

## McNemar Testi

- McNemar testi, iki ilgili grup arasında **ikili bir bağımlı değişken** üzerinde farklılık olup olmadığını belirlemek için kullanılır.
- Eşleştirilmiş örneklemeler t-testine benzer olduğu düşünülebilir, ancak sürekli bir bağımlı değişken yerine ikili bir değişken içindir.
- Bununla birlikte, eşleştirilmiş örneklemeler t-testinden farklı olarak, tekrarlanan ölçümlü dikotom bir değişkenin iki farklı özelliğini test etmek üzere kavramsallaştırılabilir.
- McNemar testi, ön test-son test çalışma tasarımlarını analiz etmek için kullanılmasının yanı sıra eşleştirilmiş çiftlerin ve vaka kontrol çalışmalarının analizinde de yaygın olarak kullanılmaktadır.
- İki'den fazla tekrarlanan ölçümünüz varsa Cochran's Q testini kullanabilirsiniz.

# Örnekler

- Bir dizi danışmanlık seansından (yani bir müdahaleden) önce düşük benlik saygısına sahip olan katılımcıların oranının (yüksek benlik saygısının aksine) müdahaleden sonra azalıp azalmadığını belirlemek için McNemar testini kullanabilirsiniz (yani, bağımlı değişkeniniz iki kategorisi olan "benlik saygısı düzeyi" olacaktır: "düşük" ve "yüksek").
- Alternatif olarak, bisiklet kaskı takarken kendini güvende hisseden (evet veya hayır) katılımcıların oranının bisiklet kaskı takmayanlara göre farklılık gösterip göstermediğini belirlemek için McNemar testini kullanabilirsiniz (yani bağımlı değişkeniniz "güvenlik hissi" olacaktır ve iki kategorisi vardır: "güvenli" ve "güvenli değil").

# Varsayımlar

- Varsayım #1: İki kategorili bir kategorik bağımlı değişkeniniz (yani dikotom bir değişken) ve iki ilgili grubu olan bir kategorik bağımsız değişkeniniz var.
- Varsayım #2: Bağımlı değişkeninizin iki grubu birbirini dışlamalıdır. Bu, hiçbir grubun üst üste gelemeyeceği anlamına gelir. Başka bir deyişle, bir katılımcı iki gruptan yalnızca birinde yer alabilir; aynı anda her iki grupta da yer alamaz. Örneğin, iki haftalık bir gözden geçirme döneminden (yani bir müdahaleden) önce bir sınavı geçen (sınavda başarısız olanın aksine) katılımcıların oranının müdahaleden sonra artıp artmadığını belirlemek için bir McNemar testi kullandığınızı düşünün (yani, bağımlı değişkeniniz iki kategoriye sahip olan "sınav performansı" olacaktır: "başarılı" ve "başarısız"). Bir katılımcı iki haftalık gözden geçirme döneminden önce sınava girdiğinde, sadece "başarılı" ya da "başarısız" olabilirdi. Aynı anda hem başarılı hem de başarısız olamazlar (örneğin, 100 üzerinden 60 ve üzeri not almışlarsa "başarılı", 59 ve altı not almışlarsa "başarısız" olurlar). Benzer şekilde, iki haftalık gözden geçirme döneminden sonra da katılımcı sınavı geçebilir ya da kalabilirdi.
- Varsayım #3: Vakalar (örneğin, katılımcılar) ilgili popülasyondan rastgele bir örneklemdir. Ancak uygulamada örnekleme her zaman bu şekilde gerçekleşmez.

