$$\widetilde{VAR}(\hat{\beta}_0) = \tilde{\sigma}^2 \left[\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \right]$$

$$= 63.37 \left[\frac{1}{28} + \frac{(1.72)^2}{0.225} \right]$$

$$= 63.37 \times 13.21$$

$$= 837.12$$

$$\therefore \sqrt{\widetilde{VAR}(\hat{\beta}_0)} = 28.93$$

Finalmente, dado que para una v.a. Student con 26 grados de libertad, el cuantil de orden 0.975 es $t_{(26)}^{0.975} = 2.056$, entonces los intervalos de confianza a nivel 95% para β_1 y β_0 son

$$I_{\beta_1} = [104.87, 173.87]$$

$$I_{\beta_0} = [-229.42, -110.46]$$

Como ya habíamos comentado, al tener $r_{xy}^2 = 0.72$ la regresión lineal del peso como función de la estatura explica un 72% de la variabilidad en la variable de respuesta, el valor estimado $\hat{\beta}_1 = 139.37$ nos dice que por cada centímetro de aumento en la estatura se estima un incremento de 1.39 Kilos en el peso de una persona. ¿Qué tanta certeza se puede depositar en esta afirmación?

Usando el intervalo de confianza $[104.87, 173.87]$ se puede afirmar que, con un 95% de confianza se estima que por cada centímetro de aumento en la estatura se produce un incremento de peso que esté entre 1.05 y 1.74 Kilos[(1)]. Se concluye que el intervalo revela que, con un 95% de confianza, se puede considerar que el error de estimación es menor que 350 gramos. Reportar una cuantificación de la magnitud del error en el método de estimación usado siempre es crucial, independientemente de si el resultado será considerado preciso ó no.

(1) Este resultado puede considerarse suficientemente preciso, o no, dependiendo de los objetivos de la investigación.

EJERCICIO: Produzca el análisis correspondiente y discuta el resultado en el caso de β_0.

ESTIMACION DE LA MEDIA CONDICIONAL

Existe un parámetro que es de interés en el modelo y que es función de los coeficientes β_0 y β_1. Se trata del valor esperado condicional de Y dado un valor de la variable de control $X=x$. Para estimar

$$\mathbb{E}(Y|X=x) = \beta_0 + \beta_1 x = g(\beta_0, \beta_1),$$

podemos usar la propiedad de invarianza de los estimadores de maxima verosimilitud. Esta última establece que el estimador máximo verosimil para el parámetro $g(\beta_0, \beta_1)$ es $g(\hat{\beta}_0, \hat{\beta}_1)$, así

$$\widehat{\mathbb{E}(Y|X=x)} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Es fácil verificar que $\mathbb{E}\left(\widehat{\mathbb{E}(Y|X=x)}\right) = \mathbb{E}(Y|X=x)$, es decir $\widehat{\mathbb{E}(Y|X=x)}$ es insesgado. También resulta inmediato verificar que $\widehat{\mathbb{E}(Y|X=x)}$ tiene distribución normal (es una combinación lineal de las variables de respuesta $\widehat{\mathbb{E}(Y|X=x)} = \sum_{i=1}^{N} c_i Y_i$). Además, como $\widehat{\mathbb{E}(Y|X=x)} = g(\hat{\beta}_0, \hat{\beta}_1)$ es una función de $\hat{\beta}_0$ y $\hat{\beta}_1$ y estos últimos estimadores, vistos como v.a., son independientes de la v.a. $\tilde{\sigma}^2$, se tiene que $\widehat{\mathbb{E}(Y|X=x)}$ es independiente de $\tilde{\sigma}^2$.

Para calcular la varianza de $\widehat{\mathbb{E}(Y|X=x)}$ tenemos

$$
\begin{aligned}
VAR\left(\widehat{\mathbb{E}(Y|X=x)}\right) &= VAR\left(\hat{\beta}_0 + \hat{\beta}_1 x\right) \\
&= VAR(\hat{\beta}_0) + VAR(\hat{\beta}_1)x^2 + 2x\,COV(\hat{\beta}_0, \hat{\beta}_1)
\end{aligned}
$$

