Capítulo 11 Modelos lineales y estimación por mínimos cuadrados

11.1 Introducción

En todas nuestras explicaciones anteriores acerca de la inferencia estadística supusimos que las variables aleatorias $Y_1, Y_2, \ldots, Y_n$, eran independientes y distribuidas idénticamente. Esto implica que el valor esperado de Y_i es constante, o en otras palabras $E(Y_i) = \mu$ no dependen de l valor de ninguna otra variables. En la realidad, esta suposición no es válida en muchos problemas inferencias.

En este capítulo se estudiaran los procedimientos inferenciales que se pueden usar cuando una variable aleatoria Y, llamada *variable dependiente*, tiene una media que es función de una o más variables no aleatorias $x_1, x_2, \ldots, x_k$, llamadas *variables independientes*.

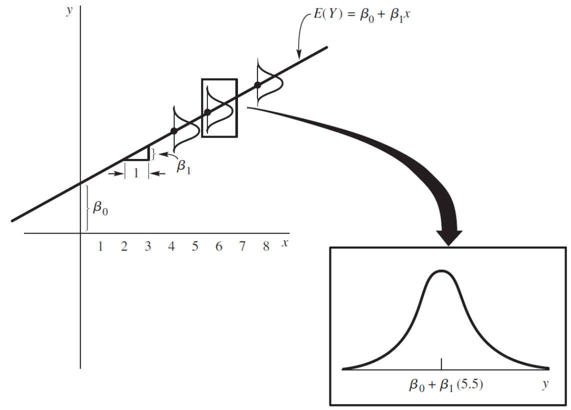
Muchos tipos diferentes de funciones matemáticas se pueden usar para modelar una respuesta que sea una función de una o más variables independientes. Éstas se pueden clasificar en dos categorías: modelos determinísticos y probabilísticos. Un modelo matemático *determinístico* no toma en cuenta ningún error para predecir o pronosticar y como función de x. Un ejemplo puede ser un modelo donde y se relaciona con x de la siguiente manera, $y = \beta_0 + \beta_1 x$. El modelo determinístico no es una representación exacta de la relación entre las dos variables.

En contraste con el modelo determinístico, los expertos en estadística usan modelos *probabilísticos*. Un ejemplo seria modelar el valor esperado de Y de la siguiente manera, $E(Y) = \beta_0 + \beta_1 x$ o bien, lo que es equivalente,

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

donde ε es una variable aleatorias que tiene una distribución de probabilidad específica con media 0. Consideramos a Y como la suma de un componente determinístico $E(Y)$ y un componente aleatorio ε. Este tipo de modelo da una descripción más precisa de la realidad que el modelo determinístico.

En la siguiente figura se puede ver una representación gráfica del modelo probabilístico $Y = \beta_0 + \beta_1 x + \varepsilon$. Cuando $x = 5.5$, hay una *población* de posibles valores de Y. La distribución de esta población está indicada en la parte principal de la gráfica y está centrada en la recta $E(Y) = \beta_0 + \beta_1 x$ en el punto $x = 5.5$. Esta población tiene una distribución con media $\beta_0 + \beta_1(5.5)$ y varianza σ^2, como se muestra en la versión amplificada de la distribución que está encerrada en el recuadro.



Esto es, en un modelo de regresión, existe una *población separada* de valores de respuesta para cada posible ajuste de la(s) variable(s) independiente(s). Todas estas poblaciones tienen la misma varianza y la forma de las distribuciones de las poblaciones son iguales; no obstante, la media de cada población depende, mediante el modelo de regresión, del ajuste de la(s) variable(s) independiente(s).

11.2 Modelos estadísticos lineales

Cuando decimos que tenemos un modelo estadístico lineal para Y, queremos decir que $E(Y)$ es una función lineal de los parámetros desconocidos β_0 y β_1 y *no* necesariamente una función lineal de x.

Si el modelo relaciona $E(Y)$ como una función lineal de β_0 y β_1 únicamente, el modelo recibe el nombre de modelo de regresión *simple*. Si más de una variable independiente, por ejemplo $x_1, x_2, \ldots, x_k$, son de interés y modelamos $E(Y)$ con

$$E(Y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

el modelo se denomina modelo de regresión lineal *múltiple*. Como $x_1, x_2, \ldots, x_k$ son consideradas como variables con valores conocidos, se supone que están medidas sin error en un experimento. De manera general se tiene la siguiente definición.

Definición 11.1

Un modelo *estadístico lineal* que relaciona una respuesta aleatoria Y con un conjunto de variables independientes $x_1, x_2, \ldots x_k$ es de la forma

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

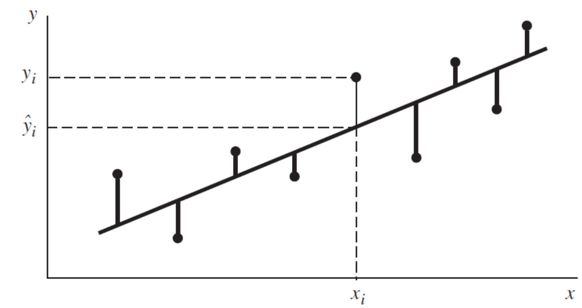
donde $\beta_0, \beta_1, \ldots, \beta_k$ son parámetros desconocidos, ε es una variable aleatoria y las variables $x_1, x_2, \ldots, x_k$ toman valores conocidos. Supondremos que $E(\varepsilon) = 0$ y por lo tanto que

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

11.3 Método de mínimos cuadrados

Un procedimiento para estimar los parámetros de cualquier modelo lineal, el método de mínimos cuadrados, se puede ilustrar con sólo ajustar una recta a un conjunto de puntos. Deseamos ajustar el modelo $E(Y) = \beta_0 + \beta_1 x$, postulamos que $Y = \beta_0 + \beta_1 + \varepsilon$, donde ε tiene alguna distribución de probabilidad con $E(\varepsilon) = 0$. Si $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores de los parámetros β_0 y β_1, entonces $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$ es claramente un estimador de $E(Y)$.

El procedimiento de mínimos cuadrados para ajustar una recta que pase por un conjunto de n puntos consiste en minimizar la suma de cuadrados de las desviaciones verticales a partir de la recta ajustada. Esto se puede ver en la siguiente figura.



donde $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ es el valor pronosticado del i-ésimo valor y (cuando $x = x_i$). Con esto la desviación (a veces llamada *error*) del valor observado y_i a partir de $\hat{y}_i$ es la diferencia $y_i - \hat{y}_i$. Este procedimiento produce estimadores con buenas propiedades y se muestran a continuación.

Estimadores de mínimos cuadrados para el modelo de regresión lineal simple

1. $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, donde $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ y $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$.
2. $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$

Demostración

Como ya se vio, la desviación del valor observado y_i a partir de $\hat{y}_i$ es la diferencia $y_i - \hat{y}_i$ y la suma de los cuadrados de las desviaciones a minimizar es

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right]^2$$

La cantidad SSE también recibe el nombre de *suma de cuadrados del error*.

