu, sonraki adımlara geçmeden önce esansiyeldir. Eğer aynı anda soruşturulmayacak çoklu erekler varsa, hatta bu daha önemlidir.

2. Tematik modelin formülasyonu:

Burada önemli olan nokta, ölçülen yanıtı potansiyel olarak etkileyen tüm faktörlerin bir listesinin hazırlanmasıdır. Bu muamele faktörlerinin seçimini ve olanaklı içkin faktörlerin teşhisini kapsar. Deneyin hacmini makul bir düzeyde tutabilmek adına, belki bazı içkin faktörleri yalnızca bir düzeyde sabitlemek gerekebilir. Mesela farklı yaşlardaki bay ve bayan deneklerden ziyade aynı yaştaki bayan denekler gibi. Böyle yapmak diğer taraftan elbette muamele etkilerine dair sonuçların uygulanabilirliğini kısıtlayacaktır, yani deneyin çıkarım uzayını daraltacaktır. Deneyin gidimini daraltmanın başka bir önemli – belki de negatif – tarafı da muamele faktörleri ve içkin faktörler arasındaki olanaklı interaksiyonları soruşturmayı gayri kabil kılmasıdır.

3. Faktör düzeylerinin seçimi:

Burada hem muamele faktörlerini hem de içkin faktörleri nazarı dikakte alıyoruz. Ve ne zaman faktör “düzey”lerine atıfta bulunuyorsak, o faktörün farklı ifadelerini kastediyoruz demektir bu. Mesela, “sıcaklık” faktörünün 200°C ve 300°C gibi farklı ayarlamaları, mesela, “kanseri tedavisi” faktörü için radyasyon ve kemoterapi gibi. Hem muamele faktörleri hem de içkin faktörler için kaç düzeyin ve hangi düzeylerin seçileceğini özenle mülhaza etmek önemlidir. Yapılacak seçim deneyin hacmini etkileyeceği kadar deneyden çıkarılacak enformasyonun tipini ve miktarını da etkileyecektir.

Mesela bir nicel muamele faktörü için, yanıt fonksiyonundaki olanaklı eğriliği değerlendirmek maksadıyla ikiden çok düzey kullanmak önemli olabilir. Dahası, düzeyler, muamelenin pratik gidimi, değişim aralığının alt ve üst limitlerini de içerecek şekilde seçilmelidir. Ve dahası, düzeyler, yanıtındaki olanaklı bir farkı yakalayabilecek kadar birbirinden uzak seçilmeli ve fakat aynı zamanda yanıtındaki arada gerçekleşecek olanaklı bir değişimin yakalanamayacağı kadar da çok uzak seçilmemelidir. Mesela, sıcaklık faktörünün 200°C ve 225°C düzeyleri arasında fark olmayabilir fakat 200°C ve 250°C yanıtındaki olanaklı bir değişimi yakalamak için birbirine yeterince uzak olabilir. Öte taraftan, 200°C ve 400°C düzeyleri birbirine çok uzak olabilir çünkü önemli değişimler 250°C ve 300°C arasında vuku buluyor olabilir.

Benzer mülhazalar aynı zamanda nitel faktörler için de geçerlidir. Muamele faktörleri için hangi düzeylerin seçileceği çoğu zaman oldukça açıktır fakat içkin faktörlerin düzeylerinin seçiminde deneyin hacmi önemli olabilir. Mesela, kirleticilerin farklı tiplerinin bitkilere olan etkilerini soruşturmak için başlangıçta deneyin kapsamı ağaçlar

ve ağaç fideleri ile kısıtlanabilir. Ve iğne yapraklıların farklı türlerini deneye dahil etmek yerine, iğne yapraklılardan bir tür ve yaprak dökkenlerden bir tür seçmek yararlı olabilir. Müteakip bir deneye o zaman sebzelerin mesela farklı tipleri gibi başka bitkiler dahil edilebilir.