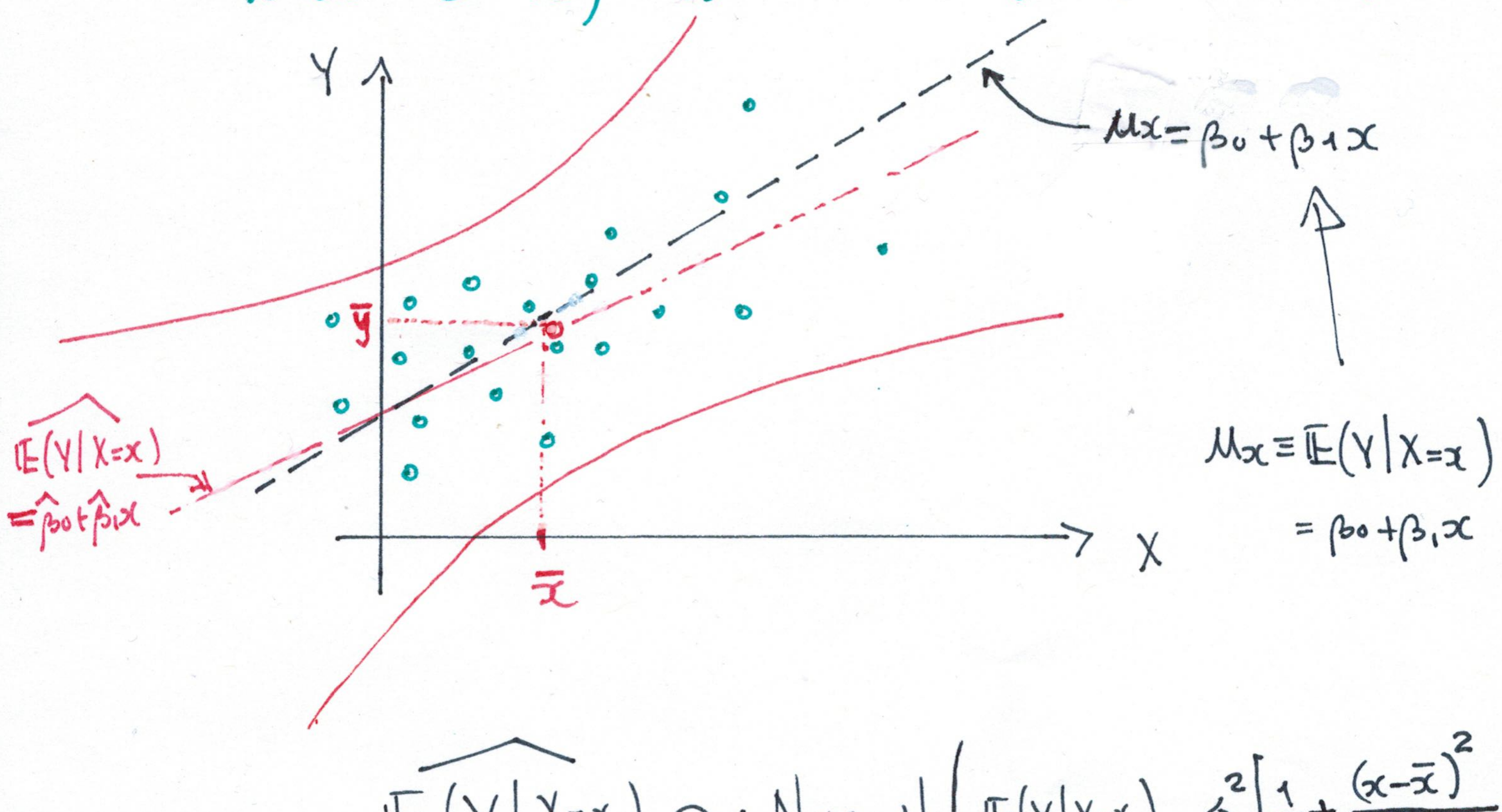
$$
= \sigma^2\left(\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right) + \frac{x^2\sigma^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} - 2\sigma^2 x\frac{\bar{x}}{\sum_{i=1}^{N}(x_i-\bar{x})^2}
$$

$$
= \sigma^2\left[\frac{1}{N} + \frac{\bar{x}^2 - 2x\bar{x} + x^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right]
$$

$$
= \sigma^2\left[\frac{1}{N} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right].
$$

Esta expresión de la varianza de $\widehat{\mathbb{E}(Y|X=x)}$, permite interpretar lo que sucede con la dispersión de $\widehat{\mathbb{E}(Y|X=x)}$ cuando el valor de x cambia:

En la medida en que x se aleja de la media $\bar{x}$ en los valores de la variable explicativa, la varianza crece en forma cuadrática. Se tiene que la varianza de $\widehat{\mathbb{E}(Y|X=x)}$, como función de x, es mínima cuando $x = \bar{x}$.



$$\widehat{\mathbb{E}(Y|X=x)} \sim \text{Normal}\left(\mathbb{E}(Y|X=x),\ \sigma^2\left[\frac{1}{N} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right]\right)$$

Por otra parte, el estimador insesgado para

$\mathrm{VAR}\left(\widehat{\mathbb{E}(Y|X=x)}\right)$ está dado por

$$\widetilde{\mathrm{VAR}}\left(\widehat{\mathbb{E}(Y|X=x)}\right) = \tilde{\sigma}^2\left[\frac{1}{N} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right]$$

<u>EJERCICIO</u>: Demuestre que el intervalo de confianza a nivel $(1-\alpha)\times 100\%$ para $\mathbb{E}(Y|X=x)$, está dado por

$$\left[\widehat{\mathbb{E}(Y|X=x)} \pm \sqrt{\widetilde{\mathrm{VAR}}\left(\widehat{\mathbb{E}(Y|X=x)}\right)}\cdot t_{(N-2)}^{1-\alpha/2}\right].$$

De regreso a los datos de estaturas y pesos, tenemos que

$$\widetilde{\mathrm{VAR}}\left(\widehat{\mathbb{E}(Y|X=x)}\right) = 63.37\left[\frac{1}{28} + \frac{(x-1.72)^2}{0.225}\right]$$

