Giriş

i) Regresyon kavramı ve model kurma: Regresyon analizi değişkenler arasındaki ilişkiyi incelemek ve modellemek için kullanılan istatistiksel bir yöntemdir. Regresyon analizinin en önemli alanlarından biri değişkenler arasındaki ilişkiyi açıklayabilecek doğru modele karar vermektir. Modele karar verebilmek için değişkenler arasındaki ilişkinin ön analizinin yapılmasıdır. Bu analize dayalı olarak model belirlenmelidir.

ii) Regresyon analizi yapabilmek için sağlıklı veri toplama yöntemleri ve bilgisayarın etkin olarak kullanılması : Bir istatistiksel analiz sonucunda sağlıklı bilgiye ulaşabilmek için analizin yapılması için kullanılacak olan verinin istatistiksel veri toplama kurallarına dayalı olarak sağlıklı bir şekilde elde edilmiş olması gerekmektedir. Aynı şekilde, regresyon analizinden sağlıklı sonuçların elde edilebilmesi içinde analizi yapabilmek için kullanılacak olan verilerin kuralara uygun olarak elde edilmiş olması gerekmektedir. Dolayısıyla, verilerin toplanabilmesi için regresyon yönteminin uygulanacağı problemin yapısı çok iyi anlaşılmalıdır. Probleme bakılarak, regresyon analizi tarihsel verilere bakılarak geriye yönelik bir çalışma olabilir. Sadece gözlemsel veriler kullanılarak yapılacak olan bir analiz olabilir. Ya ada tasarlanmış bir deney sonucunun analizi olabilir. Günümüzde verilerin boyutları oldukça büyüktür. Çok fazla değişkenle aynı anda ilgileniliyor olabilir ve örneklem hacimde oldukça büyük olabilir. Bu tür verilerin analizinde bilgisayarın kullanılması kaçınılmazdır. Ancak, Sonuçların yorumlanması ve elde edilecek olan sonuçların kullanılması analizi yapanların işidir.

iii) Regresyonun kullanım alanları: Regresyonun, mühendislik, fizik ve kimya bilimler, iktisat, yönetim, yaşam ve biyoloji bilimleri ve sosyal bilimler gibi hemen hemen tüm alanlarda farklı amaçlarla kullanılmaktadır. Bu amaçlardan bazıları şunlardır: verinin tanımlanması ve özetlenmesi, parametre kestirimleri, kestirim ve önkestirimler ve denetleme.

Koşullu beklenen değer ve regresyon kavramı

X ve Y rasgele değişkenlerinin koşullu dağılımı biliniyorsa $X = k$ verildiğinde Y'nin koşullu beklenen değeri

$$E(Y|X = x) = \begin{cases} \sum_{y \in D_y} y\, P_{Y|x}(y); \, kesikli\ durum \\ \int_{y \in D_y} y f_{y|x}(y)dy; \, sürekli\ durum \end{cases}$$

Benzer olarak,

$$E(X|Y=y)=\begin{cases}\sum_{x\in D_x} x\,P_{X|y}(x);\,kesikli\ durum\\ \int_{x\in D_x} xf_{X|y}(x)dx;\,sürekli\ durum\end{cases}$$

ile verilir.

Ayrıca $f(x,y)$ X ile Y rasgele değişkenlerinin ortak olasılık yoğunluk fonksiyonu olmak üzere regresyon analizinde

$$\mu_{y|x}=E(Y|X=x)=\int_{-\infty}^{\infty} yf_{Y|x}(Y|x)dy$$

Eşitliğine Y'nin X' e göre regresyon denklemi denir.