4. Yanıtın ölçülmesi:

Birinci madde olarak problemin ifade edilmesi, sadece gözlem ünitesini (GÜ) ve yanıt değişkenini tanımlamakla kalmaz aynı zamanda yanıt değişkeninin nasıl ölçülmesi gerektiğini de tanımlar. Bu ölçümler sürekli veya kesikli olabilir. Bazı durumlarda farklı ölçüm tipleri mümkündür ve hangisinin kullanılacağına karar vermek gerekir. Mesela, kirliğin verdiği hasarı değerlendirmek için her bitkideki hasarlı yaprak alanını ve total yaprak alanını ölçebiliriz ve sonra hasarlı yaprak alanı yüzdesini elde edebiliriz. Eğer bunu külfetli bulursan, alternatif olarak, diyelim 5-puanlık bir skorlandırma sistemi kurabilirsin ve sonra her bitkiye görsel olarak bir skor atayabilirsin, yani her bitkiyi almış olduğu hasarın miktarını en iyi yansıtan beş kategoriden birine koyarsın. Aşık ki bir sürekli ölçüm en enformasyonlu olandır. Hem daha enformasyonlu ve hem de nispeten daha basit bir ölçüm olsun istiyorsan, o zaman 5-puanlık sistem yerine diyelim 10-puanlık bir sistem seçebilirsin, ki bu şekilde detaylandırılmış bir subjektif skorlandırma sisteminin ne kadar tekrar edilebilir olduğu gerçeğini de aklında tormalısın. Diğer aşırı uçta, elbette, ikili sistem vardır – hasarlı ve hasarsız – ki bu da soruşturulan kirleticilerin arasındaki herhangi farkları tespit etmek için çok kaba kalabilir. Özetlersem, ölçümün seçimi istatistiksel bakımdan olduğu kadar uygulama/pratik bakımından da önemlidir.

5. Hata-kontrol dizaynının spesifikasyonu:

İçkin faktörlerin tespit edilmesi ve bir veya daha çok sayıdaki içkin faktörün bir kaç düzeyinin dahil edilmesi, kullanılması gereken hata-kontrol dizaynının tipini büyük ölçüde zaten belirleyecektir ki bu mesela blok dizaynının bir formu olacaktır. Dizaynı daha katıyetle belirtmek için DÜ lerini ve GÜ lerini ve muamelelerin deney ünitelerine nasıl uygulanacağını tespit etmeliyiz. Eğer birden çok muamele faktörü mevcutsa, bu bilhassa önem kazanır. Deney ünitelerinin (DÜ) farklı tipleri olabilir (bak Misal 7) ve bu yüzden farklı hata-kontrol dizaynları söz konusu olabilir, mesela blok dizaynına karşı bölünmüş-parsel dizaynı gibi.

6. Bir modelin formülasyonu ve analize ait hususlar:

Şunun önemini yeniden vurgulamak istiyorum ki bir modelin ve en azından yapılacak analizin unsurlarının detaylıca belirtilmesi, bir deneyin planlanmasında can alıcı hususlardan biridir. Bu mülahazalar bize bilimsel-araştırma hipotezinin istatistiksel bir çerçevede değerlendirilip değerlendirilemeyeceğini ve nasıl değerlendirileceğini söyleyecektir. Başka maksatların yanı sıra, enformasyonlu parametrik fonksiyonlara dair istatistiksel hipotezleri test etmek veya güven aralıklarını elde etmek için mütenasip hata terimlerini tespit edebiliriz. O zaman bu hata terimleri için yeterli serbestlik derecelerinin (s.d.) olup olmadığını değerlendirebiliriz. Bu mülahazalar, nihayetinde, bizi ya deneyin tatmin edici bir şekilde planlanmış olduğu neticesine götürür ya da başarılı bir deney için hangi değişikliklerin yapılması gerektiğini görmüş oluruz. Bu son adımın gerçekten planlama prosesinin en önemli aşaması olduğunu ne kadar vurgulasam azdır.

