



HACETTEPE
ÜNİVERSİTESİ
İSTATİSTİK BÖLÜMÜ

İST156 İSTATİSTİĞE GİRİŞ II

DERS 8-ÖRNEKLEM BÜYÜKLÜĞÜ TAHMİNİ

Ders sorumluları: Prof.Dr.Serpil AKTAŞ ALTUNAY (01 Şubesi)
Doç.Dr. Ayten YİĞİTER (02 Şubesi)

ÖRNEKLEM BÜYÜKLÜĞÜ TAHMİNİ (SAMPLE SIZE ESTIMATION)



Bir kitleden birimlerin (gözlemlerin) örnekleme nasıl seçileceğine (örnekleme yöntemine) kararı verme süreci kadar ne kadar sayıda birimin (gözlemin) yani **örneklem büyüklüğün** belirlenmesi önemli bir konudur.

NEDEN ÖRNEKLEM SEÇİLİR?

Bir araştırma kapsamına konu olan kitlenin tüm birimlerine ulaşılması, zaman ve maliyet vb. nedenlerden dolayı çoğu zaman mümkün değildir. Bu nedenle, kitleyi temsil edecek örneklem seçilir ve seçilen örneklem bilgisi kullanılarak kitlenin özellikleri (parametreleri) tahmin edilir. Örneklem seçme sürecinde iki temel konu: Örneklem büyüklüğü ve örnekleme yöntemidir.

- ❖ Bilgiyi örneklem yerine, kitleden elde etmekten daha ucuzdur.
- ❖ Bilgi örnekleme ile daha çabuk toplanır.
- ❖ Örnekleme ile daha ayrıntılı veriler elde edilebilir.
- ❖ Bazı durumlarda ise örnekleme dışında bilgiye ulaşmak mümkün değildir.

ÖRNEKLEME YÖNTEMLERİ

Olasılığa bağlı

Basit rastgele
örnekleme

Tabakalı
örnekleme

Sistematiik
örnekleme

Küme
örnekleme

Olasılığa bağı olmayan

Keyfi
Örnekleme

Kota
Örnekleme

Kartopu
örnekleme

Olasılığa Bağlı Örneklem Yöntemleri

BASİT RASTGELE ÖRNEKLEME

Bu yöntemde, kitledeki her birimin örneklem seçilme şansı eşittir. Birimlerin seçilmesi sürecinde rastgele sayılar tablosu ya da rastgele sayılar üretici algoritmalar kullanılır. Kitle homojen yapıda olduğu durumda kullanışlıdır.

TABAKALI RASTGELE ÖRNEKLEME

Tabakalı rastgele örneklemde kitledeki birimlerin, incelenecek özelliği etkileyen diğer bir özelliğine göre birbirinden farklı tabakalara ayrılır. Birimlerin seçimi her bir tabakadan basit rastgele örneklem yöntemi kullanılarak yapılır. Her bir tabaka kendi içinde homojen, tabakalar arası heterojendir.

SİSTEMATİK ÖRNEKLEME

Seçim işlemleri kolay olduğundan, özellikle kitle çok geniş olduğunda tercih edilir. Birimlerin seçiminde kitle büyüklüğü (N), örneklem büyüklüğüne (n) bölünerek kaç birimde bir birimin örneklem alınacağı belirlenir.

KÜME ÖRNEKLEMESİ

Kitledeki birimlere doğrudan ulaşmanın kolay olmadığı durumlarda bu yöntem başvurulur. Örneğin, lise öğrencileri ile yapılacak bir araştırmada okullar kümeleri oluşturur. Okullar rastgele seçilir. Tabakalı örneklem yönteminden farklı olarak kümeler aralarında homojen, küme için ise heterojen olması istenir.

Olasılığa Bağlı Olmayan Örneklem Yöntemleri

KEYFİ ÖRNEKLEME

Kitlenin dar bir alanı kapsamaması ve bilgi toplama maliyeti bakımından birimden birime farklılık yoksa ve çeşitli özellikte bulunan birimler yığının her tarafına dağılmış keyfi örneklemeyi tercih etmek uygun olabilir. Örneğin sosyal medya araçları üzerinden yapılan anket çalışmaları.

KOTA ÖRNEKLEMESİ

Kota örnekleme maliyetin düşük olması, idaresinin kolay olması ve kısa sürede tamamlanması nedeniyle tercih edilmektedir. Örneğin bir kitlede %40 erkek, %60 kadın ve örnek büyüklüğü de 300 olsun. Buna göre örnekteki kadın ve erkek sayılarının kitledeki oranları göz önüne alınarak 120 erkek ve 180 kadın olması gerekir. Toplam 10 anketör için her biri 30 anket ve 1 anketör 12 erkek ve 18 kadına anket uygular.

