1. Marco teórico

1.1. Correlación lineal

Cuando se tiene una variable controlada x y una dependiente y tenemos el modelo lineal

$$Y = \beta_0 + \beta_1 x + \epsilon \tag{1}$$

que implica entonces el modelo para análisis de rendimiento promedio:

$$E(Y) = \beta_0 + \beta_1 x. \tag{2}$$

Si la variable x es un valor observado de una variable X, al establecerse una relación funcional y al basarse en (2) se implica el modelo

$$E(Y|X = x) = \beta_0 + \beta_1 x \tag{3}$$

que supone la esperanza condicional de Y para un valor fijo de X en una función lineal del valor x. Al suponer que la variable aleatoria vectorial (X, Y) tiene una distribución normal bivariable con $E(X) = \mu_X, E(Y) = \mu_Y, V(X) = \sigma_X^2, V(Y) = \sigma_Y^2$, el coeficiente de correlación ρ puede demostrar que

$$E(Y|X = x) = \beta_0 + \beta_1 x, \qquad donde\ \beta_1 = \frac{\sigma_Y}{\sigma_X}\rho. \tag{4}$$

Si (X, Y) tiene una distribución normal bivariable, entonces la prueba de independencia es equivalente a probar si el coeficiente de correlación ρ es igual a cero. Denotando con $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ una muestra de aleatoria de distribución normal bivariante. El estimador de máxima probabilidad de ρ está dado por el coeficiente de correlación muestral:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \ \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}. \tag{5}$$

Puede expresarse r en términos de cantidades conocidas:

$$r = \frac{Sxy}{\sqrt{SxxSyy}} = \hat{\beta}\sqrt{\frac{Sxx}{Syy}}. \tag{6}$$

Cuando (X, Y) tenga una distribución normal bivariable, se sabe que

$$E(Y|X = x) = \beta_0 + \beta_1 x, \qquad donde\ \beta_1 = \frac{\sigma_Y}{\sigma_X}\rho. \tag{7}$$

Pruebas en las que los conjuntos de hipótesis que contienen β_1, por ejemplo, $H_a\colon \beta_1 = 0$ contra $H_a\colon \beta_1 > 0$ $H_a\colon \beta_1 < 0$, así como $H_a\colon \beta_1 > 0$ contra $H_a\colon \beta_1 \neq 0$ pueden estar basadas en el estadístico

$$t = \frac{\hat{\beta_1} - 0}{S/\sqrt{Sxx}}, \tag{8}$$

Wackerly, Mendenhall III y Scheaffer consideran que "parecería lógico usar r como estadístico de prueba para probar hipótesis más generales acerca de ρ, pero la distribución de probabilidad para r es difícil de obtener." Dennis D Wackerly, william Mendenhall III y Richard L Scheaffer. *Estadística matemática con aplicaciones*. Trad. por Jorge Humberto Romo Mufioz. 7.ª ed. Ciudad de México: Cenage Learning, 2009. ISBN: 9780495110811 Sin embargo, en muestras moderadamente grandes podemos probar la hipótesis $H_0\colon \rho_1 = \rho_0$ con una prueba Z en la que

$$Z = \frac{(\frac{1}{2})\ln(\frac{1+r}{1-r}) - (\frac{1}{2})\ln(\frac{1+\rho}{1-\rho})}{\frac{1}{\sqrt{n-3}}}. \tag{9}$$

Si α es la probabilidad deseada de cometer un error tipo I, la forma de la región de rechazo depende de la hipótesis alternativa. Las diversas alternativas de interés más frecuente y correspondientes regiones de rechazo son las siguientes:

$$H_a\colon \rho > \rho 0, \quad RR\colon z > z_\alpha \tag{10}$$

$$H_a\colon \rho < \rho 0, \quad RR\colon z < -z_\alpha \tag{11}$$

$$H_a\colon \rho \neq \rho 0, \quad RR\colon |z| > z_\alpha/2 \tag{12}$$

La suma de los cuadrados del error *SSE*, es es una alternativa para medie la variación en valores que permanecen sin explicación después de usar las x para ajustar el modelo de regresión lineal simple, la razón SSE/S_{yy} la proporción de la variación total en las y_i que este modelo no explica. El coeficiente de determinación se puede escribir como

$$r^2 = (\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}})^2 = (\frac{S_{xy}}{S_{xx}})(\frac{S_{xy}}{S_{yy}}) = (\frac{\hat{\beta_1}S_{xy}}{S_{yy}}) = \frac{S_{yy} - SEE}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}. \tag{13}$$

Podemos interpretar a r^2 como la proporción de la variación total en las y_i que es explicada por una variable x en un modelo de regresión lineal simple.