Bilimci ve İstatistikçi Arasında İşbirliği

Bilimsel soruşturmanın muhtelif adımlarını yukarıda tartıştım ve deneyin planlanmasına özel önem attım. Deneyin planlanması prosesi/süreci öğrenme-çalışmasının tematik bilimci/soruşturmacı ve istatistikçi arasında yakın bir işbirliğini gerektirir. Şimdi böyle bir işbirliğinin bazı hususiyetlerini, bir önceki kesimde tartışılan noktaların paralelinde ana hatlarıyla belirteceğim.

1. Ereğin ifadesi:

Bilimsel-araştırma erekleri ve hipotezleri belli bir tematik alandaki bilimsel-araştırma aktiviteleri bağlamında ortaya çıkar. Böyle olunca, bu ereklerin formülasyonu açık ki soruşturmacının başlıca sorumluluğudur. Mamafih, eğer bir deneyleme işine girişilecekse, bir istatistikçiyle iletişime geçmek için vakit kaybetmemek gerekir. Bunun başlıca sebebi bir deneyin planlanması ve icra edilmesi ait muhtelif adımlara dikkat çekmektir.

2. Tematik modelin formülasyonu:

Bu husus ta soruşturmacının başlıca sorumluluğudur. Bununla beraber, bu aşama, halihazırda, istatistikçinin yapılması düşünülen deneyin arzulan ve olanaklı çıkarım uzayı hakkında sorular sorabileceği bir zamandır. Bazen istatistikçi sorularıyla bilimsel-araştırmacının deneyinin hedeflerini açıklığa kavuşturmasına veya belki modifiye etmesine yardımcı olabilir. Bilhassa, potansiyel içkin faktörlerin mülahazası, deneyin hacmine dikkat çekecektir.

3. Faktör düzeylerinin seçimi:

Daha evvel söz konusu ettiğim üzere, bir deneye mana kazandırmak için muamele faktörlerinin ve içkin faktörlerin düzeylerinin özenle seçimi önemlidir. Burada yine istatistikçi soruşturmacının tematik bilgisine güvenmek durumundadır. İstatistiksel bir perspektiften bakınca, mesela muamele faktörlerinin belli kombinasyonlarının deneyde yer alması arzulanabilir fakat böylesi kombinasyonlar biyolojik, fiziksel veya kimyasal sebeplerle arzu edilmeyebilir hatta olanaklı olmayabilir veya tamamen uygulamaya/paratiğe matuf sebeplerle erişimi zor olabilir. Sonunda, deneyin ereklerinden ödün vermeksizin hem istatistiksel hem de tematik mülahazaları karşılayacak uzlaşmalara varılabilecektir.

4. Yanıtın ölçülmesi:

Çoğu durumda bilimsel-araştırmanın mevzusu bir boşluğa düşmez, benzer deneyler önceden yürütülmüştür. Bu nedenle, muamelelere ait yanıtın hangi prosedürlerle ölçüleceği üzerinde bir fikir birliği mevcuttur. Farklı deneylerin çıkımları arasında karşılaştırmalar yapmayı olanaklı kılmak için bu prosedürlerle genel olarak riayet edilir. Yine de bir deneyin sonuçlarını mülhaza etmek için alternatif yollar mevcut olabilir. Bunu uygulama/pratik ve ekonomi zeminlerinde düşünmek gerekir.

5. Hata-kontrol dizaynının spesifikasyonu:

Deneyin bu safhası soruşturmacı ve istatistikçi arasında yakın işbirliği gerektirir. Burada deneyin fiilen nasıl icra edilmesi gerektiğine dair sorular sorulur. Bu noktada şu sorulara cevap verilmelidir: Deney üniteleri (DÜ) nelerdir? Gözlem üniteleri (GÜ) nelerdir? DÜ leri ne kadar homojendir? DÜ leri daha homojen alt gruplara (bloklar) bölünebilir mi

