Normal Dağılımın Asimetri ve Basıklık Ölçülerinin Bulunması

Asimetri ve basıklık ölçüleri olan α_3 ve α_4 ile gösterilen kavramlar çok önemlidir. Çünkü, yapılan bütün araştırmaların çoğunda anakütlenin normal dağıldığı varsayımı ileri sürülür. Eğer bu varsayımdan kuşkuya düşülürse, normallik testi denilen α_3 ve α_4 değerleri hesaplanarak, bu değerlerin sırası ile sıfır ve üç değerlerinden ne kadar saptıkları veya sapmadıkları araştırılır.

$$\alpha_3 = \frac{\mu_3}{\sigma^3} \qquad \mu_r = E[(X-\mu)^r] \qquad M_r = E[X^r]$$

şeklinde bulunuyordu. μ_3 aritmetik ortalamaya göre üçüncü momenttir. Sıfıra göre momentler bilindiğinden dolayı, König teoremi kullanılarak μ_3 elde edilebilir.

$$\mu_r = E[(X-\mu)^r] = \sum_{j=0}^{r} \binom{r}{j} E(X^j)(-1)^{r-j}(\mu)^{r-j}$$

$$= \sum_{j=0}^{r} \binom{r}{j}(-1)^{r-j} M_j (M_1)^{r-j}$$

$$\mu_3 = \sum_{j=0}^{3} \binom{3}{j}(-1)^{3-j} M_j (M_1)^{3-j}$$

$$= M_3 - 3M_2M_1 + 2M_1^3 = \cancel{3\mu\sigma^2} + \cancel{\mu^3} - \cancel{3\mu\sigma^2} - \cancel{3\mu^3} + \cancel{2\mu^3} = 0$$

$$M_1 = E(X) = \mu$$

$$M_2 = E(X^2) = M''(0) = \sigma^2 + \mu^2$$

$$M_3 = E(X^3) = M'''(0) = 3\mu\sigma^2 + \mu^3$$

$$M'''(t) = \sigma^2(\mu + \sigma^2 t)\, e^{\mu t + \frac{\sigma^2 t^2}{2}} + 2(\mu + \sigma^2 t)\sigma^2 e^{\mu t + \frac{\sigma^2 t^2}{2}} + (\mu + \sigma^2 t)^3 e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

$$M'''(0) = \mu\sigma^2 + 2\mu\sigma^2 + \mu^3 = 3\mu\sigma^2 + \mu^3$$

Basıklık formülü

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

olarak tanımlanmıştı. Konig teoremine göre

$$\mu_r = \sum_{j=0}^{r} \binom{r}{j} (-1)^{r-j} M_j (M_1)^{r-j}$$

$$\mu_4 = M_4 - 4 M_3 M_1 + 6 M_2 M_1^2 - 3 M_1^3$$

$$\mu_4 = \sum_{j=0}^{4} \binom{4}{j} (-1)^{4-j} M_j (M_1)^{4-j}$$

$$= \binom{4}{0}(-1)^4 M_0 (M_1)^4 + \binom{4}{1}(-1)^3 M_1 (M_1)^3 + \binom{4}{2}(-1)^2 M_2 (M_1)^2$$
$$+ \binom{4}{3}(-1)^1 M_3 (M_1)^1 + \binom{4}{4}(-1)^0 M_4 (M_1)^0$$

$$= M_1^4 - 4 M_1^4 + 6 M_2 . M_1^2 - 4 . M_3 . M_1 + M_4$$
$$= M_4 - 4 M_3 M_1 + 6 M_2 M_1^2 - 3 M_1^3$$

$$M_4 = E(X^4) = M^{IV}(t=0)$$

$$M^{IV}(t) = \sigma^4 e^{\mu.t + \frac{\sigma^2.t^2}{2}} + \sigma^2 (\mu + \sigma^2.t)^2 e^{\mu.t + \frac{\sigma^2.t^2}{2}} + 2.\sigma^4 e^{\mu.t + \frac{\sigma^2.t^2}{2}}$$
$$+ 2\sigma^2 (\mu + \sigma^2.t)^2 . e^{\mu.t + \frac{\sigma^2 t^2}{2}} + 3.(\mu + \sigma^2.t)^2 . \sigma^2 . e^{\mu.t + \frac{\sigma^2 t^2}{2}}$$
$$+ (\mu + \sigma^2.t)^4 . e^{\mu.t + \frac{\sigma^2 t^2}{2}}$$

$$M^{IV}(t) = \sigma^4 + \sigma^2.\mu^2 + 2\sigma^4 + 2\sigma^2.\mu^2 + 3.\mu^2.\sigma^2 + \mu^4$$

$$= \mu^4 + 6\sigma^2.\mu^2 + 3\sigma^4$$

$$\mu_4 = 3\sigma^4$$

elde edilir. O zaman

$$\alpha_4 = \frac{3\sigma^4}{\sigma^4} = 3$$

olur.

Standart Normal Dağılım

X rassal değişkeni ortalaması μ ve standart sapması σ (varyansı σ^2) olmak üzere normal dağılmış olsun.

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$z=\frac{x-\mu}{\sigma} \Rightarrow dx=dz.\sigma$ olur. Bunlar integralde yerine konulursa

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$$

elde edilir.

Ortalaması sıfır, varyansı bir ve olasılık fonksiyonu aşağıdaki gibi tanımlanan normal dağılıma, standart normal dağılım denir ve kısaca N(0,1) olarak gösterilir.