KARTOPU ÖRNEKLEMESİ

Kitlenin sınırlarının ve kitledeki birimlerin kesin olarak belirlenemediği durumlarda kullanılan bir yöntemdir. Kartopu örneklemesinde öncelikle yığına ait birimlerden birisi ile temas kurulur. Temas kurulan birimin yardımıyla ikinci birime, ikinci birimin yardımıyla üçüncü birime gidilir. Bu şekilde, sanki bir kartopunun büyümesi gibi örneklem büyüklüğü genişler.

ÖRNEKLEME HATASI (Sampling Error)

Örnekleme hatası, bir veri toplama işleminde, kitlenin (popülasyonun) tümü yerine kitleden örneklem almanın bir sonucu olarak ortaya çıkan hatadır. Başka bir deyişle, örnekleme alınan ya da alınmayan birimlerin şansa bağlı olarak seçilmesi sonucunda ortaya çıkan hata olarak da adlandırılır. Tahminler örneklemden örnekleme farklılık gösterir ve bu tahminler arasındaki değişkenliğin ölçüsü olan «standart hata» **örnekleme hatası** olarak adlandırılır.

Örnekleme hatası, örneklemdeki birim sayısını artırmak ya da birimlerin örnekleme seçme sürecini rastgeleleştirme ile azaltılabilir.

Örneklem büyüklüğü, kabul edilebilir hata ile belirlenen anlamlılık (ya da güvenilirlik) düzeyinde araştırmanın yapılabileceği en az denek/gözlem sayısını belirleme işidir. Örneklem büyüklüğünün belirlenmesinde araştırmacı, kabullenebilecek I. Tip hata (α) ve II. Tip hata (β) miktarının ne oranda alınması gerektiğine karar vermesi gerekir.

- **I. Tip Hata** (α): Yokluk hipotezinin (H_0) doğru iken reddedilmesi olasılığıdır. $(1 - \alpha)$ ise testin **güvenirlilik** düzeyidir.
- **Güç** ($1 - \beta$): H_0 hipotezi yanlış iken, H_0 hipotezinin reddedilmesi olasılığıdır. Testin duyarlılığı (sensitivity) olarak da adlandırılır. β ise **II. Tip Hata**'dır.

KİTLE ORTALAMASININ TAHMİNİ İÇİN

Kitle varyansı (σ^2) biliniyor ise

Kitledeki birim sayısı (N)
biliniyor

$$n \geq \frac{N \times \frac{z_{\alpha}^2}{2} \times \sigma^2}{d^2(N - 1) + \frac{z_{\alpha}^2}{2} \times \sigma^2}$$

Kitledeki birim sayısı (N)
bilinmiyor

$$n \geq \frac{\frac{z_{\alpha}^2}{2} \times \sigma^2}{d^2}$$

d : tahminin gerçek değer (μ)'den sapma miktarı (hoşgörü miktarı)

KİTLE ORTALAMASININ TAHMİNİ İÇİN

Kitle varyansı (σ^2) bilinmiyor ise

Kitledeki birim sayısı (N)
biliniyor

Kitledeki birim sayısı (N)
bilinmiyor

$$n \geq \frac{N \times t_{(\frac{\alpha}{2}, sd)}^2 \times S^2}{d^2(N - 1) + t_{(\frac{\alpha}{2}, sd)}^2 \times S^2}$$

$$n \geq \frac{t_{(\frac{\alpha}{2}, sd)}^2 \times S^2}{d^2}$$

t istatistiğinin değeri yaklaşık 1,96 ~2 alınır)

Örnek 1: Bir araştırmacı bir bölgede yeni doğan bebeklerin ortalama doğum ağırlıklarını tahmin etmek istiyor. Eğer %95 güven düzeyinde bebek ağırlıklarını gerçek ortalamadan ± 250 gr sapma ile tahmin etmek isterse kaç bebek seçmelidir. (Daha önceki çalışmalarda kitle standart sapması'nın 700 gr olduğu bilgisi mevcuttur.)

$1 - \alpha = 0,95$ ve $\alpha = 0,05$ olur. $z_{\frac{\alpha}{2}} = z_{0,025} = 1.96$

$\sigma^2 = 700^2$ verilmiş $\longrightarrow n = \frac{z_{\frac{\alpha}{2}}^2 \times \sigma^2}{d^2} = \frac{1,96^2 \times 700^2}{250^2} = 30$


N=600 bebek olduğu varsayılırsa,


$$n = \frac{N \times z_{\frac{\alpha}{2}}^2 \times \sigma^2}{d^2(N - 1) + z_{\frac{\alpha}{2}}^2 \times \sigma^2} = \frac{600 \times 1,96^2 \times 700^2}{250^2 \times (600 - 1) + 1,96^2 \times 700^2} = 200$$

bebeğin seçilmesi gerekir.