Demostración

Si la SSE tiene un mínimo, ocurrirá para valores de β_0 y β_1 que satisfagan las ecuaciones $\partial\text{SSE}/\partial\hat{\beta}_0 = 0$ y $\partial\text{SSE}/\partial\hat{\beta}_1 = 0$. Tomando las derivadas parciales de la SSE con respecto a $\hat{\beta}_0$ y $\hat{\beta}_1$ e igualando a cero, obtenemos

$$\begin{aligned}\frac{\partial\text{SSE}}{\partial\hat{\beta}_0} &= \frac{\partial\left\{\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\right\}}{\partial\hat{\beta}_0}\\ &= -\sum_{i=1}^{n} 2[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]\\ &= -2\left(\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_i\right) = 0\end{aligned}$$

y

$$\begin{aligned}\frac{\partial\text{SSE}}{\partial\hat{\beta}_1} &= \frac{\partial\left\{\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\right\}}{\partial\hat{\beta}_1}\\ &= -\sum_{i=1}^{n} 2[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]x_i\\ &= -2\left(\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2\right) = 0\end{aligned}$$

Las ecuaciones $\partial\text{SSE}/\hat{\beta}_0 = 0$ y $\partial\text{SSE}/\hat{\beta}_1 = 0$ se denominan *ecuaciones de mínimos cuadrados* para estimar los parámetros de una recta.
Estas ecuaciones son lineales en $\hat{\beta}_0$ y $\hat{\beta}_1$ y por lo tanto pueden resolverse simultáneamente, despejando los estimadores de los parámetros de cada ecuación. Esto se hace primero para $\hat{\beta}_0$ a continuación.

$$\begin{aligned}-2\left(\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_i\right) &= 0\\ \sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_i &= 0\\ \sum_{i=1}^{n} y_i - \hat{\beta}_1\sum_{i=1}^{n} x_i &= n\hat{\beta}_0\\ \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}_1\frac{1}{n}\sum_{i=1}^{n} x_i &= \hat{\beta}_0\end{aligned}$$

Con esto se tiene que el estimador de mínimos cuadrados para β_0 es $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$. Se realizara el mismo calculo pero ahora utilizando la ecuación de mínimos cuadrados para $\hat{\beta}_1$.

Demostración

$$\begin{aligned}-2\left(\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2\right) &= 0\\ \sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2 &= 0\\ \sum_{i=1}^{n} x_i y_i - (\overline{y} - \hat{\beta}_1\overline{x})\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2 &= 0\\ \sum_{i=1}^{n} x_i y_i - \overline{y}\sum_{i=1}^{n} x_i + \hat{\beta}_1\overline{x}\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2 &= 0\end{aligned}$$

Despejando $\hat{\beta}_1$ de la misma manera en que se despejo $\hat{\beta}_0$ se obtiene lo siguiente.

$$\begin{aligned}\sum_{i=1}^{n} x_i y_i - \overline{y}\sum_{i=1}^{n} x_i &= \hat{\beta}_1\sum_{i=1}^{n} x_i^2 - \hat{\beta}_1\overline{x}\sum_{i=1}^{n} x_i\\ \sum_{i=1}^{n} x_i y_i - n\overline{y}\frac{1}{n}\sum_{i=1}^{n} x_i &= \hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}\frac{1}{n}\sum_{i=1}^{n} x_i\right)\\ \sum_{i=1}^{n} x_i y_i - n\overline{xy} &= \hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)\\ \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}\end{aligned}$$

Se puede demostrar que,

$$\begin{aligned}\sum_{i=1}^{n} x_i y_i - n\overline{xy} &= \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = S_{xy}\\ \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 &= \sum_{i=1}^{n}(x_i - \overline{x})^2 = S_{xx}\end{aligned}$$

Por lo tanto el estimador de mínimos cuadrados para el parámetro β_1 es

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{S_{xy}}{S_{xx}}$$

11.4 Propiedades de los estimadores de mínimos cuadrados: regresión lineal simple

Recuerde que previamente supusimos que ε es una variable aleatoria con $E(\varepsilon) = 0$. Ahora agregaremos la suposición de que $V(\varepsilon) = \sigma^2$. Esto es, estamos suponiendo que la diferencia entre la variable aleatoria Y y $E(Y) = \beta_0 + \beta_1 x$ está distribuida alrededor de cero con una varianza que no depende de x. Observe que $V(Y) = V(\varepsilon) = \sigma^2$ porque los otros términos del modelo lineal son constante.

Propiedades de los estimadores de mínimos cuadrados; regresión lineal simple

1. Los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ son insesgados, es decir, $E(\hat{\beta}_i) = \beta_i$, para $i = 0, 1$.
2. $V(\hat{\beta}_0) = c_{00}\sigma^2$, donde $c_{00} = \sum x_i^2/(nS_{xx})$.
3. $V(\hat{\beta}_1) = c_{11}\sigma^2$, donde $c_{11} = \frac{1}{S_{xx}}$.
4. $Cov(\hat{\beta}_0, \hat{\beta}_1) = c_{01}\sigma^2$, donde $c_{01} = \frac{-\overline{x}}{S_{xx}}$.
5. Un estimador insesgado de σ^2 es $S^2 = \text{SSE}/(n-2)$, donde $\text{SSE} = S_{yy} - \hat{\beta}_1 S_{xy}$ y $S_{yy} = \sum(y_i - \overline{y})^2$.

Si, además, el ε_i, para $i = 1, 2, \ldots, n$ está distribuido normalmente,

6. $\hat{\beta}_0$ y $\hat{\beta}_1$ están distribuidas normalmente.
7. La variable aleatoria $\frac{(n-2)S^2}{\sigma^2}$ tiene una distribución χ^2 con $n-2$ grados de libertad.
8. El estadístico S^2 es independiente de $\hat{\beta}_0$ y $\hat{\beta}_1$.

El resto de este capítulo se concentrara en demostrar las propiedades de los estimadores de mínimos cuadrados. Empezando por la primer propiedad.

Demostración de la primer propiedad

De la sección pasada sabemos que los estimadores de mínimos cuadrados son los siguientes.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{S_{xy}}{S_{xx}} \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$$

Primero se empezara por el estimador $\hat{\beta}_1$.

Demostración de la primer propiedad

$$
\begin{aligned}
E(\hat{\beta}_1) &= E\left[\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right] \\
&= \frac{1}{S_{xx}}E\left[\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})\right] \\
&= \frac{1}{S_{xx}}E\left[\sum_{i=1}^{n}[(x_i-\overline{x})y_i-(x_i-\overline{x})\overline{y}]\right] \\
&= \frac{1}{S_{xx}}E\left[\sum_{i=1}^{n}(x_i-\overline{x})y_i-\overline{y}\sum_{i=1}^{n}(x_i-\overline{x})\right]
\end{aligned}
$$

Se puede demostrar de manera sencilla que $\sum_{i=1}^{n}(x_i-\overline{x}) = 0$ al desarrollar la suma. Por lo tanto se tiene lo siguiente.

$$
\begin{aligned}
E(\hat{\beta}_1) &= \frac{1}{S_{xx}}E\left[\sum_{i=1}^{n}(x_i-\overline{x})y_i\right] \\
&= \frac{1}{S_{xx}}\sum_{i=1}^{n}E\left[(x_i-\overline{x})y_i\right] \\
&= \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})E\left[y_i\right]
\end{aligned}
$$

Sabemos que $Y_i = \beta_0+\beta_1 x_i+\varepsilon_i$ y por lo tanto $E(Y_i) = \beta_0+\beta_1 x_i$. Entonces se tiene lo siguiente,

$$
\begin{aligned}
E(\hat{\beta}_1) &= \frac{1}{S_{xx}}\sum_{i=1}^{n}(x_i-\overline{x})(\beta_0+\beta_1 x_i) \\
&= \frac{1}{S_{xx}}\left[\sum_{i=1}^{n}[(x_i-\overline{x})\beta_0+\beta_1(x_i-\overline{x})x_i]\right] \\
&= \frac{1}{S_{xx}}\left[\beta_0\sum_{i=1}^{n}(x_i-\overline{x})+\beta_1\sum_{i=1}^{n}(x_i-\overline{x})x_i\right]
\end{aligned}
$$