# Örnek Veri

- Bir arařtırmacı, bir müdahalenin sigara kullanımı üzerindeki etkisini arařtırmak istemiřtir. Bu varsayımsal alıřmada, 25 sigara ien ve 25 sigara imeyen olmak üzere 50 katılımcı yer almıřtır. Tüm katılımcılar, sigaraya baėlı kanserlerden kaynaklanan ölümlerin aileler üzerindeki etkisini gösteren duygusal bir video izlemiřtir. Bu video müdahalesinden iki hafta sonra, aynı katılımcılara sigara imeye devam edip etmedikleri sorulmuřtur.
- Dolayısıyla, katılımcılar müdahaleden önce sigara ien ya da imeyen olarak kategorize edilmiř ve müdahaleden sonra sigara ien ya da imeyen olarak yeniden deėerlendirilmiřtir. Aynı katılımcılar iki kez ölçüldüėünden, eřleřtirilmiř örneklemlerimiz vardır. Ayrıca, birbirini dıřlayan iki kategoriye sahip (yani "sigara ien" ve "sigara imeyen") ikili bir baėımlı deėiřkenimiz bulunmaktadır. Sonuç olarak, verileri analiz etmek için McNemar testi uygun bir seçimdir.

		SONRA	
		Hayır	Evet
ÖNCE	Sigara İçme Durumu		
	Hayır	20	5
	Evet	16	9

**Test Statistics<sup>a</sup>**

once & sonra

N	50
Exact Sig. (2-tailed)	.027 <sup>b</sup>

a. McNemar Test

b. Binomial distribution used.

İstatistiksel anlamlılık düzeyi (yani p-değeri) .05'ten küçükse (yani  $p < .05$ ), istatistiksel olarak anlamlı bir sonucunuz vardır ve müdahaleden önce ve sonra sigara içmeyenlerin oranı istatistiksel olarak önemli ölçüde farklıdır. Alternatif olarak,  $p > .05$  ise, istatistiksel olarak anlamlı bir sonucunuz yoktur ve müdahaleden önce ve sonra sigara içmeyenlerin oranı istatistiksel olarak anlamlı bir şekilde farklı değildir (yani, sigara içmeyenlerin oranı müdahale süresince değişmez). Örneğimizde,  $p = .027$  (tam p-değeri kullanılarak), bu da sigara içmeyenlerin oranının müdahaleden sonra öncesine kıyasla istatistiksel olarak önemli ölçüde farklı olduğu anlamına gelir. Başka bir deyişle, müdahalenin ardından sigara içmeyenlerin oranındaki değişim istatistiksel olarak anlamlıdır.

Raporlama: Elli katılımcı, sigaranın zararları konusunda uyarmak üzere tasarlanmış bir müdahalede yer almak üzere işe alınmıştır. Tam McNemar testi, müdahale öncesi ve sonrası sigara içmeyenlerin oranında istatistiksel olarak anlamlı bir fark olduğunu belirlemiştir,  $p = .027$ .

# Cochran's Q test,

- Cochran's Q testi, üç veya daha fazla ilgili grup arasında ikili bir bağımlı değişken üzerinde farklılık olup olmadığını belirlemek için kullanılır. Tek yönlü tekrarlanan ölçümler ANOVA'sına benzer, ancak sürekli bir bağımlı değişken yerine ikili bir değişken için veya McNemar testinin bir uzantısı olarak düşünülebilir. Cochran'ın Q testi, boylamsal çalışma tasarımlarını analiz etmek için yaygın olarak kullanılmasının yanı sıra, birden fazla farklı denemeden (diğer bir deyişle tedaviler/koşullar) geçen katılımcıları analiz ederken de kullanılır.
- Örneğin, Cochran'ın Q testini, bir dizi üç danışmanlık seansından sonra düşük benlik saygısına sahip katılımcıların oranının (yüksek benlik saygısının aksine) azalıp azalmadığını belirlemek için kullanabilirsiniz (yani, bağımlı değişkeniniz iki kategorisi olan "benlik saygısı düzeyi" olacaktır: "düşük" ve "yüksek", dört zaman noktasında ölçülür: "ilk danışmanlık seansından önce", "ikinci danışmanlık seansından sonra", "üçüncü danışmanlık seansından sonra" ve "son danışmanlık seansından sonra"). Alternatif olarak, kendini güvende hisseden (evet veya hayır) kadın katılımcıların oranının göz yaşartıcı sprey, alarm veya hiçbir şey taşımadıklarında farklılık gösterip göstermediğini belirlemek için Cochran'ın Q testini kullanabilirsiniz (yani, bağımlı değişken iki kategorisi olan "güvenlik hissi" olacaktır: "güvenli" veya "güvenli değil", üç uygulama/koşulda ölçülmüştür: "topuz", "alarm" ve "hiçbir şey").