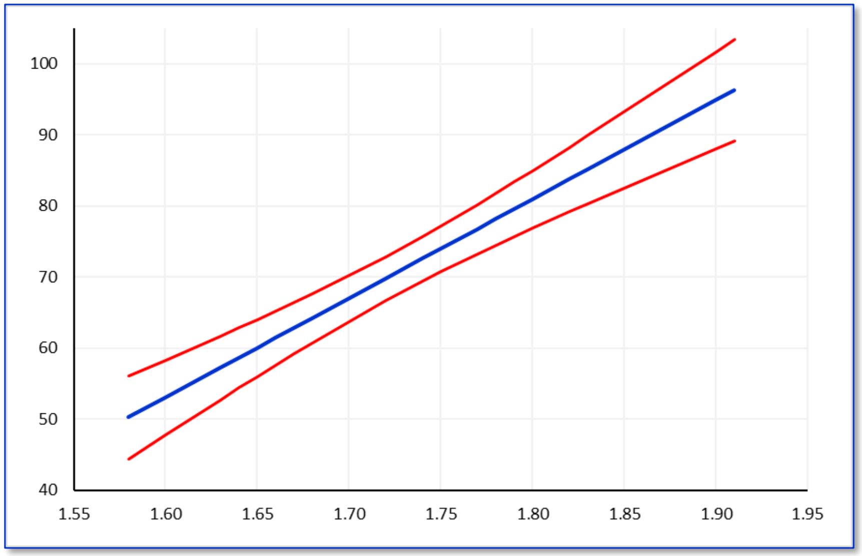
Para el valor $x=\bar{x}=1.72$ $\widetilde{\mathrm{VAR}}\left(\widehat{\mathbb{E}(Y|X=1.72)}\right) = \frac{63.37}{28}$ $= 2.26$. Como resultado, el intervalo de confianza al 95% para $\mathbb{E}(Y|X=1.72)$ está dado por $\left[70.28 \pm (1.5044)\cdot(2.056)\right] = [67.21, 73.39]$

Interpretamos este resultado como sigue: El modelo estima puntualmente que el promedio de peso para todas las personas que miden 1.72 metros de estatura es de aproximadamente 70.28 Kilos. De acuerdo con el intervalo, con una confianza de 95%, se puede afirmar que ese promedio se encuentra entre 67.2 y 73.39 Kilos

El intervalo <u>depende del valor de x</u>. Para un nivel de confianza fijo, podemos calcular los intervalos para cada uno de los valores de x en la zona de exploración. Ese conjunto de intervalos se puede visualizar en forma conjunta en una gráfica donde se forma una banda de confianza para la esperanza condicional $\mathbb{E}(Y/X=x)$

I.8. Análisis Estadístico

En la gráfica se presenta la banda de confianza al 95% para la Esperanza condicional cuando se analizan los datos de Peso y Estatura.



Banda de Confianza al 95% para el valor esperado condicional del Peso dada la Estatura

Cuando el valor de σ^2 es grande, la calidad de las inferencias (estimadores, intervalos de confianza y predicciones) es menor. Las varianzas de los estimadores para β_0 y β_1 aumentan, al igual que la varianza de $\mathbb{E}(Y|X=x)$.

Para producir un intervalo de confianza para σ^2 recordemos que

$$U \equiv \frac{(N-2)\tilde{\sigma}^2}{\sigma^2} \sim \chi^2_{(N-2)}$$

Se pueden construir 2 tipos de intervalos:

(a) Longitud mínima

Elegir la constante c tal que $\mathbb{P}(c < U) = 1-\alpha$

c coincide con el cuantil de orden α de la distribución $\chi^2_{(N-2)}$

$$\mathbb{P}\left(\chi^{2,\alpha}_{(N-2)} < \frac{(N-2)}{\sigma^2}\tilde{\sigma}^2\right) = 1-\alpha$$

$$\mathbb{P}\left(\sigma^2 < \frac{(N-2)\tilde{\sigma}^2}{\chi^{2,\alpha}_{(N-2)}}\right) = 1-\alpha$$

Una vez que los datos son observados y el estimador $\tilde{\sigma}^2$ se calcula, el intervalo es

$$\left(0, \frac{(N-2)\tilde{\sigma}^2}{\chi^{2,\alpha}_{(N-2)}}\right)$$

Interpretación: Con una confianza de $(1-\alpha)\times 100\%$, la varianza σ^2 no es mayor a $\dfrac{(N-2)\tilde{\sigma}^2}{\chi^{2,\alpha}_{N-2}}$

(b) Intervalo usual (no óptimo)

Sean a y b números reales tales que

$$\mathbb{P}(a < U < b) = 1-\alpha,$$

entonces

$$\mathbb{P}\left(a < \frac{(N-2)\tilde{\sigma}^2}{\sigma^2} < b\right) = 1-\alpha$$

$$\Rightarrow \quad \mathbb{P}\left(\frac{(N-2)\tilde{\sigma}^2}{b} < \sigma^2 < \frac{(N-2)\tilde{\sigma}^2}{a}\right) = 1-\alpha$$

