2.5. Örnekleme Dağılımları

Örneklemin parametre içermeyen bir fonksiyonuna **istatistik** denir. Kitle ortalamasının bir tahmin edicisi olan, örneklem ortalaması $\bar{X}$ ile kitle varyansının bir tahmin edicisi olan, örneklem varyansı S^2 birer istatistiktir. N birimlik bir kitleden, belli sayıda n birimlik örneklem seçilirse, seçilen her n birimlik örneklem için hesaplanan bu ortalama ve varyans değerleri birbirinden farklı olur. Yani bu istatistikler örnekten örneğe değişkenlik gösterir. Bu nedenle, bu istatistiklerin de kendilerine ait dağılımları, beklenen değerleri ve varyansları vardır. Bir istatistiğin dağılımına, **örnekleme dağılımı** denir. Örnekleme dağılımları, hipotez testi yapmak için gerekli bilgiyi sağlar.

2.5.1.Örneklem Ortalamasının Dağılımı

Ortalaması μ, varyansı σ^2 olan bir kitleden $X_1, X_2, \ldots, X_n$ gibi n birimlik bir örneklem seçilsin. Kitle ortalaması μ'ye ilişkin sonuç çıkarımı yapılmak istendiğinde akla ilk gelen istatistik örneklem ortalaması olacaktır. Örneklem ortalaması $\bar{X}$ ile gösterilmek üzere n birimlik bir örneklem için,

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

bir rastgele değişkendir. Gözlemler alınıp yerine koyulduğunda bu artık rastgele değişkenin bir değeridir ve almış olduğu değer $\bar{x}$ ile gösterilir. $\bar{X}$ rastgele değişkeninin örnekleme dağılımı göz önüne alınsın. Beklenen değerin özelliklerinden faydalanarak,

$$E(\bar{X}) = \frac{E(\sum_{i=1}^{n} X_i)}{n} = \frac{n\mu}{n} = \mu$$

olarak elde edilir ve aynı şekilde varyans tanımından;

$$Var(\bar{X}) = Var(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

olarak elde edilir. Dikkat edilirse $\bar{X}$'nin örnekleme dağılımının varyansı örneklem büyüklüğü n arttıkça küçülür. Bu, örneklemde ne kadar çok gözlem varsa örneklem ortalamasının örnekleme dağılımı da kitle ortalamasına o kadar yakın toplanmış demektir. Yani örnek sayısı arttıkça kitleye ilişkin sonuç çıkarımı daha iyi olur. Örneklem ortalamasının varyansı $\sigma_{\bar{X}}^2$ ile gösterilmek üzere $\bar{X}$'nin standart sapmasına standart hata denir ve

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

ile hesaplanır.

Bahsedilen bu sonuçlar herhangi bir dağılım varsayımı olmadan tüm rastgele değişkenler için geçerlidir.

2.5.1.1.Örneklem Ortalamasının Örnekleme Dağılımının Şekli

Örnekleme dağılımının şekli rasgele değişkenin ait olduğu kitlenin dağılımının şekli ve örneklemin büyüklüğü ile belirlenir. Kitlenin dağılımı normal ise, $\bar{X}$ örneklem ortalamasının dağılımı da normal olacaktır. Eğer, kitlenin dağılımı normal değilse $\bar{X}$ örneklem ortalamasının dağılımının şekli örneklem büyüklüğüne bağlıdır. Örneklem büyüklüğü küçük ise, $\bar{X}$ örneklem ortalamasının örnekleme dağılımının şekli, kitle dağılımının şekline benzeyecektir.
Örneklemlerin seçildiği kitlenin dağılımı bilinmiyor ise (sonlu ya da sonsuz) $\bar{X}$ örneklem ortalamasının örnekleme dağılımı, örneklem sayısı n'nin büyük olması koşulu ile merkezi limit teoreminden dolayı yaklaşık olarak normal dağılım olacaktır.

2.5.2.Örneklem Varyansının Dağılımı

Ortalaması μ, varyansı σ^2 olan bir kitleden $X_1, X_2, \dots, X_n$ gibi n birimlik bir örneklem seçilsin. Kitle varyansının yansız bir tahmin edicisi olan örneklem varyansı S^2 aşağıdaki gibi tanımlanır;

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Bunun yansız bir tahmin edici olmasından dolayı,

$$E(S^2) = \sigma^2$$

dir. Örneklem varyansı S^2'nin örnekleme dağılımını belirleyebilmek için kitle dağılımına ilişkin daha fazla bilgiye ihtiyaç vardır. Uygulamalarda, genellikle kitle dağılımının normal olduğu varsayımı kullanılır.

Kitle dağılımının normal dağılım olması durumunda örneklem ortalaması ve örneklem varyansının dağılımlarına bakılsın.