1.2. Modelo de base de datos relacional

código	CHAR	fecha	DATE	estado	INTEGER
MX01		2020-01-01		1	
MX02		2020-01-02		1	
MX03		2020-01-03		0	

[Dat12]

1.3. Probabilidad condicional

La probabilidad condicional tiene las mismas características que la probabilidad, pero $P(\cdot|B)$ actualiza nuestra incertidumbre acerca de los eventos para reflejar la evidencia observada en B.

Si A y B son eventos con $P(B) > 0$ entonces la *probabilidad condicional* de A dado B denotado por $P(A|B)$, se define como

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{14}$$

Se denomina probabilidad a priori de A a $P(A)$ y probabilidad a posteriori de A a $P(A|B)$ y es importante mencionar que $P(A|B) \neq P(B|A)$.

La probabilidad condicional es la razón de dos probabilidades y sus consecuencias, la primera de ellas se obtiene moviendo el denominador en la definición al otro lado de la ecuación, para cada evento A y B con posibilidades positivas,

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A). \tag{15}$$

se le conoce como teorema de la *probabilidad de la intersección de dos eventos*. Aplicando repetidamente el teorema (15) aplicado a la intersección de n eventos obtenemos el teorema de *probabilidad de la intersección de n eventos*. Para cualquier evento $A_1, \ldots, A_n$ con probabilidad $P(A_1, A_2, \ldots, A_{n-1}) > 0$,

$$P(A_1, A_2, \ldots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \ldots P(A_n|A_1, \ldots, A_{n-1}). \tag{16}$$

Otro teorema que relaciona a $P(A|B)$ con $P(B|A)$ es la regla *regla de Bayes*:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{17}$$

que se origina directamente del teorema (15) y a su vez se origina directamente de la definición de probabilidad condicional, sin embargo, la regla de Bayes tiene importantes aplicaciones e implicaciones en probabilidad y estadística, ya que en ocasiones es más fácil encontrar $P(B|A)$ que $P(A|B)$ o viceversa. Otra forma de escribir la regla, es en términos de sus odds COMO SE DICE EN ESPAÑOL, ¿cuota?, las odds de un evento A son

$$odds(A) = P(A)\frac{P(A)}{P(A^c)}. \tag{18}$$

Al tomar la expresión $P(A|B)$ y la dividimos entre la $P(A^c|B)$, ambas de la regla de Bayes llegamos a la *forma de odds de la regla de Bayes* que para cualquier eventos A y B con posibilidades positivas, las odds de A after conditioning on B son

$$\frac{P(A|B)}{P(A^c|B)} = \frac{P(B|A)}{P(B|A^c)}\frac{P(A)}{P(A^c)}. \tag{19}$$

En este caso las odds a posteriori $P(A|B)/P(A^c|B)$ son iguales a las odds a priori $P(A)/P(A^c)$ por el factor $P(B|A)/P(B|A^c)$ lo que se le conoce en estadística como *función de verosimilitud*.

La ley de Bayes es usada en ocasiones en conjunto con la *ley de probabilidad total*, que es es esencial para descomponer problemas complicados de probabilidad en problemas partes: Si $A_1, \ldots, A_n$ es una partición de una muestra del espacio S, con $P(A_i) > 0$ para todo i, entonces

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i). \tag{20}$$

Prueba: como los A_i forman una partición de S, podemos descomponer B como

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n). \tag{21}$$

como las partes están disjuntas, podemos agregar sus posibilidades para obtener $P(B)$:

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \ldots + P(B \cap A_n). \tag{22}$$

Aplicando el teorema (15) a cada $P(B \cap A_i)$ obtenemos

$$P(B) = P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n). \tag{23}$$

La *ley de probabilidad total* dice que para obtener la probabilidad incondicional de B, tenemos que dividir el espacio muestra en cortes disjuntos A_i, encontrar la probabilidad condicional de B en cada corte, después tomar la suma pesada de las probabilidades condicionales, donde los pesos son las probabilidades $P(A_i)$.