Una vez que los datos son observados se calcula $\tilde{\sigma}^2$ y el intervalo es

$$\left[\frac{(N-2)\tilde{\sigma}^2}{b}, \frac{(N-2)\tilde{\sigma}^2}{a}\right]$$

existen muchas formas de elegir a y b, pero lo usual es $a = \chi^{2,\alpha/2}_{(N-2)}$ y $b = \chi^{2,1-\alpha/2}_{(N-2)}$

PRONOSTICOS

Supongase que se considera hacer inferencias sobre la respuesta Y cuando se fija el valor de la variable explicativa $X = x_*$, en específico pensando en la situacion en la cual x_* no esté en $\{x_1, x_2, \ldots, x_N\}$. Denotaremos por Y_* al valor que se observará[1] cuando $X = x_*$. Dados los supuestos del modelo, tenemos que

$$Y_* | X = x_* \sim N\left(\beta_0 + \beta_1 x_*, \sigma^2\right) \ldots \ldots (P1)$$

y además, condicionalmente a $X = x_*$, Y_* es independiente de $Y_1, Y_2, \ldots, Y_N$

(1) Como Y_* no se ha observado, le llamaremos "valor futuro de Y", estamos pronosticando lo que podría valer

Una idea de lo que puede valer Y_* está dada por el pronóstico[1] puntual

$$\boxed{\hat{Y}_* | X = x_* \equiv \hat{Y}_*} = \hat{\beta}_0 + \hat{\beta}_1 x_*$$

cuyo valor, coincide con el del estimador puntual $\widehat{\mathbb{E}(Y|X=x_*)} = \hat{\mu}_* = \hat{\beta}_0 + \hat{\beta}_1 x_*$.

Ya habíamos discutido el hecho de que $\widehat{\mathbb{E}(Y|X=x)}$ es una combinación lineal de v.a. Normales independientes (tiene distribución normal), su valor esperado es $\mathbb{E}(Y|X=x_*) = \beta_0 + \beta_1 x_*$ y su varianza es $VAR\left(\widehat{\mathbb{E}(Y|X=x_*)}\right) = \sigma^2 \left[\frac{1}{N} + \frac{(x_* - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right]$. Entonces

$$\hat{Y}_* | X = x_* \overset{(2)}{\sim} N\left(\beta_0 + \beta_1 x_*, \ \sigma^2\left[\frac{1}{N} + \frac{(x_* - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right]\right)$$

.... : (P2)

(2) <u>De hecho</u> $\hat{Y}_* | X_1 = x_1, \ldots, X_N = x_N ; X = x_* \sim N\left(\beta_0 + \beta_1 x_*, \sigma^2\left[\frac{1}{N} + \frac{(x_* - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right]\right)$

(1) Cuando se habla de cantidades <u>no</u> aleatorias como $\beta_0, \beta_1, \sigma^2$ decimos que se estiman puntualmente con $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$ funciones de la muestra. Cuando la cantidad es aleatoria se habla de pronosticar su valor

El siguiente argumento nos dice que, condicionalmente a $X_1=x_1, \dots, X_N=x_N; X_*=x_*$, las v.a. $\hat{Y}_*$ y Y_* en (P2) y (P1) son independientes:

① De la página 20, usando los supuestos del modelo

$$f_{Y_1,\dots,Y_N,Y_* \mid X_1=x_1,\dots;X_N=x_N;X=x_*}(y_1,\dots,y_N,y_* \mid x_1,\dots,x_N,x_*)$$

$$=\prod_{j=1}^{N} f_{Y_j \mid X_j=x_j}(y_j \mid x_j) \times f_{Y_* \mid X=x_*}(y_* \mid x_*)$$

es decir el valor futuro Y_* se produce con el mismo mecanismo que generó a $Y_1,\dots,Y_N$ y condicionalmente a $X_1=x_1; X_2=x_2; \dots; X_N=x_N, X=x_*$ se tiene que $Y_*, Y_1, \dots, Y_N$ son independientes.

② El pronóstico puntual $\hat{Y}_*=\hat{\beta}_0+\hat{\beta}_1 x_*$ es función de las v.a. $Y_1,\dots,Y_N$, por ①, condicionalmente a $X_1=x_1; X_2=x_2; \dots; X_N=x_N, X=x_*$, Y_* es independ. de $\hat{Y}_*$

En otras palabras: $\hat{Y}_* \mid X_1=x_1, \dots, X_N=x_N, X=x_*$ es independiente de $Y_* \mid X_1=x_1, \dots, X_N=x_N, X=x_* = Y_* \mid X=x_*$

Como consecuencia, dado $X_1 = x_1, \ldots, X_N = x_N$, $X = x_*$

$$Y_* - \hat{Y}_* \sim N\left(0, \sigma^2\left[1 + \frac{1}{N} + \frac{(x_* - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right]\right),$$

pero además Y_*, $\hat{\beta}_0$ y $\hat{\beta}_1$ son condicionalmente[1] independientes[2] de S^2. Por tanto $Y_* - \hat{Y}_*$ es condicionalmente independiente de S^2. Entonces

$$Z \equiv \frac{Y_* - \hat{Y}_*}{\sigma\left[1 + \frac{1}{N} + \frac{(x_* - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right]^{1/2}} \sim N(0,1) \quad \text{y}$$

$$U \equiv \frac{(N-2)}{\sigma^2} S^2 \sim \chi^2_{(N-2)}$$

son v.a. independientes (condicionalmente), de manera que $t \equiv Z / \sqrt{U/(N-2)} \sim \text{Student}_{(N-2)}$

(1) condicionalmente a $X_1 = x_1, \ldots, X_N = x_N$, $X = x_*$.