Como ya se sabe, $\sum_{i=1}^{n}(x_i-\overline{x}) = 0$. Además se puede demostrar que $S_{xx} = \sum(x_i-\overline{x})^2 = \sum(x_i-\overline{x})x_i$ desarrollando la suma. Por lo tanto se tiene lo siguiente.

$$
E(\hat{\beta}_1) = \frac{1}{S_{xx}}\left[\beta_1 S_{xx}\right] = \beta_1
$$

Demostración de la primer propiedad

Ahora se realizara la demostración para $\hat{\beta}_0$ usando la demostración para el estimador $\hat{\beta}_1$.

$$
\begin{aligned}
E(\hat{\beta}_0) &= E[\overline{Y}-\hat{\beta}_1\overline{x}] = E[\overline{Y}]-E[\hat{\beta}_1\overline{x}] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}Y_i\right]-\overline{x}E[\hat{\beta}_1] \\
&= \frac{1}{n}\sum_{i=1}^{n}E[Y_i]-\overline{x}\beta_1 \\
&= \frac{1}{n}\sum_{i=1}^{n}(\beta_0+\beta_1 x_i)-\overline{x}\beta_1 \\
&= \frac{1}{n}\sum_{i=1}^{n}\beta_0+\frac{1}{n}\sum_{i=1}^{n}\beta_1 x_i-\overline{x}\beta_1 \\
&= \frac{1}{n}n\beta_0+\beta_1\frac{1}{n}\sum_{i=1}^{n}x_i-\overline{x}\beta_1 \\
&= \beta_0
\end{aligned}
$$

Demostración de la segunda y tercer propiedad

Se empezara por la varianza del estimador $\hat{\beta}_1$. Usando el Teorema 5.12 y el hecho de que $Y_1,\ldots,Y_n$ son independientes se tiene lo siguiente.

$$
\begin{aligned}
V(\hat{\beta}_1) &= V\left(\frac{\sum_{i=1}^{n}(x_i-\overline{x})Y_i}{S_{xx}}\right) \\
&= \left[\frac{1}{S_{xx}}\right]^2\sum_{i=1}^{n}V[(x_i-\overline{x})]Y_i \\
&= \frac{1}{S_{xx}^2}\sum_{i=1}^{n}(x_i-\overline{x})^2 V(Y_i)
\end{aligned}
$$

Como $V(Y_i) = \sigma^2$, para $i = 1,2,\ldots,n$. Entonces se tiene,

$$
V(\hat{\beta}_1) = \frac{1}{S_{xx}^2}S_{xx}\sigma^2 = \frac{\sigma^2}{S_{xx}}
$$

Ahora se obtendrá la varianza del estimador $\hat{\beta}_0$. Esta esta dada por lo siguiente.

$$
\begin{aligned}
V(\hat{\beta}_0) &= V(\overline{Y}-\hat{\beta}_1\overline{x}) \\
&= V(\overline{Y})+V(\hat{\beta}_1\overline{x})-2Cov(\overline{Y},\hat{\beta}_1\overline{x}) \\
&= V(\overline{Y})+\overline{x}^2V(\hat{\beta}_1)-2\overline{x}Cov(\overline{Y},\hat{\beta}_1)
\end{aligned}
$$

Esto se obtiene del Teorema 5.12 y resultados derivados de este teorema.

Demostración de la segunda y tercer propiedad

Ahora se desarrollara cada varianza y covarianza del resultado anterior individualmente para facilitar los cálculos. Empezamos por $V(\overline{Y})$

$$
V(\overline{Y}) = V\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}V(Y_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2
$$

Lo que resulta en $V(\overline{Y}) = \frac{\sigma^2}{n}$. Ahora se va a encontrar $Cov(\overline{Y},\hat{\beta}_1)$, para hacer esto se necesita reescribir a $\hat{\beta}_1$ como,

$$
\hat{\beta}_1 = \sum_{i=1}^{n}c_iY_i \qquad \text{donde} \qquad c_i = \frac{x_i-\overline{x}}{S_{xx}}
$$

Observamos que $\sum c_i = 1$. Entonces, la covarianza se puede escribir de la siguiente manera.

$$
Cov(\overline{Y},\hat{\beta}_1) = Cov\left[\frac{1}{n}\sum_{i=1}^{n}Y_i,\sum_{i=1}^{n}c_iY_i\right]
$$

Usando el Teorema 5.12 se tiene,

$$
Cov(\overline{Y},\hat{\beta}_1) = \sum_{i=1}^{n}\left(\frac{c_i}{n}\right)V(Y_i)+\sum_{i\neq j}\sum\left(\frac{c_j}{n}\right)Cov(Y_i,Y_j)
$$

Como Y_i y Y_j, donde $i \neq j$, son independientes, $Cov(Y_i,Y_j) = 0$. También, $V(Y_i) = \sigma^2$ y por lo tanto,

$$
Cov(\overline{Y},\hat{\beta}_1) = \frac{\sigma^2}{n}\sum_{i=1}^{1}c_i = \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{S_{xx}}\right) = 0
$$

Ahora se sustituyen estos valores en la varianza de $\hat{\beta}_0$ para obtener lo siguiente.

$$
\begin{aligned}
V(\hat{\beta}_0) &= V(\overline{Y})+\overline{x}^2V(\hat{\beta}_1)-2\overline{x}Cov(\overline{Y},\hat{\beta}_1) \\
&= \frac{\sigma^2}{n}+\overline{x}^2\left(\frac{\sigma^2}{S_{xx}}\right) \\
&= \sigma^2\left(\frac{1}{n}+\frac{\overline{x}^2}{S_{xx}}\right)
\end{aligned}
$$

Demostración de la cuarta propiedad

Ahora se demostrara la tercer propiedad en la que se encuentra la covarianza entre los dos estimadores, $\hat{\beta}_1$ y $\hat{\beta}_0$.

$$\begin{aligned}
Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\overline{Y} - \hat{\beta}_1\overline{x}, \hat{\beta}_0) \\
&= Cov(\overline{Y}, \hat{\beta}_1) - Cov(\hat{\beta}_1\overline{x}, \hat{\beta}_1) \\
&= Cov(\overline{Y}, \hat{\beta}_1) - \overline{x}Cov(\hat{\beta}_1, \hat{\beta}_1) \\
&= Cov(\overline{Y}, \hat{\beta}_1) - \overline{x}Var(\hat{\beta}_1) \\
&= 0 - \overline{x}\frac{\sigma^2}{S_{xx}} \\
&= \sigma^2\frac{-\overline{x}}{S_{xx}}
\end{aligned}$$

Esto se logra con el uso del Teorema 5.12 y resultados derivados de este. Observamos que los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ están correlacionados y por lo tanto son dependientes a menos que $\overline{x} = 0$.

Las expresiones anteriores dan las varianzas para los estimadores de mínimos cuadrados en términos de σ^2, la varianza del término de error ε. Por lo general el valor de σ^2 es desconocido y necesitaremos hacer uso de observaciones muestrales para estimar σ^2. Debido a que ahora estamos usando $\hat{Y}_i$ para calcular $E(Y_i)$, parece natural basar una estimación de σ^2 en SSE $= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$.

