Normal bi-variada y regresión lineal

Ejercicio 1

$Y =$ sistólica y $X =$ diastólica.

$X \sim N(73, 8^2)$ y la Y tiene distribución condicional $N(1.6x, 10^2)$ cuando $X = x$.

1. Cuando $X = 73$ la Y tiene distribución normal $N(116.8, 10^2)$. Su densidad condicional es entonces
$$p_Y(y|X = 73) = \frac{1}{\sqrt{2\pi 10^2}} e^{-(y-116.8)^2/(2(10)^2)}$$

2. Estandarizando y usando la tabla vemos que $\mathbf{P}(Y < 115|X = 73) = 0.4286$.

3. $\mathrm{E}(Y|X = 73) = 116.8$.

4. La densidad conjunta es
$$p(x,y) = p_Y(y|X = x)p_X(x) = \frac{1}{2\pi(80)} e^{-\frac{(y-1.6x)^2}{2(10)^2} - \frac{(x-73)^2}{2(8)^2}}$$
que es una normal bi-variada.

   De la teoría general sabemos que la distribución condicional de Y dado que $X = x$ es normal de desvío $\sigma_Y \sqrt{1-\rho^2}$ y media
$$\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_Y) = \frac{\rho\sigma_Y}{\sigma_X}x + \mu_Y - \frac{\rho\sigma_Y}{\sigma_X}\mu_X$$
Sabemos que $\mu_X = 73$ y $\sigma_X = 8$. Luego
$$\frac{\rho\sigma_Y}{8} = 1.6, \quad \mu_Y = 116.8, \quad \sigma_Y\sqrt{1-\rho^2} = 10$$
La primer ecuación implica que ρ es positivo. Despejando ρ obtenemos
$$\frac{12.8}{\rho}\sqrt{1-\rho^2} = 10$$
Entonces $\rho^2 = \frac{(12.8)^2}{(12.8)^2+(10)^2} = 0.62$, o sea $\rho = 0.79$. Luego $\sigma_Y = 16.2$.

Ejercicio 2

Sea X la densidad de la goma e Y la elasticidad. Tenemos entonces que:

- $X \sim N(\mu_X, \sigma_X^2)$ con $\mu_X = 1$ y $\sigma_X = 0.2$.
- $Y \sim N(\mu_Y, \sigma_Y^2)$ con $\mu_Y = 100$ y $\sigma_Y = 8$.
- Coeficiente de correlación: $\rho = -0.8$.

1. Sean $U = \frac{X-\mu_X}{\sigma_X} = \frac{X-1}{0.2} = 5(X-1)$ y $V = \frac{Y-\mu_Y}{\sigma_Y} = \frac{Y-100}{8}$.

   Entonces (U,V) es una normal bivariada estándar con correlación $\rho = -0.8$ y sabemos que $\mathrm{E}(V|U = u) = \rho u$.

   Sabiendo que la densidad es $X = 0.75$, resulta que $U = 5(0.75-1) = -1.25$ y por lo tanto $\mathrm{E}(V|U = -1.25) = -0.8 \times -1.25 = 1$. Volviendo a normalizar, tenemos que la predicción para la elasticidad es $\mathrm{E}(Y|X = 0.75) = 8 \times 1 + 100 = 108$

2. Buscamos un intervalo $I = [108-\varepsilon, 108+\varepsilon]$ tal que

$$P(Y \in I|X = 0.75) = P(|Y - 108| < \varepsilon|X = 0.75) = 0.95$$

Observar que por definición:

$$\begin{aligned} Y - 108 &= Y - \mathrm{E}(Y|X = 0.75) \\ &= \sigma_Y V + \mu_Y - (\sigma_Y \rho U + \mu_Y) = \sigma_Y (V - \rho U) \\ &= \sigma_Y \sqrt{1-\rho^2} Z \end{aligned}$$

con Z normal estándar independiente de U.

Por lo tanto, ε tiene que ser tal que:

$$\begin{aligned} P(Y \in I|X = 0.75) &= P(|\sigma_Y \sqrt{1-\rho^2} Z| < \varepsilon|U = -1.25) \\ &= P\left(|Z| < \frac{\varepsilon}{8\sqrt{1-(0.8)^2}}\right) = 0.95 \end{aligned}$$

donde en la penúltima igualdad usamos que U y Z son independientes.

Utilizando al tabla de la normal "al revés" resulta que $\frac{\varepsilon}{8\sqrt{0.36}} = 1.96$, de donde $\varepsilon = 1.96 \times 4.8 = 9.41$.

El intervalo resulta entonces $I = [108 - 9.41, 108 + 9.41] = [98.6, 117.4]$.