veya bölmek gerekir mi? Deneyin farklı aşamalarda icra edilmesi gerekiyor mu, yani farklı zamanlarda veya yerlerde? Muameleler DÜ lerine nasıl atanacak? Muamelelerin DÜ lerine atanmasında bir farklılık olacak mı? Bu sorulara ve belki ilave sorulara verilecek cevaplar mütenasip hata-kontrol dizaynının seçimine yardımcı olacaktır veya belli bir hata-kontrol dizaynının deneyin gereksinimlerine göre modifiye edilmesine yardımcı olacaktır. Bu mülahazaların önemli bir hususiyeti burada halihazırda seçilmiş içkin faktörlere ilaveten spesifik-olmayan faktörlerin de teşhisidir. Sayfa xx teki figürde gösterilen istatistiksel üçgenin bir köşesini temsilen, hata-kontrol dizaynının geliştirilmesi deneyin fiilen üzerinde icra edileceği sahneyi kurar. Bundan dolayı, dizaynın tüm noktaları üzerinde soruşturmacı ve istatistikçi arasında tam bir fikir birliğinin sağlanması gerekir. İki taraf arasında sağlıklı bir iletişimin sağlanması ve yanlış anlamalara mahal verilmemesi bakımından, deneyin fiziksel konfigürasyonunun bir şematik resmini göstermek üzere “Dört Deney Durumu” başlıklı alt kısımda dört farklı deney durumu için vermiş olduğum tablolardakine benzer bir diyagramın çizilmesini şiddetle tavsiye ediyorum.

6. Bir modelin formülasyonu ve analize dair hususlar:

Deneyden elde edilecek verinin analizinde kullanılacak model, büyük ölçüde muamele ve hata-kontrol dizaynları tarafından belirlenir ki bunlara ilişkin hususları yukarıda tartıştım. Aslında, her hata-kontrol dizaynı için bir doğrusal model belli varsayımlar altında türetilir. Bu varsayımlar belli interaksyonların mahiyetine ilişkin olabilir. Tematik bilgi, bilhassa, hangi muamele-içkin faktör interaksyonlarının gözardı edilebileceğine karar vermede çok yardımcı olabilir.

İstatistiksel analizin temel unsurları ve muhtelif unsurların bilimsel-araştırma hipotezinin muhtelif hususlarıyla nasıl bağlantılı olduğu üzerinde ne kadar düşünsek azdır.

7. Deneyin icrası:

Her ne kadar deneyin icrasından soruşturmacı sorumlu olsa bile idealinde istatistikçinin de yardımını dışlamamak gerekir. Her ikisi de üzerinde anlaşılan protokolün takip edildiğinden emin olmalıdır. Mesela, yanlışlığı veya karışmayı önlemek için mütenasip muamele rastgeleleştirmesinin yürütülmesi önemlidir. Aynı zamanda, özenli bir planlamaya rağmen eğer deneyin orjinal formunda icra edilebilemeyeceği ortaya çıkarsa, orjinal soruları tamamen değilse bile çoğunlukla hala cevaplayabilecek şekilde deneyi modifiye etmenin veya daraltmanın yollarını aramak gerekecektir. Soruşturmacı ve istatistikçi arasında yakın bir konsültasyon yapılmadan, keyfi bir daraltma arzu edilmeyen vargılara ve aslında deneyin başarısızlığına yol açabilecektir.

8. Verinin toplanması ve kaydedilmesi:

Verinin formal analizinden önceki son adımdır bu. Önemli olan sadece veriyi tesis edilen protokolü takip ederek özenle ve mümkün olduğunca tam olarak toplamak değil aynı zamanda analiz için kullanılacak yazılımın gerektirimlerini de dikkate alarak aynı özenle kodlamak ve organize etmektir. Her ne kadar istatistiksel yazılım paketleri kayıp gözlemleri çoğu durumda hesaba katabiliyor olsa da bu bizi özensizliğe teşvik etmemelidir çünkü kayıp gözlemler analizde ve analiz sonuçlarının yorumlanmasında gereksiz komplikasyonlara yol açabilmektedir.

