Modelos de regresión lineal multivariada

- Análisis de regresión es la metodología estadística para predecir valores de una o varias variables respuestas (dependientes), a partir de una colección de variables predictoras (independientes)
- También se puede usar para valorar el efecto que tienen las variables predictoras (independientes) sobre las respuestas, es decir, las variables que tienen mayor influencia en la respuesta
- Primero revisaremos el modelo de regresión múltiple para la predicción de una variable sencilla (modelo de regresión clásico)
- Luego el modelo lo generalizaremos para la predicción de varias variables dependientes (modelo de regresión multivariada)

- Sea $z_1, z_2, \ldots, z_r$ un conjunto de r variables predictoras que se cree están relacionadas con una variable de respuesta Y. Por ejemplo, podríamos tener

$$Y = \text{Valor actual de la vivienda en el mercado}$$

y

$$\begin{aligned} z_1 &= \text{metros cuadrados de la superficie habitable} \\ z_2 &= \text{ubicación (indicador de la zona de la ciudad)} \\ z_3 &= \text{valor de la vivienda el año pasado} \\ z_4 &= \text{calidad de la construcción (precio por metro cuadrado)} \end{aligned}$$

- Se asume que el precio de la vivienda está relacionada con estas 4 variables

- Los modelos de regresión clásicos nos dicen que Y está compuesta de una media, la cual depende de manera continua de un conjunto de variables independientes $z_i's$, y de un error aletorio ε
- ε representa el error de medición y los efectos de otras variables que no están explícitamente consideradas en el modelo.
- Los valores de las variables predictoras registradas a partir del experimento son tratados como *fijos*.
- El error y por tanto la respuesta son consideradas como variables aleatorias cuyo comportamiento es caracterizado por un conjunto de suposiciones distribucionales
- El término *lineal* se refiere al hecho de que la media es una función lineal de los parámetros desconocidos $\beta_0, \beta_1, \ldots, \beta_r$

- El modelo de regresión lineal con una única respuesta tiene la forma:

$$Y = \beta_0 + \beta_1 z_1 + \cdots + \beta_r z_r + \varepsilon$$

$$[Respuesta] = [\text{media}(\text{depende de } \; z_1, z_2, \ldots, z_r)] + [error]$$

- Tomando n observaciones independientes de Y y los valores asociados de z_i, el modelo completo se define como:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 z_{11} + \beta_2 z_{12} + \cdots + \beta_r z_{1r} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 z_{21} + \beta_2 z_{22} + \cdots + \beta_r z_{2r} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 z_{n1} + \beta_2 z_{n2} + \cdots + \beta_r z_{nr} + \varepsilon_n \end{aligned}$$

- Donde se asume que los términos del error ε_j tienen las siguientes propiedades:
  1. $E(\varepsilon_j) = 0$
  2. $Var(\varepsilon_j) = \sigma^2(\text{constante})$
  3. $Cov(\varepsilon_j, \varepsilon_k) = 0, \; j \neq k$

- En notación matricial, el modelo se puede escribir como:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{nr} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

o

$$\boldsymbol{Y}_{(n\times 1)} = \boldsymbol{Z}_{(n\times(r+1))}\beta_{((r+1)\times 1)} + \varepsilon_{(n\times 1)}$$

y las especificaciones sobre las ε son ahora

1. $E(\varepsilon) = \mathbf{0}$
2. $Cov(\varepsilon) = E(\varepsilon\varepsilon^{'}) = \sigma^2\boldsymbol{I}$

- Obsérvese que el *uno* en la primera columna de la *matriz de diseño* Z es el multiplicador del término constante β_0. Es habitual introducir la variable artificial $z_{j0} = 1$, tal que

$$\beta_0 + \beta_1 z_{j1} + \beta_2 z_{j2} + \cdots + \beta_r z_{jr} = \beta_0 z_{j0} + \beta_1 z_{j1} + \beta_2 z_{j2} + \cdots + \beta_r z_{jr}$$

- Cada columna de $\boldsymbol{Z}$ consiste de los n valores de la variable predictora correspondiente, mientras que la j-ésima fila de $\boldsymbol{Z}$ contiene los valores para todas las variables predictoras sobre el j-ésimo ensayo.