3. Tenemos que calcular

$$\mathrm{E}(Y|X > 0.8) = \mathrm{E}(8V + 100|U > -1) = 8\mathrm{E}(V|U > -1) + 100$$

Utilizaremos el siguiente resultado: si (U,V) normal bivariada estándar con correlación ρ entonces,

$$\mathrm{E}(V|a < U < b) = \frac{\frac{\rho}{\sqrt{2\pi}} \cdot \left(e^{-a^2/2} - e^{-b^2/2}\right)}{\Phi(b) - \Phi(a)}$$

En este caso $a = -1$ y $b = +\infty$, de donde $e^{-b^2/2} = 0$ y $\Phi(b) = 1$, es decir:

$$\mathrm{E}(V|U > -1) = \frac{\frac{-0.8}{\sqrt{2\pi}} \cdot e^{-1/2}}{1 - \Phi(-1)} = \frac{-0.8e^{-1/2}}{\sqrt{2\pi}\Phi(1)} = \frac{-0.8e^{-1/2}}{\sqrt{2\pi}0.8413} = -0.23.$$

Resulta entonces que la predicción es $\mathrm{E}(Y|X > 0.8) = -8 \times 0.23 + 100 = 98.16$.

Ejercicio 3

Dado que $X = 60$, la Y tiene distribución normal de parámetros

$$\mu = \rho \frac{\sigma_Y}{\sigma_X}(60 - \mu_X) + \mu_Y = 145 \text{ y } \sigma = \sigma_Y \sqrt{1-\rho^2} = 20\frac{\sqrt{7}}{4} = 13.23.$$

1. Entonces

$$\mathbf{P}(Y \geq y|X = 60) = \mathbf{P}\left(\frac{Y - 145}{13.23} \geq \frac{y - 145}{13.23}\right) = 0.05,$$

por lo que $\frac{y-145}{13.23} = 1.645$, de donde $y = 166.76$.

2. Basta encontrar y' tal que $\mathbf{P}(Y \leq y'|X = 60) = 0.05$. Los mismos cálculos que en la parte a) dan $y = 145 - 1.645 \cdot 13.23 = 123.24$. Por lo tanto un intervalo que sirve es $[123.24, 166.76]$.

Ejercicio 4

Llamemos X al puntaje en la primera prueba e Y al puntaje en la segunda.

1. Usaremos la recta de regresión. Recordar que su fórmula es

$$\mathrm{E}(Y|X = x) = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X) = 0.54x + 39.9$$

   Nuestro pronóstico es entonces $\mathrm{E}(Y|X = 60) = 73.5$.

2. Queremos calcular $\mathbf{P}(|Y - \mathrm{E}(Y|X = 60)| > 5|X = 60)$. Definimos las variables estandarizadas

$$U = \frac{X - \mu_X}{\sigma_X}, \quad V = \frac{Y - \mu_Y}{\sigma_Y}$$

   Por definición existe Z normal estándar, independiente de U tal que

$$V = \rho U + \sqrt{1 - \rho^2} Z$$

   Si $X = 60$, entonces $U = -0.5$. Además,

$$Y - \mathrm{E}(Y|X = 60) = \sigma_Y(V - \rho(-0.5)) = \sigma_Y \sqrt{1 - \rho^2} Z.$$

   Entonces

$$\begin{aligned}\mathbf{P}(|Y - \mathrm{E}(Y|X = 60)| > 5|X = 60) &= \mathbf{P}\left(|\sigma_Y \sqrt{1 - \rho^2} Z| > 5|U = -0.5\right) \\ &= \mathbf{P}(|Z| > 5/(0.8 \cdot 9)) = \mathbf{P}(|Z| > 0.694) \\ &= 0.4877\end{aligned}$$

3. Seguiremos trabajando con las variables estandarizadas. Debemos calcular

$$\frac{\mathbf{P}(V > 0, U > 0)}{\mathbf{P}(U > 0)} = 2\mathbf{P}(V > 0, U > 0)$$

   Usando que $V = \rho U + \sqrt{1 - \rho^2} Z$, queremos calcular

$$\mathbf{P}\left(Z > -\frac{\rho}{\sqrt{1 - \rho^2}} U, U > 0\right)$$

   que es un sector de ángulo $\theta + \pi/2$ en el plano. La tangente de θ es $\rho/\sqrt{1 - \rho^2} = 0.75$, de donde $\theta = 0.6435$. Por la simetría rotacional (U y Z son independientes), la probabilidad buscada es

$$\frac{\pi/2 + \theta}{2\pi} = \frac{1}{4} + 0.1024 = 0.3524.$$