Örnek 2: Sosyal yardım planlaması için bir bölgede yaşayan hanelerin ortalama geliri tahmin edilmek isteniyor. Eğer %95 güvenle düzeyinde hanehalkının gelirinin gerçek ortalamadan $\pm 100\text{TL}$ ve $\pm 50\text{TL}$ sapma ile tahmin etmek isterse kaç hane seçmelidir. ($\sigma=1000\text{TL}$)

$1 - \alpha = 0,95$ ve $\alpha = 0,05$ olur. $z_{\frac{\alpha}{2}} = z_{0,025} = 1,96$

$d=100$  $n = \frac{z_{\alpha}^2 \times \sigma^2}{d^2} = \frac{1,96^2 \times 1000^2}{100^2} = 384,16 \cong 384$

$d=50$  $n = \frac{z_{\alpha}^2 \times \sigma^2}{d^2} = \frac{1,96^2 \times 1000^2}{50^2} = 1536,64 \cong 1537$

hanenin seçilmesi gerekir.

Örnek 3: Bir spor kulübündeki sporcuların günlük ortalama protein tüketimleri ± 5 gr. sapma ile %95 güven düzeyinde tahmin edilmek isteniyor. Kulüpte kaç sporcu olduğu bilinmiyorsa ve pilot bir çalışma sonucunda spor yapanların günlük protein alımlarının standart sapmasının 20 gr. olduğu hesaplanmışsa, en az kaç sporcu seçilmelidir?

$1 - \alpha = 0,95$ ve $\alpha = 0,05$, $d=5$, $t_{\text{Tablo}} = 2$ ve $S=20$ 'dir
(σ^2 bilinmiyor) .

$$n = \frac{t_{\alpha}^2 \times S^2}{d^2} = \frac{2^2 \times 20^2}{5^2} = 64$$

Günlük ortalama protein tüketimini tahmin etmek için en az 64 sporcu seçilmelidir.



Örneklem büyüklüğü formüllerinde t ya da Z kullanılmaktadır. Ancak çözümlerde her ikisi de yaklaşık 2 değerini alır. σ^2 'nin bilinmediği durumda da onun tahmin edicisi S^2 kullanıldığından formüllerde aynı sonuca ulaşılır.

KİTLE ORANININ TAHMİNİ İÇİN

Kitledeki birim sayısı (N)
biliniyor



$$n = \frac{N \times \frac{z_{\alpha}^2}{2} \times P(1 - P)}{d^2(N - 1) + \frac{z_{\alpha}^2}{2} \times P(1 - P)}$$

Kitledeki birim sayısı (N)
bilinmiyor



$$n = \frac{\frac{z_{\alpha}^2}{2} \times P(1 - P)}{d^2}$$

d : tahminin gerçek değer (P)'den sapma miktarı (hoşgörü miktarı)

Örnek 4: Bir bölgede 16-25 yaş aralığındaki gençlerde obezite oranı tahmin edilmek isteniyor. %95 güvenle gençlerdeki obezite oranının tahmini %4 sapma ile tahmin edebilmek için örneklem büyüklüğü kaç olmalıdır?

Kitledeki obezite oranı P bilinmiyor ise, $P=1-P=0,5$ alınır.

$1 - \alpha = 0,95$ ve $\alpha = 0,05$ olur. $z_{\frac{\alpha}{2}} = z_{0,025} = 1.96$

$$n = \frac{z_{\frac{\alpha}{2}}^2 \times P(1 - P)}{d^2} = \frac{1,96^2 \times 0,5 \times 0,5}{0,04^2} = 600,25 \cong 600$$

$N=5000$ genç olduğu varsayılırsa,

$$n = \frac{N \times z_{\frac{\alpha}{2}}^2 \times P(1 - P)}{d^2(N - 1) + z_{\frac{\alpha}{2}}^2 \times P(1 - P)} = \frac{5000 \times 1,96^2 \times 0,5 \times 0,5}{0,04^2 \times (5000 - 1) + 1,96^2 \times 0,5 \times 0,5} = 536$$

gencin seçilmesi gerekir.

Örnek 5: Bir seçim kampanyası araştırma şirketi ile ortak yürütülmektedir. B partisinin oy oranını %95 güvenle ve gerçek oy oranından %3 farkla (hoşgörü ile) tahmin etmek istiyor. Araştırma şirketi 4milyonluk seçmenin bulunduğu bir şehir için kaç birimlik örneklem seçmelidir?

Kitledeki oy oranı P bilinmiyor ise, $P=1-P=0,5$ alınır.