El Modelo de regresión lineal clásico se resume como

$$\boldsymbol{Y}_{(n\times 1)} = \boldsymbol{Z}_{(n\times(r+1))}\beta_{((r+1)\times 1)} + \varepsilon_{(n\times 1)}$$

$$E(\varepsilon) = \mathbf{0}_{n\times 1} \quad \text{y} \quad Cov(\varepsilon) = \sigma^2_{n\times n}\boldsymbol{I}$$

donde β y σ^2 son parámetros desconocidos y la matriz de diseño $\boldsymbol{Z}$ tiene en la j-ésima fila $[z_{j0}, z_{j1}, \ldots, z_{jr}]$

Ejemplo: Modelos de regresión lineal clásico

Determinar el modelo de regresión lineal para ajustar una línea recta

$$\text{Respuesta media} = E(Y) = \beta_0 + \beta_1 z_1$$

a los datos

z_1	0	1	2	3	4
y	1	4	3	8	9

Estimación de los parámetros por mínimos cuadrados

- Uno de los objetivos del análisis de regresión es desarrollar una ecuación que nos permite predecir la respuesta, para los valores dados de las variables predictoras
- Es decir, debemos determinar los valores de los *coeficientes de regresión* β y de *la varianza del error* σ^2 que sean consistentes con los datos disponibles.
- Sea $\boldsymbol{b}$ un conjunto de valores de prueba para β, consideremos la diferencia $y_j - (b_0 + b_1 z_{j1} + \cdots + b_r z_{jr})$ entre la respuesta observada y_j y los valores $b_0 + b_1 z_{j1} + \cdots + b_r z_{jr}$, que se esperaría si $\boldsymbol{b}$ fuera el *verdadero* vector de parámetros.
- El método de mínimos cuadrados selecciona $\boldsymbol{b}$ que minimiza la suma de los cuadrados de las distancias:

$$S(\boldsymbol{b}) = \sum_{j=1}^{n} [y_j - (b_0 + b_1 z_{j1} + \cdots + b_r z_{jr})]^2 = (\boldsymbol{y} - \boldsymbol{Zb})'(\boldsymbol{y} - \boldsymbol{Zb})$$

Estimación de los parámetros por mínimos cuadrados

- Los coeficientes $\boldsymbol{b}$ obtenidos por el criterio de mínimos cuadrados se denominan **estimadores de mínimos cuadrados de los parámetros de regresión** β, que se denotan por $\widehat{\beta}$.
- Entonces a partir de los coeficientes $\widehat{\beta}$ podemos estimar o ajustar las respuestas observadas y_j, dados los valores observados de $z_{j1}, ..., z_{jr}$, es decir

$$\widehat{y_j} = \widehat{\beta_0} + \widehat{\beta_1} z_{j1} + \cdots + \widehat{\beta_r} z_{jr}$$

- Las desviaciones

$$\widehat{\varepsilon_j} = y_j - \widehat{y_j} = y_j - (\widehat{\beta_0} + \widehat{\beta_1} z_{j1} + \cdots + \widehat{\beta_r} z_{jr}), j = 1, ..., n$$

  son llamados los *residuales*.
- El vector de residuales $\widehat{\varepsilon} = \boldsymbol{y} - \boldsymbol{Z}\widehat{\beta}$ contiene la información sobre el parámetro desconocido restante σ^2

Resultado: Sea $\boldsymbol{Z}$ de rango completo $r+1 \leq n$. Los estimadores de mínimos cuadrados de β están dados por:

$$\widehat{\beta} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$$

Sea $\widehat{\boldsymbol{y}} = \boldsymbol{Z}\widehat{\beta} = \boldsymbol{H}\boldsymbol{y}$ denota los *valores ajustados* de $\boldsymbol{y}$, donde $\boldsymbol{H} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. Entonces los residuales

$$\widehat{\varepsilon} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$$

satisfacen $\boldsymbol{Z}'\widehat{\varepsilon} = \boldsymbol{0}$ y $\widehat{\boldsymbol{y}}'\widehat{\varepsilon} = 0$. También, la

$$\text{suma de cuadrados de residuales} = \sum_{j=1}^{n}[y_j - (\widehat{\beta}_0 + \widehat{\beta}_1 z_{j1} + \cdots - \widehat{\beta}_r z_{jr})]^2 = \widehat{\varepsilon}'\widehat{\varepsilon}$$

$$= \widehat{\boldsymbol{y}}'[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}']\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{Z}\widehat{\beta}$$