1.4. Variables aleatorias y sus distribuciones

Dado un experimento con una muestra en espacio S, una *variable aleatoria* es una función del mismo espacio S para los números reales $\mathbb{R}$. Una variable aleatoria X asigna el valor numérico $X(s)$ a cada resultado posible s del experimento. La aleatoriedad viene del hecho que tenemos un experimento aleatorio (con probabilidades descritas por la función de probabilidad P). Las variables aleatorias simplifican la notación y expanden la habilidad de cuantificar y resumir resultados de experimentos.

Se dice que una variable X es discreta cuando si hay una lista finita de valores $a_, a_2, \ldots, a_n$ o un una lista infinita de valores $a_, a_2, \ldots$ de tal forma que $P(X = a_j$ para algún $j) = 1$. Si X es una variable aleatoria discreta, entonces el grupo infinito o contable de valores x tal que $P(X = x)$ se llama *soporte* de X. En contraste una variable aleatoria continua puede tomar cualquier valor real en un intervalo.

La forma más natural de expresar la distribución de variables aleatorias discretas es la *función de probabilidad*[BH19] (PMF, por sus siglas en inglés) que, para una X discreta, es la función p_X dada por $p_X(x) = P(X = x)$. El teorema de *funciones de probabilidad válidas* dice que cuando X es una variable aleatoria con soporte $x1, x2, \ldots$, la función de probabilidad p_X de x debe satisfacer los siguiente criterios:

- No negativo $p_X(x) > 0$ si $x = x_j$ para un j, y $p_X(x) = 0$, de otra forma;
- Suma 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$.

el primer criterio es verdadero porque la probabilidad es no negativa, el segundo es verdadero ya que X debe tomar *algún* valor, y los eventos $X = xj$ están disjuntos, entonces

$$\sum_{j=1}^{\infty} P(X = x_j) = P\left(\bigcup_{j=1}^{\infty} \{X = x_j\}\right) = P(X = x_1 \text{ ó } X = x_2 \text{ ó } \ldots) = 1. \qquad (24)$$

1.4.1. Distribución de Bernoulli

Una variable aleatoria tiene la *distribución de Bernoulli* con un parámetro p si $P(X = 1) = p$ y $P(X = 0 = 1 - p)$, cuando $0 < p < 1$. Se escribe como $X \sim Bern(p)$, el símbolo $\sim$ significa "distribuido como" y la probabilidad p es el *parámetro*, que determina qué distribución de Bernoulli específica tenemos.

Supóngase que se realizan n procesos/ensayos/pruebas(trials) Bernoulli independientes, cada uno con probabilidad p de éxito. X sea el número de éxitos, la distribución X se llama *distribución binominal* con parámetros n y p; se escribe $X \sim Bin(p, n)$. $Bern(p)$ es la misma distribución que $Bin(1, p)$. Bernoulli es un caso especial de binominal, si $x \sim Bin(1, p)$, entonces la función de probabilidad de X es

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{25}$$

para $k = 0, 1, \ldots, n$ (y por otra parte $P(X = k) = 0$).

Maybe[BH19]

Fuentes

[BH19] Joseph K. Blitzstein y Jessica Hwang. *Introduction to Probability [Introducción a la probabilidad]*. 2.ª ed. Texts in Statistical Science [Textos en ciencia estadistica]. Florida: Chapman y Hall/cRC, 2019. ISBN: 9781138369917.

[Dat12] C. J. Date. *SQL and Relational Theory: How to Write Accurate SQL Code [SQL y teoria relacional: Cómo escribir código SQL correcto]*. 2.ª ed. California: 0'Reilly Media, 2012. ISBN: 9781449316402.

[WMIS09] Dennis D Wackerly, william Mendenhall III y Richard L Scheaffer. *Estadística matemática con aplicaciones*. Trad. por Jorge Humberto Romo Mufioz. 7.ª ed. Ciudad de México: Cenage Learning, 2009. ISBN: 9780495110811.