Örneklem sayısı büyük olduğunda, örneklem ortalaması $\bar{X}$'nin dağılımı için merkezi limit teoremi kullanılarak sonuç çıkarımı yapılabilir. Fakat örneklem varyansı için aynı şey geçerli değildir. Kitlenin normal dağılıma sahip olduğu varsayımı altında örneklem varyansı S^2'nin dağılımı ile ilgili sonuç çıkarımı yapılabilir.

$X_1, X_2, \dots, X_n$ ortalaması μ, varyansı σ^2 olan normal dağılıma sahip bir kitleden rastgele çekilen örneklem olmak üzere,

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

rastgele değişkeni $(n-1)$ serbestlik dereceli χ^2 **(**ki-kare) dağılımına sahiptir. Serbestlik derecesi $(n-1)$ olan χ^2 dağılımına sahip bir rastgele değişkenin beklenen değeri $(n-1)$ ve varyansı $2(n-1)$'dir. Buna göre;

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = (n-1)$$

ve beklenen değerin özelliğinden dolayı,

$$\frac{(n-1)}{\sigma^2}E(S^2) = (n-1)$$

$$E(S^2) = \sigma^2$$

olur. S^2'nin varyansını bulmak için varyansın özelliğinden yararlanılırsa;

$$Var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

olup

$$\frac{(n-1)^2}{\sigma^4} Var(S^2) = 2(n-1)$$

$$Var(S^2) = \frac{2\sigma^4}{(n-1)}$$

olarak elde edilir.

2.5.3.Örneklem Oranının Örnekleme Dağılımı

Uygulamalarda kitlede belirli özelliğe sahip olanların oranlarının bilinmesi istenebilir. Bu oranı belirlemek için kitlede istenen özelliğe sahip birimlerin sayılması gerekir ki kitle büyük olduğunda bu sayma işi zor ve hatta bazı durumlarda imkansız olabilir. Bu nedenle kitleden örneklem çekilerek kitle oranı örneklemden tahmin edilir. Bu oranla ilgili çıkarım yapabilmek için oranın örnekleme dağılımının bilinmesi gerekir.

Bir kitlede belirli özelliğe sahip olanların sayısı X ile gösterilmek üzere X, binom dağılımına sahip bir rasgele değişken olacaktır. Bu binom dağılımının parametreleri n ve p olmak üzere, $E(X) = np$ ve $Var(X) = np(1-p)$ dir. Belirli özelliğe sahip olanların oranına P denilirse, örneklemden elde edilen $P = X/n$ oranı, p için bir tahmin edici olarak alınabilir. P tahmin edicisinin örneklemden elde edilen değeri (yani p için bir tahmin) $\hat{p}$ ile gösterilsin. P'nin beklenen değer ve varyansı,

$$E(P) = E(\frac{X}{n}) = \frac{np}{n} = p$$

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

olarak elde edilir.

$$Var(P) = \frac{p(1-p)}{n}$$

olmak üzere daha önce bahsedildiği gibi P'nin standart sapması P'nin standart hatasıdır ve,

$$\sigma_p = \frac{p(1-p)}{n}$$

biçiminde hesaplanır.

Küçük örneklemlerde oran fazla bir anlam ifade etmeyeceğinden oranla ilgili bir tahmin söz konusu olduğunda örneklem hacminin büyük olduğu varsayılır ve merkezi limit teoremi gereği, P'nin örnekleme dağılımın ortalaması p ve varyansı $\frac{p(1-p)}{n}$ olan normal dağılıma yakınsadığı söylenebilir. Buna göre,

$$Z = \frac{P - p}{\frac{p(1-p)}{n}}$$

olup $Z \sim N(0,1)$' dir. P'nin örnekleme dağılımı, kitle parametresi p' nin güven aralığının bulunmasında ve hipotez testlerinde kullanılacaktır.

Örnek: Bir köyde yaşayanların %70 inin kendi tarlasına sahip olduğu biliniyor.

a) Bu köyde yaşayanların %75'inden fazlasının kendi tarlasına sahip olması olasılığı nedir?
b) Bu köyde yaşayanlardan kendi tarlasına sahip olanların oranının %55 ile %85 arasında olması olasılığı nedir?

a) İlk olarak P rasgele değişkeninin örnekleme dağılımı belirlenmelidir.

$$E(P) = p = 0.70$$

$$Var(P) = \frac{p(1-p)}{n} = \frac{0.70 \times 0.30}{100} = 0.0021$$

olup yukarıda verilen bilgilerden $P \sim N(0.70, 0.0021)$ olduğu söylenebilir. Buna göre istenen olasılık,

$$P(P \geq 0.75) = P(\frac{P - 0.70}{\sqrt{0.0021}} \geq \frac{0.75 - 0.70}{\sqrt{0.0021}})$$

$$= P(Z \geq \frac{0.75 - 0.70}{\sqrt{0.0021}})$$

olarak standart normal dağılım tablosundan bulunur.

b) a) şıkkına benzer biçimde,

$$P(0.55 \leq P \leq 0.85) = P(\frac{0.55 - 0.70}{\sqrt{0.0021}} \leq \frac{P - 0.70}{\sqrt{0.0021}} \leq \frac{0.75 - 0.70}{\sqrt{0.0021}})$$

$$= P(\frac{0.55 - 0.70}{\sqrt{0.0021}} \leq Z \leq \frac{0.75 - 0.70}{\sqrt{0.0021}})$$

olarak hesaplanır.