Çıkarımın Genel Prensipleri ve İstatistiksel Analiz Tipleri

Önceki kesimlerde yaptığımız tartışma şunu açığa çıkardı ki dizyn-edilmiş bir deneyden elde edilen verinin analizi ağırlıklı bir doğrusal modele dayanır ki bu modelin de deneyin kendi yapısını yansıtması gerekir. Aslında, böyle bir doğrusal modelin formülasyonu, bütün bu deneyleme uğraşında çok önemli bir husustur ve ilerleyen bölümlerde bunu yeniden ele alacağım. Mütenasip bir doğrusal model formüle edildikten bir sonraki adım bu modele ilişkin varyans analizinin veya bazı durumlarda görülebileceği üzere varyans analizlerinin yapılmasıdır. İşte bu bize, ele almış olduğumuz belirli durum için mütenasip gördüğümüz istatistiksel çıkarımları yapmaktaki bir dayanak sağlayacak.

Genel Model

Daha evvel zikretmiş olduğum üzere, bir deney dizaynına ait bir doğrusal model genel olarak bileşenlerin üç tipini kapsar: muamele bileşenleri, dizayn bileşenleri ve hata bileşenleri. Bir doğrusal model sayfa xx ve sayfa xx te yazılmış olanlardan daha formel olarak şöyle yazılabilir:

$$Y = I\mu + \sum_{i=1}^t X_i \tau_i + \sum_{j=1}^b U_j \beta_j + \sum_{k=1}^c Z_k \epsilon_k + \sum_{l=1}^d W_l \eta_l$$

ki bu modelde yer alan sembollerin temsil ettiği terimler şunlardır:

Y : gözlemlerin $s \times 1$ boyutlu vektörü,

μ : genel aritmetik-ortalama,

$\tau_i = (\tau_{i1}, \tau_{i2}, \dots, \tau_{ia_i})'$: “muamele etkilerinin” $a_i \times 1$ boyutlu vektörü ($i = 1, 2, \dots, t$),

$\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jb_j})'$: “bloklama etkilerinin” $b_j \times 1$ boyutlu vektörü ($j = 1, 2, \dots, b$),

$\epsilon_k = (\epsilon_{k1}, \epsilon_{k2}, \dots, \epsilon_{kc_k})'$: deney hatalarının $c_k \times 1$ boyutlu vektörü ($k = 1, 2, \dots, c$),

$\eta_l = (\eta_{l1}, \eta_{l2}, \dots, \eta_{ld_l})'$: gözlem hatalarının $d_l \times 1$ boyutlu vektörü ($l = 1, 2, \dots, d$),

I : $s \times 1$ boyutlu birim vektör,

X_i : $s \times a_i$ ($i = 1, 2, \dots, t$) boyutlu matris ki muamele faktörleri ve bunların interaksyonları (ve muamele-içkin faktör interaksyonları) gibi muamele yapısını temsil eder,

U_j : $s \times b_j$ ($j = 1, 2, \dots, b$) boyutlu matris ki hata-kontrol dizaynına ait hususları yani içkin ve spesifik-olmayan faktörlerle telkin edilen muhtelif bloklama düzeneklerini yansıtır,

Z_k ve W_l : sırasıyla $s \times c_k$ ($k = 1, 2, \dots, c$) ve $s \times d_l$ ($l = 1, 2, \dots, d$) boyutlu matrisler ki hata yapısını yansıtır ki bu yapı DÜ lerinin ve GÜ lerinin mahiyeti ve bunlara ve bunların ölçüm ve gözlem proseslerine ilişkin muhtelif hata tipleri tarafından belirlendiği kadar kısmen bloklama düzenekleri ve rastgeleleştirilmenin muhtelif safhaları tarafından belirlenir.

Varyans Analizinin (ANOVA) Anahatları

“Dört Deney Durumu” başlıklı alt kesimde Deney III için kurulan modele (Sayfa 19) dayanarak, çok kesinlikli terimlerle olmasa dahi, varyans analizinin genel yapısını anahatlarıyla aşağıdaki tabloda verebiliriz:

Kaynak		s.d.
Total		$s - 1$
DÜ leri arasında		$m - 1$
	Muameleler arasında	
Muamele dizaynı	$\begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_t \end{bmatrix}$	t_1 t_2 \vdots t_t
	Muameleler içindeki DÜ leri arasında	
Hata-kontrol dizaynı	$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_b \\ \epsilon_1 \\ \vdots \\ \epsilon_c \end{bmatrix}$	d_1 d_2 \vdots d_b e_1 \vdots e_c
	DÜ leri içinde $s - m$	
Gözlem Dizaynı	$\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_d \end{bmatrix}$	o_1 \vdots o_d