# Varsayımları

- Varsayım #1: Birbirini dışlayan iki gruba sahip tek bir bağımlı değişkeniniz var (yani, değişken ikilidir).
- Varsayım #2: Üç veya daha fazla kategorik, ilişkili gruptan oluşan bir bağımsız değişkeniniz var (yani, bağımsız değişkeniniz sıralı veya nominal bir değişkendir).
- Varsayım #3: Vakalar (örneğin, katılımcılar) ilgili popülasyondan rastgele bir örneklemdir. Ancak uygulamada örnekleme her zaman bu şekilde gerçekleşmez.
- Varsayım #4: Örneklem büyüklüğünüz, Cochran'ın Q testi tarafından üretilen asimptotik p-değerini yorumlayabilmeniz için yeterince büyük olmalı (yani, p-değeri, üç veya daha fazla ilgili grubunuz arasındaki oranlarda istatistiksel olarak anlamlı bir fark olup olmadığını belirlemek için başvurduğunuz değerdir). Örneklem büyüklüğünüz yetersizse, asimptotik p-değeri doğru olmayabilir, ancak Cochran's Q testinin daha doğru bir kesin p-değeri üretecek "kesin" bir versiyonunu çalıştırabilirsiniz.

# Örnek

- Bir okul öğretmeni, öğrencilerin çalışmak için daha fazla zamanı olduğunda geçme oranlarının artıp artmadığını incelemek istemiştir. Bu varsayımsal çalışmada, 60 öğrenci yer almak üzere işe alınmıştır. Tüm öğrencilere önce mevcut bilgilerini test etmek için bir "sürpriz sınav" yapıldı. Ardından iki hafta sonra bir "deneme sınavı" yapılmış ve iki hafta sonra da bir "final sınavına" girmişlerdir. Öğrencilerin sınavlardaki performansları "başarılı" ya da "başarısız" olarak değerlendirilmiştir.
- Aynı öğrenciler üç kez ölçüldüğü için (yani "sürpriz", "deneme" ve "final" sınavı sırasında), tekrarlanan ölçümlerimiz var. Ayrıca, birbirini dışlayan iki kategoriye sahip (yani, "Başarısız" ve "Geçti") ikili bir bağımlı değişkenimiz vardır. Sonuç olarak, Cochran'ın Q testi verileri analiz etmek için uygun bir seçimdir.

# Örnek Rapor

- Altmış katılımcı, öğrencilerin çalışmak için daha fazla zamanı olduğu için üç sınavda (sürpriz, deneme ve final sınavı) geçme oranlarının değişip değişmediğini incelemek üzere çalışmaya katılmak üzere işe alınmıştır. Cochran's Q testi, zaman içinde sınavlarını geçen öğrencilerin oranında istatistiksel olarak anlamlı bir fark olduğunu belirlemiştir,  $\chi^2(2) = 49.515$ ,  $p < .0005$ .

# Wilcoxon İşaretli Sıralar Testi

- Wilcoxon işaretli sıralar testi, bağımlı t-testine eşdeğer parametrik olmayan bir testtir. Wilcoxon işaretli sıralar testi verilerde normallik varsaymadığından, bu varsayımın ihlal edildiği ve bağımlı t-testinin kullanımının uygun olmadığı durumlarda kullanılabilir. Aynı katılımcılardan gelen iki puan kümesini karşılaştırmak için kullanılır. Bu, bir zaman noktasından diğerine puanlardaki herhangi bir değişikliği araştırmak istediğimizde veya bireyler birden fazla koşula tabi tutulduğunda ortaya çıkabilir.