(2) Ya se había revisado parcialmente en las páginas 71-72.

Puesto que

$$t = \frac{Y_* - \hat{Y}_*}{\tilde{\sigma}\left(1 + \frac{1}{N} + \frac{(x_* - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right)^{1/2}}$$

Encontrar el cuantil $t_{(N-2)}^{1-\alpha/2}$ tal que

$$\mathbb{P}\left(-t_{(N-2)}^{1-\alpha/2} < t < t_{(N-2)}^{1-\alpha/2}\right) = 1-\alpha$$

justifica la igualdad

$$\mathbb{P}\left(Y_* \in \mathbf{I}_*\right) = 1-\alpha, \quad ---- \quad (P3)$$

donde $\mathbf{I}_*$ es el intervalo de predicción o pronóstico

$$\mathbf{I}_* = \left(\hat{Y}_* - \tilde{\sigma}\left(1+\frac{1}{N}+\frac{(x_*-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right)^{1/2}\cdot t_{(N-2)}^{1-\alpha/2},\ \hat{Y}_* + \tilde{\sigma}\left(1+\frac{1}{N}+\frac{(x_*-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right)^{1/2}\cdot t_{(N-2)}^{1-\alpha/2}\right)$$

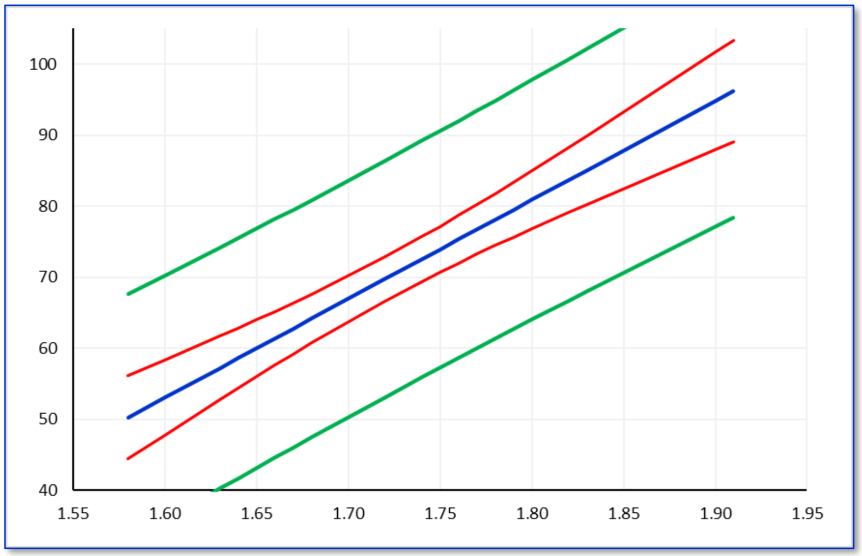
☆ $\mathbf{I}_*$ no es un intervalo de confianza. Como Y_* es una v.a. la relacion (P3) se interpreta como que la probabilidad de que una futura observación Y_* esté en $\mathbf{I}_*$ vale $1-\alpha$. No obstante esta afirmación se hace pensando en que $\hat{Y}_*$ y $\tilde{\sigma}$ son aleatorias ($\mathbf{I}_*$ es aleatorio). Una vez que se observan $(x_1, y_1), \ldots, (x_N, y_N)$ y se calculan $\hat{Y}_*$ y $\hat{\sigma}$, entonces I_* ya no es aleatorio y la afirmación ★ es una aproximación

En la figura que sigue[1] se muestran en línea azul la estimación de la media condicional $\widehat{\mathbb{E}(Y|X=x_*)}$ para x_* en la zona de exploración. Estas estimaciones coinciden con los pronósticos $\hat{Y}_*|X=x_* = \hat{\beta}_0 + \hat{\beta}_1 x_*$. En color rojo se muestran los intervalos de confianza al 95% para $\mathbb{E}(Y|X=x)$ (véase el ejercicio en la página 83), al graficar cada intervalo para cada valor de x en la zona de exploración, se obtienen las bandas.

En color verde se muestran los intervalos de pronóstico al 95% para un valor futuro Y_* que denotamos I_* en la página 92

(1) para los datos de estaturas y pesos

Como ejemplo del diferencial de longitud, en la siguiente gráfica se exhiben *la banda de pronóstico para una observación futura del Peso, dada la Estatura* al 95% (en verde) y *la banda de confianza para la media Poblacional del Peso dada al misma Estatura*, también el 95% (en rojo).