KT(Total) temelde, m DÜ olduğunu varsayarak, sırasıyla $m - 1$ ve $s - m$ serbestlik dereceleri ile KT(DÜ Arasında) ve KT(DÜ leri İçinde) larına parçalanır. KT(DÜ leri Arasında) daha sonra sırasıyla $t_i = \sum t_i$ ve $m - t$ serbestlik dereceleri ile KT(Muameleler Arasında) ve KT(Muameleler İçindeki DÜ leri Arasında) larına parçalanır. KT(Muameleler arasında) nın da yeniden parçalanması olanaklıdır ve mesela belli muamele kontrastlarına dair hipotezlerin testi ile ilgilenildiği zaman veya muameleler faktöriyel bir yapıda olduğu zaman bu arzu edilebilir. Böyle bir parçalama büyük ölçüde muamele dizaynı tarafından belirlenir. KT (Muameleler İçindeki DÜ leri Arasında) nın parçalanmasını hata-kontrol dizaynı belirler ki bu dabloklama faktörleri ve farklı rastgeleleştirmelerin bir fonksiyonu olarak deney hatalarına ilişkin muhtelif kareler toplamlarına yol açar. Muameleler hakkında istatistiksel çıkarımlar yapmak için elbette farklı KT(Deney Hatası) kullanılacaktır. Nihayetinde, KT (DÜ leri içinde) muhtelif örnekleme ve alt-örnekleme tiplerince yani gözlem yapısı tarafından belirlenir.

Bu genel tartışmayı “Deney ve Bir Model” başlıklı alt kesimde verdiğim misal üzerinden aydınlatmaya çalışacağım. Sayfa xx teki model eşitliği, şu mütakabiliyetler yoluyla Sayfa xx teki formda ifade edilebilir:

Model (Sayfa 6)	Model (Sayfa 19)
P_i	τ_1
F_k	β_1
$(PF)_{ik}$	τ_2
C_j	ε_1
ε_{ijk}	ε_2
η_{ijkl}	η_1

Kirleticiler bakımından odaların DÜ leri olması kirletici etkileri hakkındaki hipotezlerin testi için mütenasip hata terimini (Sayfa xx teki tabloda Hata 1 olarak belirtildiği üzere) $KT(\varepsilon_1)$ in sağladığını göstermektedir. Diğer taraftan, $KT(\varepsilon_2)$ kirletici \times aile interaksiyonu etkileri hakkındaki hipotezlerin testi için hata terimini (Sayfa xx teki tabloda Hata 2 olarak belirtildiği üzere) sağlamaktadır. Örnekleme hatası ise $KT(\eta_1)$ tarafından sağlanmaktadır.

Yukarıda misallendirilerek aydınlatıldığı üzere, varyans analizinin önemli bir hususiyeti muamele ve blok etkileri gibi sistematik etkilerin rastgele etkilerden veya hata etkilerinden ayrılmasıdır. Bu sadece hipotez testi bağlamında değil aynı zamanda güven aralıklarının tesisi ve muamele karşılaştırmaları için standart hataların elde edilmesi için de önemlidir. Genel formdaki (Sayfa xx) model eşitlikleri ve hata bileşenlerinin özellikleri (veya bunlar hakkındaki varsayımlar) ile beraber münasipçe dizayn edilmiş bir deneyin varyans analizi bizim $\sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2, \dots, \sigma_{\varepsilon_c}^2$ ve $\sigma_{\eta_1}^2, \sigma_{\eta_2}^2, \dots, \sigma_{\eta_c}^2$ varyans bileşenlerini (veya onların doğrusal fonksiyonlarını) kestirebilmemizi sağlar ki bunları sonra yukarıda zikredildiği üzere kullanabiliriz. Bu varyans bileşenleri hakkında bilgi sahibi olmak ileriki deneysel stratejileri tesis etmek için de çoğu zaman oldukça yararlıdır. Mesela her muamele için tekrarlamaların mütenasip sayılarını ve DÜ leri içindeki örnekleme miktarını belirlemek gibi. Özetleyecek olursak, deney verisinden istatistiksel çıkarım, ister test etme formunda ister kestirim formunda olsun, bir altta-yatan doğrusal modele dayanır. Böyle bir doğrusal modelin varlığında en küçük kareler metodu kullanılarak varyans analizi tablosunun yanısıra münasebetli parametrelerin kestirimleri de elde edilir. Bütün bunlarda, rastgeleleştirme teorisinin unsurları önemli rol oynar.