Demostración de la quinta propiedad

Como se esta basando la estimación de σ^2 en la suma de cuadrados del error entonces el estimador insesgado propuesto es el siguiente.

$$S^2 = \left(\frac{1}{n-2}\right)\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \left(\frac{1}{n-2}\right)\text{SSE}$$

Para demostrar que es un estimador insesgado se necesita demostrar que $E(S^2) = \sigma^2$. Vemos que

$$E(S^2) = E\left[\left(\frac{1}{n-2}\right)\text{SSE}\right] = \left(\frac{1}{n-2}\right)E(\text{SSE})$$

por lo que nos necesitamos concentrar en encontrar $E(\text{SSE})$. Esto se hace a continuación.

$$\begin{aligned}
E(\text{SSE}) &= E\left[\sum (Y_i - \hat{Y}_i)^2\right] \\
&= E\left[\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right] \\
&= E\left[\sum (Y_i - \overline{Y} + \hat{\beta}_1\overline{x} - \hat{\beta}_1 x_i)^2\right]
\end{aligned}$$

Demostración de la quinta propiedad

$$\begin{aligned}
E(\text{SSE}) &= E\left[\sum [(Y_i - \overline{Y}) - \hat{\beta}_1(x_i - \overline{x})]^2\right] \\
&= E\Big[\sum (Y_i - \overline{Y})^2 + \hat{\beta}_1^2 \sum (x_i - \overline{x})^2 \\
&\quad - 2\hat{\beta}_1 \sum (x_i - \overline{x})(Y_i - \overline{Y})\Big]
\end{aligned}$$

Sabemos que $\sum (x_i - \overline{x})(Y_i - \overline{Y}) = S_{xy}$ de la sección pasada. Además se puede demostrar que $S_{xy} = \sum (x_i - \overline{x})^2\hat{\beta}_1$ al desarrollar la suma. También se puede demostrar que $\sum (Y_i - \overline{Y})^2 = \sum Y_i^2 - n\overline{Y}^2$. Con esto se tiene lo siguiente.

$$\begin{aligned}
E(\text{SSE}) &= E\Big[\sum Y_i^2 - n\overline{Y}^2 + \hat{\beta}_1^2 \sum (x_i - \overline{x})^2 \\
&\quad -2\hat{\beta}_1^2 \sum (x_i - \overline{x})^2\Big] \\
&= E\left[\sum Y_i^2 - n\overline{Y}^2 - \hat{\beta}_1^2 S_{xx}\right] \\
&= \sum E(Y_i^2) - nE(\overline{Y}^2) - S_{xx}E(\hat{\beta}_1^2)
\end{aligned}$$

Lo anterior se puede simplificar al observar que para cualquier variable aleatoria U, $E(U^2) = V(U) + [E(U)]^2$. Entonces se tiene lo siguiente.

$$\begin{aligned}
E(\text{SSE}) &= \sum \left\{V(Y_i) + [E(Y_i)]^2\right\} - n\left\{V(\overline{Y}) + \right. \\
&\quad \left. [E(\overline{Y})]^2\right\} - S_{xx}\left\{V(\hat{\beta}_1) + [E(\hat{\beta}_1)]^2\right\} \\
&= \sum \left\{\sigma^2 + [\beta_0 + \beta_1 x_i]^2\right\} - n\left\{\frac{\sigma^2}{n} + \right. \\
&\quad \left. [\beta_0 + \beta_1\overline{x}]^2\right\} - S_{xx}\left\{\frac{\sigma^2}{S_{xx}} + \beta_1^2\right\} \\
&= n\sigma^2 + \sum [\beta_0 + \beta_1 x_i]^2 - \sigma^2 - n[\beta_0 + \beta_1\overline{x}]^2 \\
&\quad - \sigma^2 - S_{xx}\beta_1^2 \\
&= n\sigma^2 + \sum [\beta_0^2 + \beta_1^2 x_i^2 + 2\beta_0\beta_1 x_i] - \sigma^2 \\
&\quad - n[\beta_0^2 + \beta_1^2\overline{x}^2 + 2\beta_0\beta_1\overline{x}] - \sigma^2 - S_{xx}\beta_1^2 \\
&= n\sigma^2 + \sum \beta_0^2 + \sum \beta_1^2 x_i^2 + \sum 2\beta_0\beta_1 x_i - \sigma^2 \\
&\quad - n[\beta_0^2 + \beta_1^2\overline{x}^2 + 2\beta_0\beta_1\overline{x}] - \sigma^2 - S_{xx}\beta_1^2 \\
&= n\sigma^2 + n\beta_0^2 + \beta_1^2 \sum x_i^2 + 2n\beta_0\beta_1\frac{1}{n}\sum x_i \\
&\quad - \sigma^2 - n\beta_0^2 - \beta_1^2 n\overline{x}^2 - 2n\beta_0\beta_1\overline{x} - \sigma^2 - S_{xx}\beta_1^2
\end{aligned}$$

Demostración de la quinta propiedad

$$\begin{aligned}
E(\text{SSE}) &= n\sigma^2 + n\beta_0^2 + \beta_1^2 \sum x_i^2 + 2n\beta_0\beta_1\overline{x} \\
&\quad - \sigma^2 - n\beta_0^2 - \beta_1^2 n\overline{x}^2 - 2n\beta_0\beta_1\overline{x} - \sigma^2 - S_{xx}\beta_1^2 \\
&= n\sigma^2 + \beta_1^2 \sum x_i^2 - \sigma^2 - \beta_1^2 n\overline{x}^2 - \sigma^2 - S_{xx}\beta_1^2 \\
&= n\sigma^2 - 2\sigma^2 + \beta_1^2 \sum x_i^2 - \beta_1^2 n\overline{x}^2 - \beta_1^2 S_{xx} \\
&= (n-2)\sigma^2 + \beta_1^2\left(\sum x_i^2 - n\overline{x}^2\right) - \beta_1^2 S_{xx} \\
&= (n-2)\sigma^2 + \beta_1^2 S_{xx} - \beta_1^2 S_{xx} \\
&= (n-2)\sigma^2
\end{aligned}$$

Sustituyendo este resultado en $E(S^2)$ se tiene lo siguiente,

$$\begin{aligned}
E(S^2) &= E\left[\left(\frac{1}{n-2}\right)\text{SSE}\right] = \left(\frac{1}{n-2}\right)E(\text{SSE}) \\
&= \left(\frac{1}{n-2}\right)(n-2)\sigma^2 = \sigma^2
\end{aligned}$$

Del resultado anterior se vio una demostración de como escribir la suma de cuadrados del error de una manera más sencilla. Esto es,

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \overline{y})^2 - \hat{\beta}_1\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = S_{yy} - \hat{\beta}_1 S_{xy}$$

Este resultado se obtiene de manera similar al método utilizado en la demostración de la quinta propiedad por lo que no se incluirá aquí.

Las últimas tres propiedades no se demostrarán aquí. Solo se explicara su deducción.

Como la forma de las distribuciones muestrales para $\hat{\beta}_0$ y $\hat{\beta}_1$ dependen de la distribución del término de error ε. Debido a que con frecuencia se presenta la distribución normal en la naturaleza entonces es razonable suponer que ε está distribuido normalmente con media 0 y varianza σ^2. Si se garantiza esta suposición de normalidad, se deduce que Y_i está distribuida normalmente con media $\beta_0 + \beta_1 x_i$ y varianza σ^2. Entonces, como $\hat{\beta}_0$ y $\hat{\beta}_1$ son *funciones lineales* de $Y_1, Y_2, \ldots, Y_n$, los estimadores están distribuidos normalmente, con las medias y varianzas determinadas previamente. Además si se garantiza la suposición de normalidad, se puede demostrar que,

$$\frac{(n-2)S^2}{\sigma^2} = \frac{\text{SSE}}{\sigma^2}$$

tiene una distribución χ^2 con $n-2$ grados de libertad.