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \quad , -\infty<z<\infty$$

Binom Dağılımına Normal Yaklaşım

Binom dağılımı incelemiş ve X elde edilen başarıların sayısını göstermek üzere

$$E(X)=n.p \text{ ve } V(X)=n.p.q$$

olarak tanımlanmıştı. Binomun, normal dağılıma yaklaşımında belirleyici olan değerler p ve n değerleridir.

p değeri 1/2'ye yaklaştıkça ve bağımsız deney sayılarını gösteren n'nin değeri de arttıkça, binom dağılımı normal dağılıma yaklaşır.

Teorem Bir tek deneydeki başarı olasılığı p olmak üzere, X rassal değişkeni n tane bağımsız deneydeki başarıların sayısını göstersin. O zaman standart hale getirilmiş binom rassal değişkeni Z; E(X)=n.p ve V(X)=n.p.q olmak üzere, deneylerin sayısı yeterli bir sayıda olmak üzere, ortalaması 0 ve varyansı 1 olmak üzere yaklaşık bir normal dağılıma sahiptir. Yani,

$$Z=\frac{X-E(X)}{\sqrt{V(X)}} \quad ; \quad Z=\frac{X-n.p}{\sqrt{n.p.q}} \sim N(0,1)$$

Böylece, X rassal değişkeninin sahip olacağı başarı sayılarının olasılığı binom dağılımı yerine, normal dağılımdan gidilerek bulunabilir.

$$P(a \leq X \leq b) \cong P\left[\frac{a-n.p}{\sqrt{n.p.q}} \leq \frac{X-n.p}{\sqrt{n.p.q}} \leq \frac{b-n.p}{\sqrt{n.p.q}}\right]$$

$$\cong P\left[\frac{a-n.p}{\sqrt{n.p.q}} \leq Z \leq \frac{b-n.p}{\sqrt{n.p.q}}\right]$$

Eğer n, yeterli derecede büyükse, bu yaklaşım için sorun yoktur. Fakat n yeterli derecede büyük değilse, bir süreklilik düzeltmesine gerek vardır.

Süreklilik düzeltmesi

n yeterli derecede büyük değilse, bir süreklilik düzeltmesine gerek vardır. Bunun nedenlerinden birisi de, binom dağılımı kesikli bir dağılım iken, normal dağılımın sürekli bir dağılım olmasıdır.

Hangi durumda süreklilik düzeltmesinin kullanılabileceğine karar vermek için n'i sınırlayan iki tane yaklaşım vardır.

a) $p < \frac{1}{2}$ ise $n > \frac{9.q}{p}$ olması durumunda,

veya

$p > \frac{1}{2}$ ise $n > \frac{9.p}{q}$ olması durumunda

süreklilik düzeltilmesine gerek yoktur.

b) $n \geq 50$ ise süreklilik düzeltmesine gerek yoktur.

$20 \leq n < 50$ ise süreklilik düzeltmesine gerek vardır.

Süreklilik düzeltmesi şu şekilde uygulanır.

$$P(X \leq a) \cong P\left[Z \leq \frac{a+0.5-n.p}{\sqrt{n.p.q}}\right] \qquad P(X < a) \cong P\left[Z < \frac{a-0.5-n.p}{\sqrt{n.p.q}}\right]$$

$$P(X \geq a) \cong P\left[Z \geq \frac{a-0.5-n.p}{\sqrt{n.p.q}}\right] \qquad P(X > a) \cong P\left[Z > \frac{a+0.5-n.p}{\sqrt{n.p.q}}\right]$$

Bu dört olasılıktan gidilerek, aralık halinde bulunan olasılıkların değerleri de kolaylıkla yazılabilir.

Örneğin $P(a \leq X \leq b)$ olasılığı

$$P(a \leq X \leq b) \cong P\left(\frac{a-0.5-n.p}{\sqrt{n.p.q}} \leq Z \leq \frac{b+0.5-n.p}{\sqrt{n.p.q}}\right)$$

süreklilik düzeltmesi ile yukarıdaki gibi bulunur.

Poisson Dağılımına Normal Yaklaşım

X rassal değişkeni bir Poisson değişkeni ise,

$$E(X) = V(X) = \lambda$$

şeklinde elde edilir.

Teorem

X, poisson dağılımına sahip bir rassal değişken ise, λ'nın çok büyük olması durumunda, standart hale getirilmiş Poisson değişkeni Z, $E(X)=V(X)=\lambda$ olmak üzere, sıfır ortalama ve bir varyansı ile yaklaşık bir normal dağılıma sahiptir.

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} \quad ; \quad Z = \frac{X - \lambda}{\sqrt{\lambda}} \sim N(0,1)$$

Poisson dağılımından yararlanılarak elde edilecek olan olasılıklar, λ'nın yeterli derecede büyük olması durumunda aşağıdaki eşitliklede bulunabilir.

$$P(a \leq X \leq b) \cong P\left[\frac{a-\lambda}{\sqrt{\lambda}} \leq Z \leq \frac{b-\lambda}{\sqrt{\lambda}}\right]$$

Eğer süreklilik düzeltmesi kullanılırsa, bu olasılık

$$P(a \leq X \leq b) \cong P\left[\frac{a-0.5-\lambda}{\sqrt{\lambda}} \leq Z \leq \frac{b+0.5-\lambda}{\sqrt{\lambda}}\right]$$

olur.