Banda de Pronóstico al 95% para el valor del Peso dada la Estatura

En ocasiones, parte del análisis estadístico para el modelo de regresión lineal simple requiere contrastar hipótesis sobre los parámetros desconocidos del modelo.

Sea $\Theta \equiv (\beta_0, \beta_1, \sigma^2)$, en general se pueden considerar las hipótesis

$$H_0 : \Theta \in \omega,$$
$$H_1 : \Theta \in \omega^c,$$

donde ω es un subconjunto conocido de $\mathbb{R} \times \mathbb{R} \times (0, +\infty)$. Aunque en general las hipótesis de interés dependen de los objetivos específicos del estudio, existe un par de hipótesis cuyo contraste es obligatorio en todo modelo de regresión lineal simple, estas hipótesis son

$$H_0 : \beta_1 = 0,$$
$$H_1 : \beta_1 \neq 0.$$

Del modelo de regresión, se tiene que

$$Y|X=x \sim N(\beta_0+\beta_1 x, \sigma^2),$$

(MODELO COMPLETO) [MC]

pero si la hipótesis Ho es cierta, entonces tenemos el Modelo Reducido [MR]

$$Y \sim N(\beta_0, \sigma^2),$$

el cual **no** depende del valor de **X**.

El modelo reducido es un caso particular del modelo completo, si dada la información de los datos no se rechaza la hipótesis Ho, entonces para efectos prácticos, ambos modelos proveen una descripción igualmente razonable de la variable Y. Si no se rechaza Ho, se puede concluir que no se cuenta con evidencia estadística de asociación lineal entre X y Y, en otras palabras se puede abandonar el modelo de regresión lineal simple.

Una forma de desarrollar una prueba de hipótesis es usando el Cociente de Verosimilitudes Generalizado, se rechaza Ho si y sólo si los datos pertenecen a la región

$$\mathcal{C} = \{ \mathbf{Y} \in \mathbb{R}^N \,/\, \Lambda < c \}$$

donde

$$\Lambda = \frac{\sup_{H_0} L(\beta_0, \beta_1, \sigma^2 \mid (x_1, y_1), \dots, (x_N, y_N))}{\sup_{H_0 \cup H_1} L(\beta_0, \beta_1, \sigma^2 \mid (x_1, y_1), \dots, (x_N, y_N))}$$

$$= \frac{\sup_{[MR]} L(\beta_0, \beta_1, \sigma^2 \mid (x_1, y_1), \dots, (x_N, y_N))}{\sup_{[MC]} L(\beta_0, \beta_1, \sigma^2 \mid (x_1, y_1), \dots, (x_N, y_N))}$$

Numerador: Verosimilitud correspondiente al modelo reducido

Denominador: Verosimilitud correspondiente al modelo completo

c es una constante tal que

(α) ..... $\sup_{H_0} \mathbb{P}(\Lambda < c) = \alpha$ $\left\{ \begin{array}{l} \text{De hecho} \\ \mathbb{P}(\Lambda < c \mid \overset{H_0}{\overline{\beta_1 = 0}}) = \alpha \end{array} \right.$

α valor de significancia de la prueba.

La función de verosimilitud para el modelo completo esta dada por

(α) se usa si el dominio especificado en Ho es un conjunto

$$L(\beta_0, \beta_1, \sigma^2 \mid (x_1, y_1), \ldots, (x_N, y_N))$$
$$= (2\pi\sigma^2)^{-N/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2}.$$

Esta función se maximiza al evaluarla en los estimadores de maxima verosimilitud, de donde el denominador en Λ se escribe

$$\text{Sup}_{H_0 \cup H_1} L(\beta_0, \beta_1, \sigma^2) = L(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$$
$$= (2\pi\hat{\sigma}^2)^{-N/2} e^{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2},$$

pero como $\hat{\sigma}^2 = \left(\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right)/N$, entonces

$$\text{Sup}_{H_0 \cup H_1} L(\beta_0, \beta_1, \sigma^2) = (2\pi\hat{\sigma}^2)^{-N/2} e^{-N/2}.$$

Usando el modelo completo con la restriccion $\beta_1 = 0$ obtenemos el modelo reducido, como consecuencia la verosimilitud del modelo reducido resulta

(2) $$L(\beta_0, 0, \sigma^2 \mid (x_1, y_1), \ldots, (x_N, y_N)) = (2\pi\sigma^2)^{-N/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \beta_0)^2}$$

Para maximizar (2) como función de β_0 y σ^2, notamos que esta función es la verosimilitud de variables independientes e idénticamente distribuidas $N(\beta_0, \sigma^2)$, entonces como función de β_0 y σ^2 (2) se maximiza al evaluar en los estimadores de máxima verosimilitud

$$\hat{\beta}_0^{\circ} = \bar{y} \quad \text{y} \quad \hat{\sigma}_0^2 = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0^{\circ}\right)^2 .$$

Luego

$$\begin{aligned}
\sup_{H_0} L(\beta_0, \beta_1, \sigma^2) &= L(\hat{\beta}_0^{\circ}, 0, \hat{\sigma}_0^2) \\
&= \left(2\pi\hat{\sigma}_0^2\right)^{-N/2} e^{-\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^{N}\left(y_i-\hat{\beta}_0^{\circ}\right)^2} \\
&= \left(2\pi\hat{\sigma}_0^2\right)^{-N/2} e^{-N/2} .
\end{aligned}$$

