

## HAFTA 10

### Gerçek permütasyon dağılımının yaklaşımı:

Mümkün test istatistiklerinin toplam sayısı, genel olarak iki örnekli durum için  $i$ . tedaviyi alan deneklerin toplam sayısı ve  $N = n_1 + n_2$  olmak üzere  $\binom{N}{n_1} = \frac{N!}{n_1!(N-n_1)!} = \frac{N!}{n_1!n_2!}$  = toplam test istatistik sayısı, çok büyük hatta ölçülü  $n_1$  ve  $n_2$  için olabilir.

Böyle bir durumda yaklaşık permütasyon dağılımı Monte Carlo Simülasyon çalışmasıyla elde edilir. Bu mümkün permütasyonlarının bir kısmı örnek olarak alınır, her bir durum için test istatistiğinin değeri hesaplanır ve gerçek permütasyon dağılımının bir yaklaşımı gibi bu permütasyon dağılımı kullanılır.

**Örnek 5.2 (permütasyon)**  $T$  hücreleri (veya  $T$ -lenfositler) bağışıklık sisteminde önemli rol olan beyaz kan hücreleridir. Bir tipi yüzeyi üzerinde  $CD4$  olarak adlandırılan moleküllere sahiptir; bu “yardımcı” hücreler virüsler gibi belirli mikro-organizmalara vücudun cevap vermesini düzenlerler. *HIV* kendi kendini bu  $CD4$  molekülüne bağlayabilir, izin verilen virüs bu hücrelere girer ve bulaştırır.  $CD4$  sayısı kanda dolaşan  $CD4$   $T$  hücrelerinin kaç tanesinin bir yansımasıdır. Yetişkinlerde normal  $CD4$  sayısı her milimetreküp kanda 500 ile 1500 aralığındadır. İleri *AIDS*'li benzer hastaların katıldığı küçük bir iki tedavili denemede denekler  $A$  ve  $B$  tedavilere rasgele atansın.  $CD4$  sayısına uyan artışlar (tedavi uygulandıktan sonra altı hafta boyunca alınan  $CD4$  sayıları) Tablo 5.10'da verildi. Gözlenen test istatistiği

$$T = \bar{Y}_A - \bar{Y}_B = -64$$

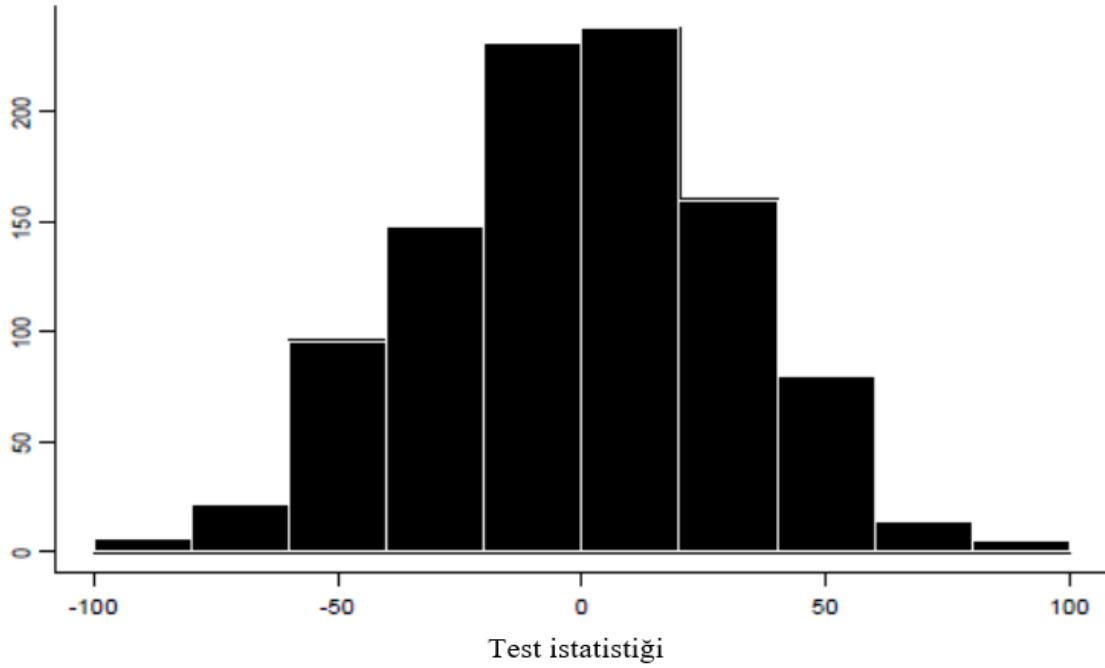
bulundu.