Örnek

$$f(x,y)=\begin{cases}6x\,, & 0<x<y<1\\ 0\,, & dy\end{cases}$$

a) $\mu_{Y|x}=?$

b) $\mu_{X|y}=?$

Çözüm

a) $\mu_{Y|x}=E(Y|X=x)=\int_{-\infty}^{\infty} yf_{y|x}(Y|x)dy$

$$f_{y|x}(y)=\frac{f(x|y)}{f(x)}$$

$$f(x)=\int_{y=x}^{1} f(x,y)dy=\int_{y=x}^{1} 6xdy=6xy|_x^1=6x(1-x),\quad 0<x<1$$

$$f_{y|x}(y)=\frac{6x}{6x(1-x)}=\frac{1}{1-x}$$

$$\mu_{Y|x} = E(Y|X=x) = \int_{y=x}^{1} y f_{y|x}(y)dy = \int_{y=x}^{1} \frac{y}{1-x}dy = \frac{1}{1-x}\frac{y^2}{2}\ |_x^1$$

$$= \frac{1}{1-x}\frac{1-x^2}{2} = \frac{x+1}{2} = \frac{1}{2}+\frac{1}{2}x$$

b) $\mu_{X|y} = E(X|Y=y) = \int_{-\infty}^{\infty} y f_{y|x}(Y|x)dy$

$$f_{x|y}(x) = \frac{f(x|y)}{f(y)}$$

$$f(y) = \int_0^y f(x,y)dx = \int_0^y 6x dx = 3x^2|_0^y = 3y^2, \quad 0 < y < 1$$

$$f_{x|y}(y) = \frac{6x}{3y^2}$$

$$\mu_{X|y} = E(X|Y=y) = \int_0^y x f_{X|y}(x)dx = \int_0^y x\frac{6x}{3y^2}dx = \frac{1}{3y^2}2x^3\ |_0^y = \frac{2y^3}{3y^2} = \frac{2y}{3}$$

Ödev 1

y \ x	0	1	2	
0	1/12	1/6	1/24	7/24
1	1/4	1/4	1/40	21/40
2	1/8	1/20	0	7/40
3	1/120	0	0	1/120
	56/120	28/60	8/120	1

a) $\mu_{X|1}$ =?

b) $\mu_{Y|0}$ =?

Ödev 2

$$f(x,y) = \begin{cases} c\dfrac{1}{9}, & 0 < y < x < 3 \\ 0, & \text{ö.}d. \end{cases}$$

a) c sabit sayısını bulunuz.

b) $f(x)$ ve $f(y)$ marjinal olasılık yoğunluk fonksiyonlarını bulunuz.

c) $E(Y/X = x)$ koşullu beklenen değerini hesaplayınız.

Normal Dağılım ve Özellikleri

Bir X rasgele değişkeninin olasılık yoğunluk fonksiyonu,

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

biçiminde olduğunda, X rasgele değişkenine normal dağılıma sahiptir denir. $\mu \in R$ ve $\sigma^2 \in (0,\infty)$ olmak üzere sırasıyla dağılımın konum ve ölçek parametreleridir. $\mu = 0$ ve $\sigma^2 = 1$ olan dağılıma standart normal dağılım denir ve olasılık yoğunluk fonksiyonu aşağıdaki gibi ifade edilir.

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right], \; -\infty < z < +\infty$$

Burada,

$Z = \dfrac{X-\mu}{\sigma}$ şeklinde ifade edilir.

__Örnek__

Bir X rasgele değişkeni için $X \sim N(\mu = 13.5\,, \sigma^2 = 2.25)$ olduğu bilindiğine göre aşağıdaki soruları cevaplandırınız.

a) $P(X > 18) = ?$

b) $P(X < 12) = ?$

__Çözüm__

a) $P(X > 18) = P\left(Z > \dfrac{18-13.5}{1.5}\right) = P(Z > 3) = 1 - P(Z \le 3) = 1 - 0.9987 = 0.0013$

b) $P(X < 12) = P\left(Z < \frac{12 - 13.5}{1.5} \right) = P(Z < -1) = 0.1587$

<u>Ödev</u>

Bir hastanede yapılan kan ölçümlerinde ölçülen bir madde 2 mg. ortalama ve 0,1 mg. standart sapma ile normal dağılıma uymaktadır. Bu maddenin 1,8 ile 2,15mg. dışına düşmesi kişinin hasta olduğunu göstermektedir. Bu verilere göre hasta kişi oranını bulunuz.