# Varsayımları

- Varsayım #1: Bağımlı değişkeniniz sıralı veya sürekli düzeyde ölçülmelidir.
- Varsayım #2: Bağımsız değişkeniniz iki kategorik, "ilişkili gruplar" veya "eşleştirilmiş çiftlerden" oluşmalıdır. "ilişkili gruplar" her iki grupta da aynı deneklerin bulunduğunu gösterir. Her grupta aynı deneklerin bulunmasının mümkün olmasının nedeni, her bir deneğin aynı bağımlı değişken üzerinde iki kez ölçülmüş olmasıdır.

- Varsayım #3: İlgili iki grup arasındaki farkların dağılımının (yani, bağımsız değişkenin her iki grubunun puanları arasındaki farkların dağılımı; örneğin, "mavi aydınlatmalı" bir oda ile "kırmızı aydınlatmalı" bir odadaki tepki süresi) simetrik bir şekle sahip olması gerekir. Farklılıkların dağılımı simetrik olarak şekilleniyorsa, çalışmanızı Wilcoxon işaretli sıralar testini kullanarak analiz edebilirsiniz. Ancak, SPSS Statistics kullanarak kendi verilerinizi analiz ederken bu varsayım ihlal edilirse (yani karşılanmazsa) şaşırmayın. Bununla birlikte, verileriniz bu varsayımı karşılamasa bile, genellikle bunun üstesinden gelmek için bir çözüm vardır; örneğin, farklılıkların simetrik şekilli bir dağılımını elde etmek için verilerinizi dönüştürmek (tercih edilen bir seçenek değildir) veya Wilcoxon işaretli sıralama testi yerine bir işaret testi yapmak gibi.

# Örnek Rapor

- Wilcoxon işaretli sıralar testi, 4 haftalık, haftada iki kez uygulanan akupunktur tedavisinin mevcut bel ağrısı olan bireylerde bel ağrısında istatistiksel olarak anlamlı bir değişiklik yaratmadığını göstermiştir ( $Z = -1.807$ ,  $p = 0.071$ ). Nitekim, medyan Ağrı Skoru değerlendirmesi hem tedavi öncesinde hem de sonrasında 5.0 idi.

# Friedman Testi

- Friedman testi, tekrarlanan ölçümlerle tek yönlü ANOVA'nın parametrik olmayan alternatifidir. Ölçülen bağımlı değişken sıralı olduğunda gruplar arasındaki farklılıkları test etmek için kullanılır. Ayrıca, tekrarlanan ölçümlerle tek yönlü ANOVA'yı çalıştırmak için gerekli varsayımları ihlal eden sürekli veriler için de kullanılabilir (örneğin, normallikten belirgin sapmaları olan veriler).

# Varsayımları

- Varsayım #1: Üç veya daha fazla farklı durumda ölçülen bir grup.
- Varsayım #2: Grup, popülasyondan rastgele bir örneklemdir.
- Varsayım #3: Bağımlı değişkeniniz sıralı veya sürekli düzeyde ölçülmelidir.
- Varsayım #4: Örneklerin normal dağılıma sahip olması GEREKMEZ.

# Post Hoc Testleri

- Farklılıkların gerçekte nerede oluştuğunu incelemek için, ilgili grupların farklı kombinasyonları üzerinde ayrı Wilcoxon işaretli sıralama testleri yapmanız gerekir. Wilcoxon testlerinden elde ettiğiniz sonuçlar üzerinde Bonferroni düzeltmesi kullanmanız gerekir çünkü çoklu karşılaştırmalar yapıyorsanız, bu da bir sonucu anlamlı ilan etmemeniz gerekirken anlamlı ilan etme olasılığınızı artırır (Tip I hata). Neyse ki Bonferroni düzeltmesini hesaplamak çok kolaydır; sadece başlangıçta kullandığınız anlamlılık düzeyini (bu durumda 0,05) alın ve çalıştırdığınız test sayısına bölün. Örneğin 3 karşılaştırma var ise,  $0,05/3 = 0,017$ 'lik yeni bir anlamlılık düzeyine sahibiz. Bu, p değerinin 0,017'den büyük olması durumunda istatistiksel olarak anlamlı bir sonuca sahip olmadığımız anlamına gelir.