**Tablo 5.10:** *AIDS* hastalarının iki grubu için  $CD4$  sayılarının artışı

<b>A</b>	<b>A</b>	<b>B</b>	<b>B</b>
221	134	212	383
107	179	221	169
102	180	207	137
68	144	121	114
71	180	225	287

Böylece iki tedavi arasında bir farklılık olup olmadığını test etmek için basitçe test istatistiklerinin sayısı sayılabilir. Mümkün  $\binom{20}{10}=184756$  permütasyondan her birinden mutlak 64 değerinden büyük veya eşit olanların sayısı mümkün permütasyon sayısı 184756'ya bölünür. Bu gerçek permütasyon dağılımına dayanan iki yönlü olasılık değeridir. İki yönlü olasılık değerlerinin küçük değerleri  $H_0$  hipotezine karşın kanıt olarak kullanılır. (Yani her hasta için A ve B tedavilerinin aynı etkiye sahip olduklarını gösterir.)

**Simülasyon:** Bu örnekte sadece 20 denekle mümkün permütasyon sayısı oldukça büyüktür. Daha önce bahsedildiği gibi yaklaşık gerçek permütasyon dağılımı için mümkün permütasyonların bir kısmı üretilebilir ve yokluk hipotezinin testi için bulunan test istatistiğine ilişkin karar bunların sonucuna bağlıdır. Basitlik için  $B=100$  permütasyon (bu değiştirilebilir.) kullanıldı. R programı ile yapılan simülasyon sonucu



**Şekil 5.4:** AIDS hastalarının iki grubu için denemelerdeki farklılıkların tahmin edilen permütasyon dağılımı

**Sonuçlar:** R programının sonuçları ve yaklaşık permütasyon dağılımı Şekil 5.4 de verilen histogramla gösterilmektedir. Bu hastalar üzerinde iki tedavinin farklı etkilere sahip olduğunu gösteren güçlü bir kanıttır.

Gözlenen test istatistiği: -64

$p$ -değeri: 0.032 (tahmini başarı olasılığı)

**Açıklama:** Bu program tekrar çalıştırılırsa, hafif farklılıklarla sonuçlar elde edileceği not edilmelidir. Bunun ışığı altında gerçek olasılık değeri için yaklaşık bir güven aralığı

$$0.032 \pm 1.96 \sqrt{\frac{(0.032)(0.968)}{1000}} \Leftrightarrow (0.021, 0.043)$$

rapor edilmesi daha iyidir. Bunun Wald tipi bir aralık olduğu hatırlanmalı ve burada “kitle” sonludur. (Mümkün 0-1 değerleri sadece 184756 adettir) Buna rağmen, CD4 sayılarına göre iki tedavi arasındaki farklılıklar gerçek görülmektedir.

**Açıklama:** Yukarıdaki olasılık dağılımının değerlendirilmesinde deneyde bireylerin seçilmesi koşullandırıldı. Yani, bu değer sabit alındı. Rasgelelik bireylere tedavilerin atanma şanslarıyla oluşturuldu. Test istatistiğinin olasılık dağılımı çıkarılarak bu gibi bir deneyde kullanılan olağan istatistiksel modelle karşılaştırmaktır. (Model bazlı sonuç çıkarım için)

$$Y_{11}, Y_{12}, \dots, Y_{1n_1} \sim N(\mu_1, \sigma_1^2)$$

$$Y_{21}, Y_{22}, \dots, Y_{2n_2} \sim N(\mu_2, \sigma_2^2)$$

$\sigma_1^2 = \sigma_2^2 = \sigma^2$  olduğunda

$$H_0 : \mu_1 - \mu_2 = 0$$

hipotezi  $\bar{Y}_{i+} = i$ . örnek ortalaması ve pooled örnek varyansı

$$S_p^2 = \frac{(n_1 - 1)^2 S_1^2 + (n_2 - 1)^2 S_2^2}{n_1 + n_2 - 2}$$

olmak üzere test istatistiği

$$t = \frac{\bar{Y}_{1+} - \bar{Y}_{2+}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

dir.  $H_0$  hipotezi altında  $t \sim t_{n_1+n_2-2}$  dağılımına sahiptir. Böylece bu referans dağılımını kullanılarak  $H_0$  hipotezine karşın güçlü kanıtın ölçüsüdür.

