

## Orneklem Buyuklugu

Bir arastirmaci  $n$  bagimsiz deney baz alinarak elde edilen binom parametresi  $p$ 'yi tahmin etmek istiyor, fakat kac tane  $n$  kullanmasi gerektigini bilmiyor. Tabii ki daha buyuk  $n$  degerleri daha iyi sonuclar verecektir, her deneyin bir masrafi vardir. Bu iki gereklilik nasil birbiri ile uzlastirilir?

Yeterli olacak en az kesinligi, duyarlilik (precision) bulmak icin  $Z$  transformasyonu kullanilabilir belki. Diyelim ki  $p$  icin maksimum olurluk tahmini olan  $X/n$ 'in en azindan  $100(1 - \alpha)\%$  olasilikta  $p$ 'nin  $d$  kadar yakininda olmasini istiyoruz. O zaman alttaki denklemi tatmin eden en ufak  $n$ 'i buldugumuz anda problemimizi cozduk demektir,

$$P\left(-d \leq \frac{X}{n} - p \leq d\right) = 1 - \alpha \quad (1)$$

Tahmin edici  $X/n$ 'nin kendisi de bir rasgele degiskendir. Bu degisken normal olarak dagilmistir, cunku  $X$  Binom olarak dagilmis ise, bu dagilim ayri Bernoulli dagilimlerinin toplamina esittir. Toplamin aritmetik ortalamasi ise Merkezi Limit Kanunu'na gore normal'e yaklasir. O zaman, standardize etmek icin  $X/n$ 'den beklentiyi cikartip standart sapmaya bolmek gerekir. Beklenti zaten cikartilmis durumda (sansa bak!), beklentinin ne oldugunu kontrol edelim tabii, ezbere yapmayalim bu isi, eger her Bernoulli'yi  $X_i$  olarak temsil edersek,

$$X = X_1 + \dots + X_n$$

$$X/n = 1/n(X_1 + \dots + X_n)$$

$$E[X/n] = E[1/n(X_1 + \dots + X_n)]$$

$$= 1/nE[(X_1 + \dots + X_n)]$$

$$= n(1/n)p = p$$

Varyans icin

$$\text{Var}(X/n) = \frac{1}{n^2}\text{Var}(X) = \frac{1}{n^2}np(1-p) = \frac{1}{n}p(1-p)$$

Binom dagilimlar icin  $\text{Var}(X) = np(1-p)$  oldugunu biliyoruz. Standart sapma ustteki ifadenin karekoku, yani

$$\text{Std}(X/n) = \sqrt{p(1-p)/n}$$

Simdi standardize edelim,

$$P\left(\frac{-d}{\sqrt{p(1-p)/n}} \leq \frac{\frac{X}{n} - p}{\sqrt{p(1-p)/n}} \leq \frac{d}{\sqrt{p(1-p)/n}}\right) = 1 - \alpha$$

$$P\left(\frac{-d}{\sqrt{p(1-p)/n}} \leq Z \frac{d}{\sqrt{p(1-p)/n}}\right) = 1 - \alpha$$

Daha onceki z-skoru iceren esitsizlikleri hatirlarsak, ustteki ifade

$$\frac{d}{\sqrt{p(1-p)/n}} = z_{\alpha/2}$$

O zaman

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

Fakat bu bir nihai sonuc olamaz, cunku  $n$ ,  $p$ 'nin bir fonksiyonun haline geldi ve  $p$  bilinmeyen bir deger. Fakat biliyoruz ki  $0 \leq p \leq 1$ , ve  $p(1-p) \leq \frac{1}{4}$ .

Bunu kontrol edelim,  $p(1-p)$  hangi  $p$ 'de maksimize olur?  $p$ 'ye gore turev aliriz, sifira esitleriz,  $(p-p^2)' = 1-2p = 0$ ,  $p = 1/2$ . Ve hesabi yaparsak,  $1/2(1-1/2) = 1/4$ . Demek ki  $p(1-p)$  degeri  $1/4$ 'ten daha buyuk olamaz. Buna gore, ustteki formule  $p(1-p)$  yerine onun olabilecegi en buyuk degeri koyarsak,

$$\frac{z_{\alpha/2}^2 1/4}{d^2} = n$$

$$n = \frac{z_{\alpha/2}^2}{4d^2}$$

Ornek

Buyuk bir sehirde cocuklarin kacta kacinin asisini almis olup olmadigini anlamak icin bir anket gerceklestirilecek. Anketi duzenleyenler orneklem orani olan  $X/n$ 'in en az 98% oranda gercek oran  $p$ 'nin 0.05 yakininda olmasini istiyorlar. Orneklem ne kadar buyuk olmalidir?

Burada  $100(1-\alpha) = 98$ , o zaman  $\alpha = 0.02$ , demek ki  $z_{\alpha/2} = z_{0.02/2} = z_{0.01}$  degerine ihtiyacimiz var. Python ile

```
from scipy.stats.distributions import norm
print norm.ppf(0.99)
```

2.32634787404

Tüm hesap için

$$n = \frac{(2.33)^2}{4(0.05)^2} = 543$$

Demek ki kabul edilebilir en ufak değer 543.