

ÖRN: Bir TV firması yayına koyduğu bir programın izlenme oranını ve sayısını tahmin etmek istiyor. Bu bağlamda bölge sosyo-ekonomik yapısına göre 3 tabakaya ayrılıyor ve 2000 kişiyle görüşülüyor. İstenen tahmini ve güven sınırlarını bulunuz.

Sosyo-Ekonomik Yapı	N_h	n_h	İzlenme oranı	W_h
Fakir	110000	1100	0,22	0,550
Orta	75000	650	0,14	0,375
Zengin	15000	250	0,07	0,075

ÇÖZÜM: Tabakalı rastgel örnekleemeye göre izlenme oranını tahmini

$$p_{tb} = \sum_{h=1}^L W_h p_h$$

$$= 0,550 \cdot 0,22 + 0,375 \cdot 0,14 + 0,075 \cdot 0,07$$

$$= 0,1825$$

Bu programı izleyenlerin sayısının tahmini de

$$\hat{A}_{tb} = N p_{tb}$$

$$= 200000 \times 0,1825$$

$$= 36500$$

kişidir. p_{tb} 'nin varyansı:

$$\hat{V}(p_{tb}) = \sum_{h=1}^L W_h^2 \frac{p_h q_h}{n_h - 1}$$

$$= 0,55^2 \frac{0,22 \times 0,78}{1099} + 0,375^2 \frac{0,14 \times 0,86}{649} + 0,075^2 \frac{0,07 \times 0,93}{249}$$

$$= 0,0000743$$

$$sh(p_{tb}) = 0,0086$$

$$P\left(p_{tb} - \left[t_{\alpha/2} sh(p_{tb}) + \frac{1}{2n}\right] \leq P \leq p_{tb} + \left[t_{\alpha/2} sh(p_{tb}) + \frac{1}{2n}\right]\right) = 1 - \alpha$$

$$P(0,1825 - 0,0177 \leq P \leq 0,1825 + 0,0177) = 0,95$$

$$P(0,1648 \leq P \leq 0,2002) = 0,95$$

Bu programın bu gölgede izlenme oranı %95 güvenle %16,48 ile %20,02 arasında değişmektedir.

$$\hat{V}(\hat{A}_{tb}) = \hat{V}(Np_{tb}) = 2180000^2 \hat{V}(p_{tb})$$

$$sh(\hat{A}_{tb}) = 1725,0611$$

$$P(33049,8 \leq A \leq 39950,122) = 0,95$$

Bu bölgede bu diziyi seyredenlerin sayısı 0,95 günde 33050 ile 39950 arasında değişiklik göstermektedir.

7. ÖRNEKLEM GENİŞLİĞİNİN TABAKALARA DAĞITIMI

7.1 Sürekli Veriler için Dağıtım

Tabakalı örneklemede tabakalardan seçilecek alt örneklem genişliklerinin (n_h 'ların) belirlenmesi önem arz eder. Bu bağlamda maliyet ve varyansın en küçük yapılar örneklem genişliklerinin bulunması gerekir. Bununla birlikte maliyet fonksiyonu tabakalı örneklemede

$$M = M_0 + \sum_{h=1}^L m_h n_h$$

şeklinindedir. Bu formülde M_0 genel giderleri göstermekte m_h ise h -ıncı tabakada birim başına yapılacak masraftır.

Tabakalı örneklemede örneklem genişliğini tabakalara dağıtımları aşağıdadır.

1)

$$n_h = n \frac{N_h S_h / \sqrt{m_h}}{\sum_{h=1}^L N_h S_h / \sqrt{m_h}}, \quad h = 1, 2, \dots, L$$

veya

$$n_h = n \frac{W_h S_h / \sqrt{m_h}}{\sum_{h=1}^L W_h S_h / \sqrt{m_h}}, \quad h = 1, 2, \dots, L$$

Bu dağıtıma optimal dağıtım ya da en iyi dağıtım denir.

2) Optimal dağıtımda $m_1 = m_2 = \dots = m_L = m$ ise yani her bir tabakada maliyet aynı ise

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}, \quad h = 1, 2, \dots, L$$

veya

$$n_h = n \frac{W_h S_h}{\sum_{h=1}^L W_h S_h}, \quad h = 1, 2, \dots, L$$

olur. Bu dağıtıma da Neyman dağıtımını denir.

3) Neyman dağıtımında her bir tabakada varyans aynı ise $S_1^2 = S_2^2 = \dots = S_L^2 = S^2$ ise

$$n_h = n \frac{N_h}{\sum_{h=1}^L N_h}, \quad h = 1, 2, \dots, L$$

veya

$$n_h = n W_h, \quad h = 1, 2, \dots, L$$

4) Eşit dağıtım

$$n_h = \frac{n}{L}$$

olmakla birlikte tercih edilen bir dağıtım değildir.

Süreklili Verilerde Optimal Dağıtımda

$$n_h = n \frac{W_h S_h \sqrt{m_h}}{\sum_{h=1}^L W_h S_h \sqrt{m_h}}$$

$$V(\bar{y}_{tb}) = \sum_{h=1}^L W_h^2 (1 - f_h) \frac{S_h^2}{n_h}$$

$$= \sum_{h=1}^L W_h^2 \frac{S_h^2}{n_h} - \sum_{h=1}^L W_h \frac{S_h^2}{n_h}$$

$$V_{opt}(\bar{y}_{tb}) = \frac{1}{n} \left(\sum_{h=1}^L W_h S_h \sqrt{m_h} \right) \left(\sum_{h=1}^L \frac{W_h S_h}{\sqrt{m_h}} \right) - \frac{1}{N} \sum_{h=1}^L W_h S_h^2$$

Eğer $V(\bar{y}_{tb})$ önceden biliniyorsa

$$n = \frac{\left(\sum_{h=1}^L \frac{W_h S_h}{\sqrt{m_h}} \right) \left(\sum_{h=1}^L \frac{W_h S_h}{m_h} \right)}{V_{opt}(\bar{y}_{tbk}) + \frac{1}{N} \sum_{h=1}^L W_h S_h^2}$$