**Kişisel yorum:** Tedaviler hakkında sonuç çıkarım için permütasyon dağılımlarının kullanılması biraz kısıtlanır. Eninde sonunda, sonuçlar deneysel bir örnekten, büyük kitleye genişletmekle ilgilenilir. Permütasyon dağılımı ile sonuç çıkarımının kapsamı sınırlıdır, gerçekte deneydeki bu denekler hakkında sadece açıklama yapmaktır. Böylece, rasgeleliğin önemi sadece tartışılan dağılımdan bağımsız istatistiksel testlerin mantıklı bir biçimde kullanılması beceri değildir, fakat “nedensel çıkarsama” yapmaya izin verir. Bu, rasgele bir klinik deneme sonuçlarının hastalığın sonuçları üzerinde müdahale nedenlerini anlamada kullanılabilir. Bu rasgele olmayan denemelere veya sadece birleştirilerek çıkarsama yapıldığındaki epidemiyolojik çalışmalara tam terstir.

### Rasgeleliğin dezavantajları:

- Hastalar veya doktor, tedaviye karar verme şansının olduğu bir denemeye katılmaya ilgi duymayabilirler.
- Hasta-doktor ilişkisi denemeye müdahale edebilir.
- Kaynakların bir kısmı kontrol grubuna genişletilebilir. (Örneğin, tarihi kontroller)

### 5.2. Tarihi kontrolleri birleştirme:

**Senaryo:** Tedaviye cevap vermenin büyük olduğu belirli hasta kitlesi için standart tedavinin olduğu varsayalım. Yeni bir tedavinin değerlendirilmesi için  $n$  hastadan oluşan bir örneğe bu yeni tedavinin verilmesi ve tedaviye alınan cevabın karşılaştırılması için kontrol tedavisi üzerinde alınan tarihi veri neden alınmasın? Rasgele klinik deneme  $n$  hastanın alınması, rasgele  $\frac{n}{2}$  adet hastaya yeni tedavi ve diğer  $\frac{n}{2}$  adet hastaya da kontrol veya standart tedavinin atanmasını önerir. Rasgele klinik denemeye karşın tarihi kontrolleri kullanan denemenin görece yararlılığını değerlendirmenin bir temeli için hiyerarşik model kullansın.

**Tarihi kontroller:** Farklı deneysel çalışmalar değişebilir tedavi cevap oranları ile farklı çalışma kitlelerinden rasgele örnekleri içerdiği düşünülebilir. Bu bir hiyerarşik model formülasyonuna neden olur. Verilen herhangi bir kitle çalışması için A tedavisi ile B tedavisi karşılaştırıldığında cevap verme olasılığı  $\Delta$  (sabit bir parametre) kadar artar.

$$P_B = \text{B tedavisi verilen kitlede cevap oranı}$$

Bu varsayımlara göre cevap oranı, bu aynı çalışma kitlesine de A tedavisi verilmiş ise  $P_A = P_B + \Delta$  dir. Böylece  $P_B$  çalışmadan çalışmaya değişkenliğinin olasılığı için sağlanan bir rasgele değişken olarak en iyi kabul edilebilir. Özellikle,

$$E(P_B) = \theta_B \text{ ve } Var(P_B) = \sigma_\theta^2$$

olsun.  $P_A = P_B + \Delta$  olduğundan

$$E(P_A) = \theta_A = \theta_B + \Delta \text{ ve } Var(P_A) = \sigma_\theta^2$$

olarak elde edilir.

**Amaç:** Tedavi farklarını gösteren  $\Delta$  parametresini tahmin edebilmek için bir çalışmanın tasarlanması amaçtır. Bunun için iki seçenek düşünülebilir.