11.5 Inferencias respecto a los parámetros β_i

De la sección pasada sabemos que los estimadores son insesgados y tienen distribución normal con las varianzas calculadas en tal sección. Entonces se puede construir una prueba de la hipótesis $H_0 : \beta_i = \beta_{i0}$, donde β_{i0} es un valor específico de β_i, usando el estadístico de prueba $Z = \dfrac{\hat{\beta}_i - \beta_{i0}}{\sigma\sqrt{c_{ii}}}$ donde $c_{00} = \dfrac{\sum x_i^2}{nS_{xx}}$ y $c_{11} = \dfrac{1}{S_{xx}}$

Sin embargo para esto se necesitaría conocer σ o tener una buena estimación basada en un número adecuado de grados de libertad. Por lo general no se tiene esta información por lo que se necesita calcular un estimación a partir de los datos particulares. Se puede utilizar S^2, calculada en la sección anterior, para estimar σ^2 y demostrar que

$$T = \frac{\hat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}$$

tiene una distribución t de Student con $n-2$ grados de libertad. De esto se puede desarrollar una prueba de hipótesis para los parámetros.

Prueba de hipótesis para β_i

$H_0 : \beta_i = \beta_{i0}$.

$$H_a : \begin{cases} \beta_i > \beta_{i0} & \text{(región de rechazo de cola superior)} \\ \beta_i < \beta_{i0} & \text{(región de rechazo de cola inferior)} \\ \beta_i \neq \beta_{i0} & \text{(región de rechazo de dos colas)} \end{cases}$$

Estadístico de prueba: $T = \dfrac{\hat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}$.

$$\text{RR} : \begin{cases} t > t_\alpha & \text{(alternativa de cola superior)} \\ t < -t_\alpha & \text{(alternativa de cola inferior)} \\ |t| > t_{\alpha/2} & \text{(alternativa de dos colas)} \end{cases}$$

donde

$c_{00} = \dfrac{\sum x_i^2}{nS_{xx}}$ y $c_{11} = \dfrac{1}{S_{xx}}$

Observe que t_α está basada en $(n-2)$ grados de libertad.

Como se explico en el capítulo anterior, se puede desarrollar un intervalo de confianza con el mismo estadístico de prueba.

Un intervalo de confianza $100(1-\alpha)$ para β_i

$$\hat{\beta}_i \pm t_{\alpha/2} S\sqrt{c_{ii}}$$

donde

$$c_{00} = \frac{\sum x_i^2}{nS_{xx}} \qquad \text{y} \qquad c_{11} = \frac{1}{S_{xx}}$$

11.6 Inferencias respecto a funciones lineales de los parámetros del modelo: regresión lineal simple

Además de realizar inferencias acerca de una β_i individual a menudo nos interesa realizar inferencias acerca de funciones lineales de los parámetros del modelo β_0 y β_1. En particular $E(Y) = \beta_0 + \beta_1 x$.

Suponemos que se desea hacer una inferencia acerca de la función lineal $\theta = a_0\beta_0 + a_1\beta_1$, donde a_0 y a_1 son constantes. Entonces

$$\hat{\theta} = a_0\hat{\beta}_0 + a_1\hat{\beta}_1$$

es un estimador insesgados de θ. Esto se puede demostrar fácilmente con el uso del Teorema 5.12,

$$E(\hat{\theta}) = a_0 E(\hat{\beta}_0) + a_1 E(\hat{\beta}_1) = a_0\beta_0 + a_1\beta_1 = \theta$$

Aplicando el mismo teorema se puede determinar que la varianza de $\hat{\theta}$ es

$$V(\hat{\theta}) = a_0^2 V(\hat{\beta}_0) + a_1^2 V(\hat{\beta}_1) + 2a_0a_1 Cov(\hat{\beta}_0, \hat{\beta}_1)$$

donde $V(\hat{\beta}_i) = c_{ii}\sigma^2$ y $Cov(\hat{\beta}_0, \hat{\beta}_1) = c_{01}\sigma^2$, con

$$c_{00} = \frac{\sum x_i^2}{nS_{xx}} \qquad c_{11} = \frac{1}{S_{xx}} \qquad c_{01} = \frac{-\overline{x}}{S_{xx}}$$

Desarrollando los términos se obtiene lo siguiente,

$$V(\hat{\theta}) = \left(\frac{a_0 \dfrac{\sum x_i^2}{n} + a_1^2 - 2a_0a_1\overline{x}}{S_{xx}} \right) \sigma^2$$

Ya que los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ están distribuidas normalmente entonces el estimador $\hat{\theta}$ tiene distribución normal. Se puede utilizar el estadístico Z para una prueba de hipótesis acerca del valor de θ. Sin embargo se llega al mismo problema que en la sección pasada en el cual no se conoce el valor de σ^2. Entonces se puede utilizar el estimador insesgado S^2, definido en la sección 11.4, para desarrollar una prueba de hipótesis acerca de θ.

Una prueba para $\beta = a_0\beta_0 + a_1\beta_1$

$H_0 : \theta = \theta_0$

$$H_a : \begin{cases} \theta > \theta_0 \\ \theta < \theta_0 \\ \theta \neq \theta_0 \end{cases}$$

Estadístico de prueba: $T = \dfrac{\hat{\theta} - \theta_0}{S\sqrt{\left(\dfrac{a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2a_0a_1\overline{x}}{S_{xx}} \right)}}$

Una prueba para $\theta = a_0\beta_0 + a_1\beta_1$

$$\text{Región de Rechazo} : \begin{cases} t > t_\alpha \\ t < -t_\alpha \\ |t| > t_{\alpha/2} \end{cases}$$

Aquí t_α y $t_{\alpha/2}$ están basados en $n-2$ grados de libertad.

Un intervalo de confianza $100(1-\alpha)\%$ para $\theta = a_0\beta_0 + a_1\beta_1$

$$\hat{\theta} \pm t_{\alpha/2} S \sqrt{\left(\frac{a_0^2 \dfrac{\sum x_i^2}{n} + a_1^2 - 2a_0a_1\overline{x}}{S_{xx}} \right)}$$

donde la $t_{\alpha/2}$ tabulada está basada en $n-2$ grados de libertad.

Una aplicación útil de las técnicas de prueba de hipótesis e intervalo de confianza que se acaban de presentar es al estimar $E(Y)$ para un valor fijo de la variable independiente x.

Si x^* denota un valor específico de x que es de interés, entonces $E(Y) = \beta_0 + \beta_1 x^*$. Esto es un caso especial de θ con $a_0 = 1$ y $a_1 = x^*$. Por lo tanto la varianza de su estimador es la siguiente,

$$V(\hat{\theta}) = \left(\frac{a_0^2 \dfrac{\sum x_i^2}{n} + a_1^2 - 2a_0a_1\overline{x}}{S_{xx}} \right) \sigma^2 = \left(\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}} \right) \sigma^2$$

Se puede usar el estimador S^2 y de tal manera obtener el siguiente intervalo,

Un intervalo de confianza $100(1-\alpha)\%$ para $E(Y) = \beta_0 + \beta_1 x^*$

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$$

donde la $t_{\alpha/2}$ tabulada está basada en $n-2$ grados de libertad.

11.7 Predicción de un valor particular de Y mediante regresión lineal simple

En la sección anterior se estudio un método para estimar el rendimiento medio $E(Y)$ del proceso en el ajuste $x = x^*$. Ahora consideramos un problema diferente. En lugar de calcular el rendimiento medio en x^*, deseamos *predecir* la respuesta particular Y que observaremos si el experimento se ejecuta en el futuro.