Ejemplo: calculando los estimadores de mínimos cuadrados

Calcula los estimadores de mínimos cuadrados $\widehat{\beta}$, los residuales $\widehat{\varepsilon}$ y la suma de cuadrados de residuales para un modelo de línea recta

$$Y_j = \beta_0 + \beta_1 z_{j1} + \varepsilon_j$$

ajustado a los datos

z_1	0	1	2	3	4
y	1	4	3	8	9

Descomposición de la suma de cuadrados

- De acuerdo al resultado anterior $\widehat{\boldsymbol{y}}'\widehat{\varepsilon} = 0$, la suma total de las respuestas al cuadrado $\boldsymbol{y}'\boldsymbol{y} = \sum_{j=1}^{n} y_i^2$ satisface

$$\boldsymbol{y}'\boldsymbol{y} = (\widehat{\boldsymbol{y}} + \boldsymbol{y} - \widehat{\boldsymbol{y}})'(\widehat{\boldsymbol{y}} + \boldsymbol{y} - \widehat{\boldsymbol{y}}) = (\widehat{\boldsymbol{y}} + \widehat{\varepsilon})'(\widehat{\boldsymbol{y}} + \widehat{\varepsilon}) = \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + \widehat{\varepsilon}'\widehat{\varepsilon}$$

- Ya que la primer columna de $\boldsymbol{Z}$ es $\mathbf{1}$, la condición $\boldsymbol{Z}'\widehat{\varepsilon} = \mathbf{0}$ implica que

$$0 = \mathbf{1}'\widehat{\varepsilon} = \sum_{j=1}^{n} y_j - \sum_{j=1}^{n} \widehat{y}_j \quad \text{ó} \quad \overline{y} = \overline{\widehat{y}}$$

- Restando $n\overline{y}^2 = n\overline{\widehat{y}}^2$ de ambos lados de la descomposición de arriba, se obtiene la descomposición básica de la suma de cuadrados alrededor de la media:

$$\boldsymbol{y}'\boldsymbol{y} - n\overline{y}^2 = \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} - n\overline{\widehat{y}}^2 + \widehat{\varepsilon}'\widehat{\varepsilon}$$

ó

$$\sum_{j=1}^{n} (y_i - \overline{y})^2 = \sum_{j=1}^{n} (\widehat{y}_j - \overline{y})^2 + \sum_{j=1}^{n} \widehat{\varepsilon}_j^2$$

Descomposición de la suma de cuadrados

$$\sum_{j=1}^{n}(y_i-\overline{y})^2=\sum_{j=1}^{n}(\widehat{y}_j-\overline{y})^2+\sum_{j=1}^{n}\widehat{\varepsilon}_j^2$$

$$\left(\begin{array}{c}\text{suma total de cuadrados}\\ \text{alrededor}\\ \text{de la media}\end{array}\right)=\left(\begin{array}{c}\text{suma de cuadrados}\\ \text{de la regresión}\end{array}\right)+\left(\begin{array}{c}\text{suma de cuadrados}\\ \text{del residual(error)}\end{array}\right)$$

- La calidad del ajuste de los modelos se puede medir por *el coeficiente de determinación*

$$R^2=1-\frac{\sum_{j=1}^{n}\widehat{\varepsilon}_j^2}{\sum_{j=1}^{n}(y_i-\overline{y})^2}=\frac{\sum_{j=1}^{n}(\widehat{y}_j-\overline{y})^2}{\sum_{j=1}^{n}(y_i-\overline{y})^2}$$

- R^2 determina la cantidad de la varianza total en las $y_j^{\prime}s$ explicada o atribuida a las $z_1, z_2, \ldots, z_r$
- Si $R^2=1$ la ecuacion ajustada pasa a través de todos los puntos, así $\widehat{\varepsilon}_j=0$ para todo j
- Si $R^2=0$ entonces $\widehat{\beta}_0=\overline{y}$ y todos los $\widehat{\beta}_1=\widehat{\beta}_2=\cdots=\widehat{\beta}_r=0$

Propiedades muestrales de los estimadores clásicos de mínimos cuadrados

Resultado: Bajo el modelo de regresión lineal general, los estimadores de mínimos cuadrados $\widehat{\beta} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$ cumplen

$$E(\widehat{\beta}) = \beta \quad \text{y} \quad Cov(\widehat{\beta}) = \sigma^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$$

Los residuales $\widehat{\varepsilon}$ tienen las siguientes propiedades:

$$E(\widehat{\varepsilon}) = \boldsymbol{0} \quad \text{y} \quad Cov(\widehat{\varepsilon}) = \sigma^2[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-\boldsymbol{1}}\boldsymbol{Z}'] = \sigma^2[\boldsymbol{I} - \boldsymbol{H}]$$