#### 1. Rasgele klinik deneme:

- Belirli bir çalışma kitlesi seçilir ve rasgele seçilen  $\frac{n}{2}$  hastaya A tedavisi, diğer gruba da B tedavisi atansın

- Böyle bir denemede  $\hat{P}_A$  ve  $\hat{P}_B$  rasgele seçilen çalışma kitlesinde  $\frac{n}{2}$  hastaya rasgele atanan A ve B tedavilere karşılık gelen örnek cevap oranlarını göstermek üzere  $\Delta$ 'nın bir tahmin edicisi

$$\hat{\Delta}_{RCT} = \hat{P}_A - \hat{P}_B \text{ dir.}$$

## 2. Tarihi kontroller:

- B tedavisinin kullanıldığı birçok çalışmada bir sürü geçmiş veri olsun.  $E(P_B) = \theta_B$ 'nin gerçek değeri bilinsin. Böylece, rasgele seçilen çalışma kitlesi için bütün  $n$  hastaya A tedavisi verilmesidir.
- Bu durumda  $\Delta$  için tahmin edici

$$\hat{\Delta}_{HC} = \hat{P}_A - \theta_B \text{ dir.}$$

**Soru:** Bu iki tasarım nasıl karşılaştırılır? Bunun için  $\hat{\Delta}_{RCT}$  ve  $\hat{\Delta}_{HC}$ 'nin ortalamaları ve varyanslarının bulunması sorunun cevabıdır. İlk olarak, rasgele klinik deneme düşünülür:

$$\begin{aligned} E(\hat{\Delta}_{RCT}) &= E\{E(\hat{\Delta}_{RCT} | P_A, P_B)\} \\ &= E\{E(\hat{P}_A - \hat{P}_B | P_A, P_B)\} \\ &= E(P_A - P_B) = E(\Delta) = \Delta \end{aligned}$$

Bu nedenle  $\hat{\Delta}_{RCT}$  tedavi farkı  $\Delta$ 'nın yansız bir tahmin edicisidir.  $\hat{\Delta}_{RCT}$ 'nin varyansı

$$\begin{aligned} Var(\hat{\Delta}_{RCT}) &= E\{Var(\hat{\Delta}_{RCT} | P_A, P_B)\} + Var\{E(\hat{\Delta}_{RCT} | P_A, P_B)\} \\ &= E\{Var(\hat{P}_A - \hat{P}_B | P_A, P_B)\} + Var\{E(\hat{P}_A - \hat{P}_B | P_A, P_B)\} \\ &= E\{Var(\hat{P}_A | P_A, P_B) + Var(\hat{P}_B | P_A, P_B)\} + Var\{P_A - P_B\} \\ &= E\left\{\frac{P_A(1-P_A)}{n/2} + \frac{P_B(1-P_B)}{n/2}\right\} + \underbrace{Var\{\Delta\}}_0 \\ &= E\left\{\frac{P_A(1-P_A)}{n/2} + \frac{P_B(1-P_B)}{n/2}\right\} \end{aligned}$$

Şimdi,  $\Delta$  küçük ise (yani tedaviler arasında farkın olmadığını gösteren yokluk hipotezine yakındır)

$$E\{P_A(1-P_A)\} \approx E\{P_B(1-P_B)\} \approx E\{P(1-P)\}$$

söylensin, bu durumda

$$\begin{aligned} \text{Var}(\hat{\Delta}_{RCT}) &= \frac{2}{n} E\{P_A(1-P_A) + P_B(1-P_B)\} \\ &\approx \frac{4}{n} E\{P(1-P)\} \end{aligned}$$

dir.  $E\{P(1-P)\}$  için yansız bir tahmin edici

$$\frac{1}{N} \sum_{i=1}^N \frac{n_i \hat{p}_i (1 - \hat{p}_i)}{n_i - 1}$$

ile gösterilebilir.