De la sección pasada se vio como si es razonable suponer que ε está distribuido normalmente con media 0 y varianza σ^2 entonces se deduce que Y está distribuida normalmente con media $\beta_0 + \beta_1 x$ y varianza σ^2. Si estamos interesados en el valor de Y cuando $x = x^*$, emplearíamos $\widehat{Y^*} = \hat{\beta}_0 + \hat{\beta}_1 x^*$ como pronosticador de un valor particular de Y^* y también como estimador de $E(Y)$.

Si $x = x^*$, el error de predecir un valor particular de Y^*, usando $\widehat{Y^*}$ como el pronosticador, es la diferencia entre el valor real de Y^* y el valor pronosticado: error $= Y^* - \widehat{Y^*}$. Como Y^* y $\widehat{Y^*}$ son variables aleatorias distribuidas normalmente, su diferencia también está distribuida normalmente. Este error tiene la siguiente esperanza,

$$E(\text{error}) = E(Y^* - \widehat{Y^*}) = E(Y^*) - E(\widehat{Y^*})$$

y como $E(\widehat{Y^*}) = \beta_0 + \beta_1 x^* = E(Y^*)$, entonces $E(\text{error}) = 0$. Del mismo modo,

$$V(\text{error}) = V(Y^* - \widehat{Y^*}) = V(Y^*) + V(\widehat{Y^*}) - 2Cov(Y^*, \widehat{Y^*})$$

Debido a que estamos pronosticando un valor futuro de Y^* que no se emplea en el cálculo de $\widehat{Y^*}$, se deduce que Y^* y $\widehat{Y^*}$ son independientes y por lo tanto $Cov(Y^*, \widehat{Y^*}) = 0$. Entonces se tiene,

$$\begin{aligned} V(\text{error}) &= V(Y^*) + V(\widehat{Y^*}) = \sigma^2 + V(\hat{\beta}_0 + \hat{\beta}_1 x^*) \\ &= \sigma^2 + \left(\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}} \right) \sigma^2 \\ &= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}} \right] \end{aligned}$$

Como el error de pronosticar un valor particular de Y está distribuido normalmente con media 0 y varianza ya demostrada se deduce que,

$$Z = \frac{Y^* - \widehat{Y^*}}{\sigma \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}}$$

tiene una distribución normal estándar. Además, si S se sustituye por σ, se puede demostrar que

$$T = \frac{Y^* - \widehat{Y^*}}{S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}}$$

tiene distribución t de Student con $n-2$ grados de libertad. Con esto se puede construir un *intervalo de predicción* para la variable aleatoria Y^* usando métodos similares a los presentados en capítulo anteriores. Entonces un intervalo alrededor de $\widehat{Y^*}$ que en muestreo repetido contendrá el valor real de Y^* con probabilidad $1 - \alpha$ es

Intervalo de predicción de 100$(1-\alpha)$% para Y cuando $x = x^*$

$$\boldsymbol{\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}}$$

De acuerdo al resultado de la sección anterior se puede observar que los *intervalos de predicción* para el valor real de Y son más largos que los *intervalos de confianza* para $E(Y)$ si ambos están determinados por el mismo valor de x^*. Esto se debe al valor de la varianza del error.

11.8 Correlación

Los modelos vistos hasta ahorita son útiles en dos situaciones prácticas distintas. En la primera, la variable x puede ser controlada completamente por el experimentador. Desde luego, x podría variar de un experimento a otro, pero bajo el completo control del experimentador. Entonces se tiene que,

$$Y = \beta_0 + \beta_1 x + \epsilon \qquad \Rightarrow \qquad E(Y) = \beta_0 + \beta_1 x$$

En el segundo, la variable x puede ser un valor observado de una variable aleatoria X. Para esta situación usamos el modelo,

$$Y = \beta_0 + \beta_1 x + \epsilon \qquad \Rightarrow \qquad E(Y|X = x) = \beta_0 + \beta_1 x$$

Estamos suponiendo que la esperanza condiciona de Y para un valor fijo de X es una función lineal del valor x. Por lo general suponemos que la variable aleatoria vectorial (X, Y) tiene una distribución normal bivariable con $E(X) = \mu_X$, $E(Y) = \mu_Y$, $V(X) = \sigma_X^2$, $V(Y) = \sigma_Y^2$ y el coeficiente de correlación ρ, en cuyo caso se puede demostrar que,

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

La teoría para hacer inferencias acerca de los parámetros β_0 y β_1 es exactamente igual para estos dos casos.

Para el caso donde (X, Y) tiene una distribución bivariable, no siempre estamos interesados en la relación lineal de las variables aleatorias. En ocasiones nos interesa si las variables aleatorias X y Y son independientes, esto es equivalente a probar que el coeficiente de ρ es igual a cero.

Si denotamos con $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ una muestra aleatoria de una distribución normal bivariante entonces el estimador de máxima probabilidad de ρ está dado por el coeficiente de correlación muestral

$$r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \hat{\beta}_1 \sqrt{\frac{S_{xx}}{S_{yy}}}$$

Entonces una prueba para $H_0 : \rho = 0$ es equivalente a la prueba para $H_0 : \beta_1 = 0$, por lo que se puede ocupar el mismo estadístico de prueba al igual que las mismas regiones de rechazo para el caso que se necesita de acuerdo a la región de rechazo elegida. Además este estadístico se puede reescribir en términos de r como sigue,

$$T = \frac{\hat{\beta}_1 - 0}{S/\sqrt{S_{xx}}} = \frac{r\sqrt{n-2}}{\sqrt{1 - r^2}}$$

Este estadístico tiene distribución t con $n - 2$ grados de libertad. A pesar de que se puede utilizar esta prueba para el caso en que X y Y son independientes no es el caso para probar casos generales acerca de ρ. Además no se puede ocupar el estadístico r ya que su distribución de probabilidad es difícil de obtener.

Si la muestra es lo suficientemente grande, se tiene que $(1/2) \ln [1 + r)/(1 - r)]$ está distribuida normalmente en forma aproximada con media $(1/2) \ln [(1 + \rho)/(1 - \rho)]$ y varianza $1/(n - 3)$. Entonces, para probar la hipótesis $H_0 : \rho = \rho_0$ se puede ocupar una prueba Z con el estadístico de prueba,

$$Z = \frac{\left(\frac{1}{2}\right) \ln \left(\frac{1+r}{1-r}\right) - \left(\frac{1}{2}\right) \ln \left(\frac{1+\rho_0}{1-\rho_0}\right)}{\frac{1}{\sqrt{n-3}}}$$

Al usar Z se tienen las mismas hipótesis alternativas y regiones de rechazo que se mostraron en el capítulo anterior ya que Z tiene distribución normal estándar. Esto es,

$$H_a : \begin{cases} \rho > \rho_0 & \text{RR} : z > z_\alpha \\ \rho < \rho_0 & \text{RR} : z < -z_\alpha \\ \rho \neq \rho_0 & \text{RR} : |z| > z_{\alpha/2} \end{cases}$$

Una cantidad que nos puede ser de interés es r^2, denominado *coeficiente de determinación*. Si el modelo de regresión lineal simple se ajusta bien a los datos, las diferencias entre los valores observados y pronosticados son pequeñas, lo cual lleva a un valor pequeño para la suma de cuadrados del error (SSE). De manera alterna, si el modelo de regresión se ajusta mal la SSE será grande. De la sección 11.4 se obtuvo una manera alterna de escribir SSE, esta es

$$\text{SSE} = S_{yy} - \hat{\beta}_1 S_{xy}$$

S_{yy} proporciona una mediad de la variación total entre los valores y, pasando por alto las x. De manera alternativa, la SSE mide la variación en los valores y que permanecen sin explicación después de usar las x para ajustar el modelo de regresión lineal simple. Entonces la razón SSE$/S_{yy}$ da la proporción de la variación total en las y_i que no es explicada por el modelo de regresión lineal.