2.6.Bazı Örnekleme Dağılımları

2.6.1.Ki–kare (χ^2) Dağılımı

İstatistik uygulamalarında sıkça kullanılan bu dağılım aslında gamma dağılımının özel bir halidir. Gamma dağılımında parametreler, $\alpha = \frac{\nu}{2}$, $\beta = 2$ olarak alınırsa bu ν serbestlik dereceli bir ki-kare dağılımı olur. X rastgele değişkeni serbestlik derecesi ν olan bir ki-kare dağılımına sahip ise yoğunluk fonksiyonu;

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2 - 1} e^{x/2}, x > 0$$

biçimindedir ve $X \sim \chi_\nu^2$ ile gösterilir. Bu durumda X rastgele değişkeninin beklenen değeri,

$$\mu = \nu$$

ve varyansı

$$\sigma^2 = 2\nu$$

olur.

2.6.1.1.Ki-kare dağılımına uyan bazı rastgele değişkenler

- $X \sim N(\mu, \sigma^2)$ dağılımına uyan bir rastgele değişken olmak üzere $Z = \frac{X-\mu}{\sigma}$ rastgele değişkeninin $N(0,1)$ dağılımına sahiptir. Bu durumda,

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_1^2$$

olur.

- $X_1, X_2, \dots, X_n$ rastgele değişkenleri $N(\mu, \sigma^2)$ dağılımına sahip rastgele değişkenlerin bir dizisi olmak üzere,

$$Z_i = \frac{X_i - \mu}{\sigma}$$

ile gösterilirse,

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$$

olur.

- $X_1, X_2, \dots, X_n$ rastgele değişkenleri $N(\mu, \sigma^2)$ dağılımına sahip rastgele değişkenlerin bir dizisi olmak üzere,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

olur. Yani $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ ise

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

dir.

2.6.2. t Dağılımı

Z, ortalaması 0, varyansı 1 olan standart normal dağılıma sahip bir rastgele değişken, X'de serbestlik derecesi ν olan ki-kare dağılımına sahip bir rastgele değişken olsun. Z ile X birbirinden bağımsız olmak üzere;

$$T = \frac{Z}{\sqrt{\frac{X}{\nu}}}$$

şeklinde tanımlanan T rastgele değişkeni serbestlik derecesi ν olan t dağılımına sahip bir rastgele değişkendir. Serbestlik derecesi ν olan t dağılımına sahip rastgele değişkenin olasılık yoğunluk fonksiyonu,

$$f(t) = \frac{\Gamma\left({(\nu+1)}/{2}\right)}{\Gamma\left(\left({\nu}/{2}\right)\sqrt{\pi\nu}\right)}\left(1 + \frac{t^2}{\nu}\right)^{-{(\nu+1)}/{2}}, -\infty < t < \infty.$$

biçiminde yazılır.

$t -$dağılımı normal dağılım gibi ortalamaya göre simetrik bir dağılımdır, serbestlik derecesi arttıkça standart normal dağılıma yakınlaşır.

2.6.3.F Dağılımı

U ve V sırasıyla ν_1 ve ν_2 serbestlik dereceleri ile ki-kare dağılımına sahip birbirinden bağımsız iki rastgele değişken olmak üzere,

$$F = \frac{U/_{\nu_1}}{V/_{\nu_2}}$$

rastgele değişkeni ν_1, ν_2 serbestlik dereceli F dağılımına sahip bir rastgele değişkendir ve yoğunluk fonksiyonu,

$$f(f) = \frac{\Gamma\left({(\nu_1+\nu_2)}/_{2}\right)}{\Gamma\left({\nu_2}/_{2}\right)\Gamma\left({\nu_1}/_{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/_2} \frac{f^{\nu_1-2}/_2}{\left[1+\left(\frac{\nu_1}{\nu_2}\right)f\right]^{(\nu_1+\nu_2)}/_2}, f > 0$$

biçiminde gösterilir.

2.6.3.1.F dağılımı ile ilgili çeşitli özellikler:

1) $X \sim F(\nu_1, \nu_2)$ ise $1/_X \sim F(\nu_2, \nu_1)$'dir.

2) $X \sim t(\nu)$ ise $X^2 \sim F(1, \nu)$'dir.

3) $X \sim F(\nu_1, \nu_2)$ ise $F_{\nu_{1-1}, \nu_{2-1,1-}\alpha/_2} = \frac{1}{F_{\nu_{2-1}, \nu_{1-1,}\alpha/_2}}$'dir.