Daha önce verilen (Örnek 4.2) kalın bağırsağın ileri karsinomu (carcinoma) olan hastalara verilen 5-FU tedavisi tarihi kontrol olarak alınsın. En son değer 0.18 olarak bulunmuştur. Bu “kabaca bir rehber” olarak kullanılması, ileri düzeyli kalın bağırsak karsinomu olan hastalar bir çalışma kitlesi olarak alınır ve her bir tedaviye  $n$  hastadan  $\frac{n}{2}$  tanesinin atandığı bir rasgele klinik deneme oluşturulursa tahmin edilen tedavilerin farkının  $\frac{4}{n} \times 0.18 = \frac{0.72}{n}$  olan varyansa sahip olması beklenir. Tersine, eğer tarihi kontrol tasarımı kullanılırsa tedavi farkının tahmin edicisi

$$\hat{\Delta}_{HC} = \hat{P}_A - \theta_B$$

dir. Burada,  $\hat{P}_A$  = rasgele seçilen çalışma kitlesinde A tedavisi verilen  $n$  hasta içerisinde tedaviye cevap verenlerin örnek oranı ve  $\theta_B$  iyimser olarak biliniyor varsayılınsın. İlk olarak,

$$\begin{aligned} E(\hat{\Delta}_{HC}) &= E\{E(\hat{\Delta}_{HC} | P_A, P_B)\} \\ &= E\{E(\hat{P}_A - \theta_B | P_A, P_B)\} \\ &= E(P_A - \theta_B) = E(P_A) - \theta_B = \theta_A - \theta_B = \Delta \end{aligned}$$

olduğundan  $\hat{\Delta}_{HC}$  yansız bir tahmin edicidir.  $\hat{\Delta}_{HC}$  tahmin edicisinin varyansı

$$\begin{aligned} \text{Var}(\hat{\Delta}_{HC}) &= E\{\text{Var}(\hat{\Delta}_{HC} | P_A, P_B)\} + \text{Var}\{E(\hat{\Delta}_{HC} | P_A, P_B)\} \\ &= E\{\text{Var}(\hat{P}_A - \theta_B | P_A, P_B)\} + \text{Var}\{E(\hat{P}_A - \theta_B | P_A, P_B)\} \\ &= E\{\text{Var}(\hat{P}_A | P_A, P_B)\} + \text{Var}\{P_A - \theta_B\} \\ &\approx E\left\{\frac{P_A(1-P_A)}{n}\right\} + \text{Var}\{P_A\} \\ &\approx \frac{1}{n} E\{P(1-P)\} + \sigma_\theta^2 \end{aligned}$$

olacaktır. Tekrar 5-FU verisi açıklayıcı amaçlar için kullanılırsa, bu son ifade

$$\frac{0.18}{n} + 0.0435$$

olarak tahmin edilir. Daha önce  $\sigma_\theta^2$ 'nin tahmin edicisi 0.0435 olarak bulunmuştu.

**Özet:** Bu örnekte,  $\hat{\Delta}_{RCT}$  ve  $\hat{\Delta}_{HC}$  tahmin edicilerinin her ikisi  $\Delta$  için koşulsuz yansız tahmin edicilerdir. İstatiksel prensipler kullanılarak her iki tahmin edici yansız (koşulsuz) olduğundan daha iyi (veya daha etkin) tahmin edici, varyansı daha küçük olandır. Açıklayıcı örnekte, rasgele klinik deneme her ne zaman

$$\frac{0.72}{n} \leq \frac{0.18}{n} + 0.0435, \quad n \geq 13$$

daha etkindir.

**Genel olarak:** Tedavi farkı  $\Delta$  çok büyük olmadığında

$$\text{Var}(\hat{\Delta}_{RCT}) \approx \frac{4}{n} E\{P(1-P)\}$$

ve

$$\text{Var}(\hat{\Delta}_{HC}) \approx \frac{1}{n} E\{P(1-P)\} + \sigma_\theta^2$$

dır. Böylece rasgele klinik deneme

$$n \geq \frac{3}{\sigma_\theta^2} E\{P(1-P)\}$$

olduğunda bir tarihi kontrol tasarımından daha etkindir.

**Not:** Bir  $RCT$ 'in hassaslığı örnek çapı  $n$ 'in yeterince büyük seçilmesiyle keyfi olarak iyileştirilebilir. Yani tahmin edicinin varyansı arzu edildiği kadar küçültülebilir. Öte yandan bir tarihi kontrol denemesinin hassaslığı çalışmadan çalışmaya değişkenliği  $\sigma_\theta^2$  ile aşağıdan sınırlıdır.