$1 - \alpha = 0,95$ ve $\alpha = 0,05$ olur. $z_{\frac{\alpha}{2}} = z_{0,025} = 1,96$

$N=4$ milyon seçmen,

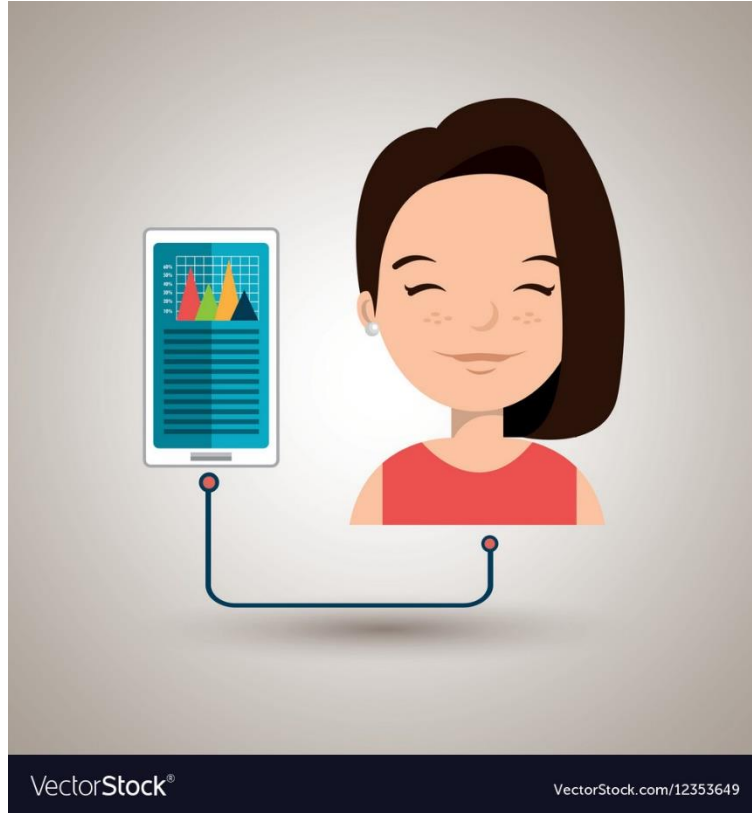
$$n = \frac{N \times z_{\frac{\alpha}{2}}^2 \times P(1 - P)}{d^2(N - 1) + z_{\frac{\alpha}{2}}^2 \times P(1 - P)} = \frac{4000000 \times 1,96^2 \times 0,5 \times 0,5}{0,03^2 \times (4000000 - 1) + 1,96^2 \times 0,5 \times 0,5}$$
$$= 1066,827 \cong 1067 \text{ seçmen}$$

Önceki seçimlerde oy oranı %60 olsun, buna göre örneklem büyüklüğü,

$$n = \frac{4000000 \times 1,96^2 \times 0,6 \times 0,4}{0,03^2 \times (4000000 - 1) + 1,96^2 \times 0,6 \times 0,4} = 1024,165$$
$$\cong 1024 \text{ seçmen}$$

Örneklem Büyüklüğü Hesaplama İçin Kullanılan Programlar

- ❖ Stata,
- ❖ Minitab,
- ❖ SPSS Sample Power,
- ❖ R



Bir sonraki derste «korelasyon konusu»
incelenecek.



KAYNAKLAR

- 1) “İstatistiksel Yöntemlere Giriş”, H.Demirhan, C.Hamurkaroğlu H.Ü.Yayınları, 2011.
- 2) “Discovering Statistics Using SPSS for Windows : Advanced Techniques for the Beginner”, Andy Field, Ref No: HA32.F54 2000.
- 3) “SPSS for Windows : An Introduction to Use and Interpretation in Research”, George A. Morgan, Orlando V. Griego, Gene W. Gloeckner. Ref No: HA 32M667 2001.
- 4) “Modern Elementary Statistics”, John.E.Freund, Prentice Hall, 2004.
- 5) “Temel İstatistik Yöntemler”, Serpil Cula, Zehra Muluk, Başkent Üniversitesi yayınları,2006.
- 6) «Genç Akademisyenler İçin Bilimsel Araştırma Teknikleri ve Yayın Etiği, Bölüm 4: Örneklem Seçimi ve Güç Analizi», Hande Konşuk Ünlü, Serpil Aktaş Altunay, Mustafa Altındış, Nobel Yayınevi, 2022.
<https://opentextbc.ca/introbusinessstatopenstax/chapter/test-of-a-single-variance/>
<https://creativemaths.net/blog/sampling-and-non-sampling-error/>
<https://slideplayer.biz.tr/slide/14065947/>