Deney Dizaynlarına Dair Başka Mülahazalar

Şimdiye kadar yaptığım bu genel tartışmada, başlıca, bir muamele dizaynı, bir hata-kontrol dizaynı ve bir gözlem (örnekleme) dizaynından oluşan bir deney dizaynının soruşturmacının sorularının cevaplanabileceği bir yoldan seçilmesi gerektiği üzerinde durdum. Dizayn ve ilişkin istatistik analiz bağlantısı üzerinde durarak bu bağlantının nasıl kurulabileceğini göstermeye çalıştım. Gördük ki çoğu durumda belirli bir bilimsel ereğe farklı deney tipleri ile ulaşmak mümkün. Bunun bir misalini “İstatistiksel Üçgen: Öğrenme-çalışması 2” başlıklı kesimde (Sayfa xx) vermiştim. Öyle bir durumda aşık bir soru şudur: Hangi deney kurulumunu seçmeliyim? Veya bu durum için “en iyi” deney dizaynı nedir?

Soru direk sorulmuş olsa bile, rakip dizaynları karşılaştırmak için geliştirilmiş farklı kriterlerin mevcudiyetinden dolayı cevap o kadar da kolay olmayabilir. Bu kriterlerden bazılarını kısaca değineceğim. Bazılarını ise daha sonra daha detaylıca tartışacağım.

En önemli kriterlerden biri optimallik veya daha doğrusu varyans-optimalliği kriteridir. Bununla, muamele etkilerinin doğrusal kombinasyonlarının kestiriminde maksimum kesinliği (bir manasıyla) kastediyorum. Ancak çoğunlukla önümüze tüm bu varyansların

minimizasyonun bir fonksiyonu ile alakalı *A*-optimalliği, *D*-optimalliği veya *E*-optimalliği gibi muhtelif optimallik kriterleri çıkacaktır ve kullanmak için en iyi kriterin hangisi olduğu her zaman açık olmayacaktır.

Bir dizaynın sahip olmasında yarar olan bir özellik ortogonalliktir. Bu özellik, soruşturmacının basitçe muamele aritmetik-ortalamalarına bakarak karşılaştırmalar yapabilmesine izin verir. Ortogonallik özelliği aynı zamanda bize kareler toplamlarının kolayca hesaplanabileceği yegane/tekil bir varyans analizi sağlar. Hızlı kompüterler çağında bu belki manalı gelmeyebilir ama diğer taraftan sonuçların daha kolay ve daha şeffaf yorumlanmasına yardım eder.

Mevcut dizaynların ve yaygınca kullanılan dizaynların çoğu ortogonaldir fakat eğer ortogonallik sağlanamazsa dengelilik özelliği sağlanmaya çalışılır. Burada varyans-dengeli dizaynlara normalleştirilmiş muamele karşılaştırmalarının aynı kesinlikle kestirilmesi manasında atıfta bulunuyorum. Dengenin diğer nosyonları, bilhassa faktöriyel deneyler bağlamında, mevcuttur ki faktöriyel deneylerin tüm bu deney dizaynı tartışması içinde önemli bir yeri vardır.

Burada anabileceğimiz mesela bağlantılılık, etkinlik ve yansızlık gibi pek çok başka kriterler ve özellikler de var fakat bunların tartışmasını daha sonra bunlara gerçekten açıkça gereksinim duyduğumuz yerde yapacağım.