Al reescribir el coeficiente de determinación se obtiene,

$$r^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 = \left(\frac{S_{xy}}{S_{xx}}\right)\left(\frac{S_{xy}}{S_{yy}}\right)$$
$$= \left(\frac{\hat{\beta}_1 S_{xy}}{S_{yy}}\right) = \frac{S_{yy} - \text{SSE}}{S_{yy}}$$
$$= 1 - \frac{\text{SSE}}{S_{yy}}$$

Así, r^2 se puede interpretar como la *proporción de la variación total en las y_i que es explicada por la variable x en un modelo de regresión lineal simple*.

11.10 Ajuste del modelo lineal mediante matrices

Hasta el momento solo se han visto resultados de modelos de regresión lineal simple. Sin embargo estos resultados se pueden generalizar para k parámetros esto se conoce como regresión lineal múltiple. La única forma práctica de manejar resultados y deducciones análogos para modelos de regresión lineal *múltiple* es por medio del álgebra de matrices.

Suponga que tenemos el modelo lineal

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

y hacemos n observaciones independientes, $y_1, y_2, \ldots, y_n$ en Y. Podemos escribir la observación y_i como

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

donde x_{ij} es el ajuste de la j-'esima variable independiente para la i-'esima observación $i = 1, 2, \ldots, n$. Ahora definimos las siguientes matrices con $x_0 = 1$:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} x_0 & x_{11} & x_{12} & \cdots & x_{1k} \\ x_0 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_0 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entonces, las n ecuaciones que representan y_i como función de las x, las β y las ϵ se pueden escribir simultáneamente como

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Ecuaciones de mínimos cuadrado y soluciones para un modelo lineal general

Ecuaciones: $(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$.
Soluciones: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

$$\text{SSE} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$$

Al igual que los resultados anteriores, se puede demostrar los resultados del modelo de regresión lineal simple usando matrices. Sin embargo para demostrar estos resultados se necesita tener conocimientos básicos del álgebra de matrices por lo que sale fuera del alcance de este resumen.

11.11 Funciones lineales de los parámetros del modelo: regresión lineal múltiple

Todos los resultados teóricos de la sección 11.4 se pueden ampliar al modelo de regresión múltiple,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

para $i = 1, 2, \ldots, n$. Suponga que $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ son variables aleatorias independientes con $E(\varepsilon_i) = 0$ y $V(\varepsilon_i) = \sigma^2$. Entonces los estimadores de mínimos cuadrados están dados por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

siempre que $(\mathbf{X}'\mathbf{X})^{-1}$ exista. Las propiedades de estos estimadores son las siguiente

Propiedades de los estimadores de mínimos cuadrados: regresión lineal múltiple

1. $E(\hat{\beta}_i) = \beta_i$, $i = 1, 2, \ldots, k$.
2. $V(\hat{\beta}_i) = c_{ii}\sigma 2$ donde c_{ii} es el elemento en la fila i y columna i de $(\mathbf{X}'\mathbf{X})^{-1}$. (Recuerde que esta matriz tiene una fila y una columna numerados con 0.)
3. $Cov(\hat{\beta}_i, \hat{\beta}_j) = c_{ij}\sigma^2$, donde c_{ij} es el elemento en la fila i y columna j de $(\mathbf{X}'\mathbf{X})^{-1}$.
4. Un estimador insesgado de σ^2 es $S^2 = SSE/[n - (k+1)]$, donde $\text{SSE} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$. (Observe que hay $k+1$ valore β_i desconocidas en el modelo.)

Si, además, las ε_i, para $i = 1, 2, \ldots, n$ están distribuidas normalmente,

5. Cada $\hat{\beta}_i$ está distribuida normalmente.
6. La variable aleatoria $\dfrac{[n-(k+1)]S^2}{\sigma^2}$ tiene una distribución χ^2 con $n - (k+1)$ grados de libertad.
7. Los estadísticos S^2 y $\hat{\beta}_i$ son independientes para cada $i = 0, 1, 2, \ldots, k$.

11.12 Inferencias respecto a funciones lineales de los parámetros del modelo: regresión lineal múltiple

Como lo vimos en las Secciones 11.5 y 11.6, podríamos estar interesados en hacer inferencias acerca de una β_i individual o de combinaciones lineales de los parámetros del modelo $\beta_0, \beta_1, \ldots, \beta_k$. Por ejemplo, queremos estimar $E(Y)$, dada por

$$E(Y) = \beta_0 + \beta_1 x_1 + \cdots \beta_k x_k$$

Suponga que deseamos hacer inferencia acerca de la función lineal

$$a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \cdots + a_k\beta_k$$

donde $a_0, a_1, a_2, \ldots, a_k$ son constantes. Definiendo la matriz $(k+1) \times 1$,

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}$$

se deduce que una combinación lineal de las $\beta_0, \beta_1, \ldots, \beta_k$ correspondiente a $a_0, a_1, \ldots, a_k$ se puede expresar como

$$\mathbf{a}'\boldsymbol{\beta} = a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \cdots + a_k\beta_k$$

Como $\mathbf{a}'\boldsymbol{\beta}$ es una combinación lineal de los parámetros del modelo, un estimador insesgado para $\mathbf{a}'\boldsymbol{\beta}$ está dado por la misma combinación lineal de los estimadores paramétricos. Esto es,

$$\widehat{\mathbf{a}'\boldsymbol{\beta}} = a_0\hat{\beta}_0 + a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + \cdots + a_k\hat{\beta}_k = \mathbf{a}'\hat{\boldsymbol{\beta}}$$

Se puede demostrar de manera sencilla que $E(\mathbf{a}'\hat{\boldsymbol{\beta}}) = \mathbf{a}'\boldsymbol{\beta}$, de la siguiente manera.

$$E(\mathbf{a}'\hat{\boldsymbol{\beta}}) = E(a_0\hat{\beta}_0 + a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + \cdots + a_k\hat{\beta}_k)$$
$$= a_0\beta_0 + a_1\beta_1 + a_2\beta_2 + \cdots + a_k\beta_k = \mathbf{a}'\boldsymbol{\beta}$$

Aplicando el Teorema 5.12 se tiene que la varianza de $\mathbf{a}'\boldsymbol{\beta}$:

$$V(\mathbf{a}'\hat{\boldsymbol{\beta}}) = V(a_0\hat{\beta}_0 + a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + \cdots + a_k\hat{\beta}_k)$$
$$= a_0^2 V(\hat{\beta}_0) + a_1^2 V(\hat{\beta}_1) + a_2^2 V(\hat{\beta}_2) + \cdots + a_k^2 V(\hat{\beta}_k)$$
$$+ 2a_0a_1 Cov(\hat{\beta}_0, \hat{\beta}_1) + 2a_0a_2 Cov(\hat{\beta}_0, \hat{\beta}_2) + \cdots$$
$$+ 2a_1a_2 Cov(\hat{\beta}_1, \hat{\beta}_2) + \cdots + 2a_{k-1}a_k Cov(\hat{\beta}_{k-1}, \hat{\beta}_k)$$

donde $V(\hat{\beta}_i = c_{ii}\sigma^2$ y $Cov(\hat{\beta}_i, \hat{\beta}_j) = c_{ij}\sigma^2$. Además se puede verificar que,

$$V(\mathbf{a}'\boldsymbol{\beta}) = \left[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\right]\sigma^2$$

Al igual que en la sección pasada, se sabe que los estimadores tienen distribución normal en muestreo repetitivo. Por esta razón $\mathbf{a}'\hat{\boldsymbol{\beta}}$ está distribuida normalmente en muestreo repetitivo con media $\mathbf{a}'\boldsymbol{\beta}$ y varianza $V(\mathbf{a}'\boldsymbol{\beta}) = \left[\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}\right]\sigma^2$. Con este resultados se puede ocupar el estadístico Z para desarrollar una prueba, sin embargo se tiene el mismo problema de no conocer a σ^2. Por esta razón se ocupa el estimador S^2 y el estadístico T para desarrollar una prueba para $\mathbf{a}'\boldsymbol{\beta}$ como se muestra a continuación.