Así obtenemos que

$$\Lambda = \frac{\left(2\pi\hat{\sigma}_0^2\right)^{-N/2} e^{-N/2}}{\left(2\pi\hat{\sigma}^2\right)^{-N/2} e^{-N/2}} = \frac{\left(\hat{\sigma}_0^2\right)^{-N/2}}{\left(\hat{\sigma}^2\right)^{-N/2}}$$

$$= \left\{\frac{\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \hat{\beta}_0^{\circ}\right)^2}\right\}^{N/2}$$

$\hat{y}_i^{\circ} \equiv \bar{y}$

$$= \left\{\frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \hat{y}_i^{\circ}\right)^2}\right\}^{N/2} = \left\{\frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}\right\}^{\frac{N}{2}}$$

En el modelo completo $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ es el pronóstico puntual para la i-ésima observación y coincide con la estimación puntual de la esperanza (condicional) de y_i. En forma análoga, en el modelo reducido $\hat{\beta}_0^0 = \bar{y}$ es la estimación de la esperanza de y_i y también es un pronóstico puntual para la i-ésima observación, de estas observaciones decimos que

$SCE_{MC} \equiv \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$ es la suma de cuadrados de los errores de los pronósticos en el modelo completo

$SCE_{MR} \equiv \sum_{i=1}^{N} (y_i - \bar{y})^2$ es la suma de cuadrados de los errores de los pronósticos en el modelo reducido

Por tanto

$$C = \left\{ Y \in \mathbb{R}^N \;\middle|\; \frac{SCE_{MC}}{SCE_{MR}} < c^{2/N} \right\}$$

$c' \equiv c^{2/N}$

$$C = \left\{ Y \in \mathbb{R}^N \;\middle|\; \frac{SCE_{MC}}{SCE_{MR}} < c' \right\}$$

Para encontrar el valor de c' hay que determinar la distribución del cociente

$$\frac{SCE_{MC}}{SCE_{MR}}$$

bajo Ho. Cuando estudiamos el coeficiente de determinación R^2 [a] se observó que

$$(U) \ldots \quad \sum_{i=1}^{N}(y_i-\hat{y}_i)^2 = \sum_{i=1}^{N}(y_i-\bar{y})^2 - \hat{\beta}_1^2 \sum_{i=1}^{N}(x_i-\bar{x})^2$$

Denotando $S_{xx} = \sum_{i=1}^{N}(x_i-\bar{x})^2$ tenemos que (U) implica

$$(V) \quad SCE_{MR} = SCE_{MC} + \hat{\beta}_1^2 S_{xx}$$

SCE_{MR} mide el error de pronóstico que se produce al usar el modelo reducido

SCE_{MC} mide el error de pronóstico que se produce al usar el modelo completo

Como el modelo reducido es un caso particular del modelo completo, es razonable que su error de pronóstico sea mayor. La ecuación (V) dice que la magnitud en la que se incrementa

(a) r_{XY}^2

el error (al usar el modelo reducido) es $\hat{\beta}_1 S_{xx}$, esta última cantidad se puede pensar como la fracción del error atribuible a dar por cierta la hipótesis H_0. Tenemos así, la igualdad fundamental en sumas de cuadrados

$$SCE_{MR} = SCE_{MC} + \overbrace{SCE_{H_0}}^{\hat{\beta}_1 S_{xx}}$$

el error (al usar el modelo reducido) es $\hat{\beta}_1 S_{xx}$, esta última cantidad se puede pensar como la fracción del error atribuible a dar por cierta la hipótesis H_0. Tenemos así, la igualdad fundamental en sumas de cuadrados

$$SCE_{MR} = SCE_{MC} + \overbrace{SCE_{H_0}}^{\hat{\beta}_1 S_{xx}}$$

(*)

Bajo H_0: $$\boxed{\frac{SCE_{MC}}{\sigma^2} = \frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sigma^2} = \frac{N}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{(N-2)}}$$

Bajo H_0: $y_1 \ldots y_N$ son v.a. i.i.d. $N(\mu, \sigma^2)$ luego $\sum_1^N (y_i-\bar{y})/\sigma^2 \sim \chi^2_{(N-1)}$ $\longrightarrow$

$$\frac{SCE_{MR}}{\sigma^2} = \frac{\sum_{i=1}^{N}(y_i-\bar{y})^2}{\sigma^2} \sim \chi^2_{(N-1)}$$

$$\frac{SCE_{H_0}}{\sigma^2} \sim \chi^2_{(1)}$$

$$\underbrace{\frac{SCE_{MR}}{\sigma^2}}_{\chi^2_{(N-1)}} = \underbrace{\overbrace{\frac{SCE_{MC}}{\sigma^2}}^{\chi^2_{(N-2)}} + \overbrace{\frac{SCE_{H_0}}{\sigma^2}}^{\chi^2_{(1)}}}_{\text{V.A. INDEPENDIENTES (2)}}$$