Daha evvel zikredildiği üzere, bir deney dizayn etmenin başlıca ereklerinden biri muamele etkileri arasındaki karşılaştırmaları olanaklı olduğunca kesinlikle kestirmektir. Buna birden çok yoldan ulaşılabilir. Mesela tekrarlama ve bloklama veya deney ve ölçüm tekniklerinin rafine edilmesi gibi. Bir diğer önemli düzenek ise final yanıtlar ile korelasyonlu olduğu bilinen GÜ lerine ait ölçümler formundaki destekleyici enformasyonu kullanmaktır. Burada kritik nokta, GÜ leri üzerinde yapılan bu gözlemlerin GÜ lerine uygulanan muamelelerden etkilenmiyor olması gerektiğidir. Bu destekleyici enformasyonun veri analizine dahil edilmesiyle “DÜ lerinin daha üniform hale getirilmesi” ve böylece değişkenliğin azaltılması sağlanmaktadır. Bu durum için kullanılacak istatistiksel tekniğe kovaryans analizi adı verilir. Kovaryans analiz her tip deney dizaynı ile birlikte kullanılabilir.

Destekleyici enformasyon her daim elverişli olamayabilir veya elverişli olsa bile her zaman avantajlı olamayabilir veya elde edilmesi pahalıya gelebilir. Oldukça aşık ki bazı durumlarda dizayn edilmiş deneyler için maliyete dair mülahazalar öne çıkar. Bu problemle ilişkili olarak yalnızca şunu söyleyebilirim: soruşturmacının ereklerini karşıladığı sürece dizaynı olanaklı olduğunca basit tutmak gerekir. Deneyin icrasını olduğu kadar, basitlik, aynı zamanda verinin analizini ve sonuçların yorumlanmasını da etkileyen bir gerektirimdir. Fakat burada basitlik teriminin manasına dikkat etmelisin belirli bir durum için en basit deney belki de en kompleks olanıdır esasen.

Basitlik nosyonu/mefhumu aynı zamanda bir diğer kavram ile de bağlantılıdır ki belki de şimdye değin zikrettiklerimin arasında en önemlisi olabilir: geçerlilik gidimi veya hedef popülasyon. Eğer geçerlilik gidimi çok dar ise deney dizaynı ziyadesiyle basit olabilir. Misal vermek gerekirse, eğer 20 günlük yaştaki loblolly çamı fidelerine ozonun etkisini 30 günlük bir periyotta günde 6 saat 5 ppm lik doza maruz bırakarak soruşturmak istiyorsak, hedef popülasyon ziyadesiyle küçüktür ve bir tamamen rastgeleleştirilmiş dizayn (“İstatistiksel Üçgen: Öğrenme-çalışması 2” başlıklı kesimde – Sayfa xx – betimlenenin bir benzeri) mütenasip bir dizayn olabilir. Lakin, eğer geçerlilik gidimini doğal ormanlara ve kirliliğin muhtelif formalarına değin genişletmeyi arzuluyorsak hem muamele dizaynı hem de hata-kontrol dizaynı çok daha karmaşık olacaktır eğer ki basitleştirilmiş varsayımlar yapılmazsa ve hedef popülasyon daraltılmazsa.

Şunu yeniden ifade etmek istiyorum ki şimdiye kadar yaptığım tartışmayı bir manasıyla yüzeysel tutmaya çalıştım çok detaylandırmaksızın. Maksadım bir deney dizayn edeceğin zaman farkında olmak zorunda olduğun, istatistiksel olan ve olmayan, hususları işaret etmektir. Dahası, gördük ki tüm gerektirimleri ve kriterleri her zaman karşılamak olanaklı değildir. Aslında, bazıları birbiriyle çelişkilidir. Öyleyse, uzlaştırma en iyi yoldur. Deney dizaynının temel fikirleri ve prensipleri iyi anlaşılmalı ve özümsemiş olsalar dahi sürekli yeni kriterler geliştirilmekte ve bu alana dahil olmaktadır.