> **Una prueba para $\mathbf{a}'\boldsymbol{\beta}$**
> $H_0 : \mathbf{a}'\boldsymbol{\beta} = (\mathbf{a}'\boldsymbol{\beta})_0$.
>
> $$H_a : \begin{cases} \mathbf{a}'\boldsymbol{\beta} > (\mathbf{a}'\boldsymbol{\beta})_0 \\ \mathbf{a}'\boldsymbol{\beta} < (\mathbf{a}'\boldsymbol{\beta})_0 \\ \mathbf{a}'\boldsymbol{\beta} \neq (\mathbf{a}'\boldsymbol{\beta})_0 \end{cases}$$
>
> Estadístico de prueba: $T = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}} - (\mathbf{a}'\boldsymbol{\beta})_0}{S\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$.
>
> $$\text{Región de rechazo} : \begin{cases} t > t_\alpha \\ t < -t_\alpha \\ |t| > t_{\alpha/2} \end{cases}$$
>
> Aquí, t_α está basada en $n-(k+1)$ grados de libertad.

> **Un intervalo de confianza $100(1-\alpha)\%$ para $\mathbf{a}'\boldsymbol{\beta}$**
>
> $$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2} S \sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$

Al igual que el caso de regresión lineal simple, el principal objetivo de realizar estos cálculos es encontrar el valor medio de Y, $E(Y)$ para valores fijos de las variables independientes $x_1, x_2, \ldots, x_k$. En particular, si x_i^* denota un valor específico de x_i, para $i = 1, 2, \ldots, k$, entonces

$$E(Y) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_k x_k^*$$

Observe que $E(Y)$ es un caso especial de $a_0\beta_0 + \cdots + a_k\beta_k = \mathbf{a}'\boldsymbol{\beta}$ con $a_0 = 1$ y $a_i = x_i^*$, para $i = 1, 2, \ldots, k$. Se puede desarrollar una inferencia alrededor de $E(Y)$ cuando $x_i = x_i^*$ para $i = 1, 2, \ldots, k$ usando las técnicas desarrolladas en esta sección.

11.13 Predicción de un valor particular de Y mediante regresión múltiple

En el caso de regresión múltiple, suponemos que se ajusta un modelo de regresión lineal como el siguiente,

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

Estamos interesados en predecir el valor de Y^* cuando $x_1 = x_1^*, x_2 = x_2^*, \ldots, x_k = x_k^*$. Este valor se puede predecir con

$$\widehat{Y^*} = \hat{\beta}_0 + \hat{\beta}_1 x_1^* + \hat{\beta}_2 x_2^* + \cdots + \hat{\beta}_k x_k^* = \mathbf{a}'\hat{\boldsymbol{\beta}}$$

donde

$$\mathbf{a} = \begin{bmatrix} 1 \\ x_1^* \\ x_2^* \\ \vdots \\ x_k^* \end{bmatrix}$$

Al igual que en la Sección 11.7 nos concentramos en la diferencia entre la variable Y^* y el valor predicho:

$$\text{error} = Y^* - \widehat{Y^*}$$

De la misma manera se puede demostrar que el error está distribuido normalmente además, si aplicamos el Teorema 5.12 y los resultados de la Sección 11.11 se puede encontrar que,

$$E(\text{error}) = 0 \qquad V(\text{error}) = \sigma^2[1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}]$$

Se puede ocupar la distribución t para encontrar un intervalo de predicción para el valor real de Y. Esto se muestra a continuación,

> **Un intervalo de predicción $100(1-\alpha)$ para Y cuando $\mathbf{x_1} = \mathbf{x_1^*}, \mathbf{x_2} = \mathbf{x_2^*}, \ldots, \mathbf{x_k} = \mathbf{x_k^*}$**
>
> $$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2} S \sqrt{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$$
>
> donde $\mathbf{a}' = [1, x_1^*, x_2^*, \ldots, x_k^*]$ y t_α está basada en una distribución t con $n-(k+1)$ grados de libertad.

11.14 Una prueba para $H_0 : \beta_{g+1} = \beta_{g+2} = \ldots = \beta_{g+k}$

Suponga que debemos ajustar un modelo que comprende sólo un subconjunto de las variables independientes en consideración, esto es ajustar un modelo reducido de la forma,

$$\text{modelo R:} Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_g x_g$$

Se calcula la suma de cuadrado de las desviaciones entre los valores observados y pronosticados de Y, SSE_R. Una vez que se hace esto se necesita ajustar el modelo lineal con *todas* las variables independientes, esto es:

$$\begin{aligned}\text{modelo C} : Y =& \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_g x_g + \beta_{g+1} x_{g+1} \\ &+ \cdots + \beta_k x_k + \varepsilon\end{aligned}$$

y determinar la suma de cuadrado de las desviaciones para este modelo, SSE_C. Ahora una pregunta interesante es como sabemos cual de estos dos modelos es el mejor para los datos. Es intuitivo que si el modelo reducido es el mejor para los datos entonces su SSE_R tendría un menor valor que la suma de cuadrados del modelo completo. Entonces nos concentramos en la diferencia entre estas dos sumas $(\text{SSE}_R - \text{SSE}_C)$.

Esta suma de cuadrados recibe el nombre de *suma de cuadrados asociada con las variables* $x_{g+1}, x_{g+2} + \cdots g_k$ ajustada para las variables $x_1, x_2, \cdots x_g$. Valores grandes de esta diferencia nos llevaría a rechazar la hipótesis nula $H_0 : \beta_{g+1} = \beta_{g+2} = \cdots = \beta_{g+k}$.

Para obtener un estadístico de prueba, *supongamos* que la hipótesis nula es verdadera y luego examinaremos las cantidades que hemos calculado, en particular observe que,

$$\text{SSE}_R = \text{SSE}_C + (\text{SSE}_R - \text{SSE}_C)$$

Se divide a la suma de cuadrado del modelo reducido en dos partes. Entonces si H_0 es verdadera, entonces

$$\chi_3^2 = \frac{\text{SSE}_R}{\sigma^2} \qquad \chi_2^2 = \frac{\text{SSE}_C}{\sigma^2} \qquad \chi_1^2 = \frac{\text{SSE}_R - \text{SSE}_C}{\sigma^2}$$

tiene distribución de probabilidad χ^2 en muestreo repetido con $(n-[g+1]), (n-[k+1])$ y $(k-g)$ grados de libertad, respectivamente. Además se puede demostrar que χ_2^2 y χ_1^2 son estadísticamente independientes.

Entonces el estadístico de prueba para la H_0 es el siguiente,

$$F = \frac{\chi_1^2/(k-g)}{\chi_2^2/(n-[k+1])} = \frac{(\text{SSE}_R - \text{SSE}_C)/(k-g)}{(\text{SSE}_C)/(n-[k+1])}$$

Si H_0 es verdadera, entonces F tiene una distribución F con $v_1 = k-g$ grados de libertad en el numerador y $v_2 = n-(k+1)$ grados de libertad en el denominador. La región de rechazo apropiad apara una probabilidad de error tipo I igual a α es

$$F > F_\alpha$$