(*) De hecho $\frac{SCE_{MC}}{\sigma^2} = \frac{(N-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(N-2)}$ es un hecho sin importar si H_0 sucede o no sucede

(2) $\dfrac{SCE_{MC}}{\sigma^2}$ es indep. de $\dfrac{SCE_{H_0}}{\sigma^2}$

porque $\hat{\beta}_1$ es indep. de $\hat{\sigma}^2$

$\therefore$ $\hat{\beta}_1^2 S_{xx}$ es indep. de $\dfrac{(N-2)\hat{\sigma}^2}{\sigma^2}$

Probemos que $\hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx})$

luego $\dfrac{\sqrt{S_{xx}}\,(\hat{\beta}_1-\beta_1)}{\sigma} \sim N(0,1)$

Pero Bajo Ho $\dfrac{\sqrt{S_{xx}}\,\hat{\beta}_1}{\sigma} \sim N(0,1)$

$$\therefore \quad \frac{S_{xx}(\hat{\beta}_1)^2}{\sigma^2} = \frac{SCE_{H_0}}{\sigma^2} \sim \chi^2_{(1)}$$

Ahora, para encontrar la región de Rechazo

$\mathcal{C}$

$$\frac{SCE_{MC}}{SCE_{MR}} < c'$$

$$\iff \frac{SCE_{MC}}{SCE_{MC} + SCE_{H_0}} < c'$$

$$\Longleftrightarrow \quad \frac{1}{1 + \frac{SCE_{Ho}}{SCE_{mc}}} < c'$$

$$\Longleftrightarrow \quad c'' \equiv \frac{1}{c'} < 1 + \frac{SCE_{Ho}}{SCE_{mc}}$$

$$\Longleftrightarrow \quad \frac{SCE_{Ho}}{SCE_{mc}} > c'' - 1 \equiv c'''$$

$$\Longleftrightarrow \quad \frac{\frac{SCE_{Ho}}{(1)\,s^2}}{\frac{SCE_{mc}}{(N-2)\,s^2}} > K$$

Pero $\frac{SCE_{Ho}}{(1)\,s^2}$ y $\frac{SCE_{mc}}{(N-2)\,s^2}$ son dos v.a. independ. con distribuciones $\chi^2_{(1)}$ y $\chi^2_{(N-2)}$ respectivamente. Por tanto

$$F \equiv \frac{\frac{SCE_{Ho}}{(1)\,s^2}}{\frac{SCE_{mc}}{(N-2)\,s^2}} \sim F_{1,\,N-2}$$

De esta forma la región de Rechazo para Ho es

$$C = \left\{ \mathbb{Y} \in \mathbb{R}^n : \left(\frac{N-2}{1}\right) \frac{SCE_{Ho}}{SCE_{mc}} > K \right\}$$

Para encontrar (dado un nivel de significancia α) el valor de K tenemos que ver que se cumple la condición

$$(3) \quad \ldots\ldots \quad \sup_{\Theta \in H_0} \left\{ \mathbb{P}(F_c > K \mid H_0) \right\} = \alpha^{(A)}$$

Para lo cual tendríamos que identificar la forma en que (condicionalmente en H_0) la probabilidad $\mathbb{P}(F_c > K)$ depende de Θ

Bajo H_0 : $\Theta = (0, \beta_0, \sigma^2)$

$$\in \{0\} \times \mathbb{R} \times (0, \infty)$$

$$= \omega$$

Pero hasta el momento hemos probado que bajo H_0 $\quad F \sim F_{1, N-2}$ que NO depende ni de β_0 ni de σ^2. Por tanto

$$\sup_{\Theta \in H_0} \left\{ \mathbb{P}\{F_c > K \mid H_0\} \right\} = \mathbb{P}\{F_c > K \mid H_0\}$$

(A) F_c es el valor de $F = \dfrac{SCE_{H_0}}{SCE_{mc}} \times \left(\dfrac{N-2}{1}\right)$ calculado usando la muestra $(x_1, y_1), \ldots, (x_N, y_N)$

y para encontrar K en (3)

$$\mathbb{P}(F_c > K \mid H_0) = \alpha$$

$$\mathbb{P}(F_c > K \mid F_c \sim F_{1,N-2}) = \alpha$$

$$\Longleftrightarrow \quad K = F^{1-\alpha}_{1,N-2}$$

$$\therefore \quad \mathcal{C} = \{ \mathbf{Y} \in \mathbb{R}^N \mid F_c > F^{1-\alpha}_{1,N-2} \}$$

TABLA ANOVA

FUENTE	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	CUADRADOS MEDIOS	F_c
H_0	SCE_{H_0}	1	$SCE_{H_0}/1$	$\dfrac{SCE_{H_0}/1}{SCE_{MC}/(N-2)}$
Modelo completo	SCE_{MC}	$N-2$	$SCE_{MC}/N-2$	
Modelo Reducido	SCE_{MR}	$N-1$		

Ejemplo: Datos estaturas y pesos

Fuente	G.L.	SCE	CME	F_c	P
H_0	1	4363.65	4363.65	68.90	0.000000009
Modelo completo	26	1646.72	63.34		
Modelo Reducido	27	6010.37			