También $E(\widehat{\varepsilon}'\widehat{\varepsilon}) = (n-r-1)\sigma^2$, definiendo

$$s^2 = \frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{n-(r+1)} = \frac{\widehat{\boldsymbol{y}}'[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-\boldsymbol{1}}\boldsymbol{Z}']\boldsymbol{y}}{n-r-1} = \frac{\boldsymbol{y}'[\boldsymbol{I} - \boldsymbol{H}]\boldsymbol{y}}{n-r-1}$$

Tenemos que

$$E(s^2) = \sigma^2$$

Además, $\widehat{\beta}$ y $\widehat{\varepsilon}$ no están correlacionados

- Para hacer inferencia sobre el modelo de regresión es necesario determinar las distribuciones muestrales de $\widehat{\beta}$ y de la suma de los cuadrados de residuales, $\widehat{\varepsilon}'\widehat{\varepsilon}$

Resultado: Sea $\boldsymbol{Y} = \boldsymbol{Z}\beta + \varepsilon$, donde $\boldsymbol{Z}$ tiene rango completo $r+1$ y $\varepsilon \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Entonces el estimador de máxima de verosimilitud de β es el mismo que el estimador de mínimos cuadrados $\widehat{\beta}$. Mas aún

$$\widehat{\beta} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} \sim N_{r+1}(\beta, \sigma^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1})$$

Y se distribuye independientemente de los residuales $\widehat{\varepsilon} = \boldsymbol{y} - \boldsymbol{Z}\widehat{\beta}$. Además

$$n\widehat{\sigma}^2 = \widehat{\varepsilon}'\widehat{\varepsilon} \quad \text{se distribuye como} \quad \sigma^2\chi^2_{n-r-1}$$

donde $\widehat{\sigma}^2$ es el estimador de máxima verosimilitud de σ^2

A partir del resultado anterior se puede construir fácilmente un elipsoide de confianza para β. Éste se expresa en términos de la matriz de covarianza estimada $s^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$, donde $s^2 = \frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{n-(r+1)}$.

Resultado: Sea $\boldsymbol{Y} = \boldsymbol{Z}\beta + \varepsilon$, donde $\boldsymbol{Z}$ tiene rango completo $r+1$ y $\varepsilon \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

- Entonces una region de confianza del $100(1-\alpha)\%$ para β esta dado por

$$(\widehat{\beta}-\beta)'\boldsymbol{Z}'\boldsymbol{Z}(\widehat{\beta}-\beta) \leq (r+1)s^2 F_{r+1,n-r-1}(\alpha)$$

  donde $F_{r+1,n-r-1}(\alpha)$ es el percentil superior del $(100\alpha)\%$ de una distribucion F con r+1 y n-r-1 grados de libertad.

- También los intervalos de confianza simultáneos del $100(1-\alpha)\%$ para las β_i están dados por

$$\widehat{\beta}_i \pm \sqrt{\widehat{Var}(\widehat{\beta}_i)}\sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}, \quad i = 0, 1, ..., r$$

donde $\widehat{Var}(\widehat{\beta}_i)$ es el i-ésimo elemento de la diagonal de $s^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ correspondientes a $\widehat{\beta}_i$

- Frecuentemente en la práctica se ignora la propiedad de *confianza simultánea* en las estimaciones de los intervalos para las $\beta_i's$
- Entonces se reemplaza $(r+1)F_{r+1,n-r-1}(\alpha)$ por el t valor, $t_{n-r-1}(\alpha/2)$, que es el percentil de una t asumiendo que los intervalos para cada β_i se obtienen de manera univariada

Pruebas de la razón de verosimilitud para los parámetros de regresión

- Parte del análisis de regresión tiene que ver con la evaluación de los efectos de determinadas variables predictoras sobre la variable respuesta.
- Una hipótesis nula de interés representaría algunos de los $z_i^{'}s$ que se creen no influyen en la respuesta Y.
- Estos predictores serán etiquetados como $z_{q+1}, z_{q+2}, ..., z_r$. Entonces la declaración de que $z_{q+1}, z_{q+2}, ..., z_r$ no influyen en Y se traduce en la hipótesis estadística

$$H_0 : \beta_{q+1} = \beta_{q+2} = \cdots = \beta_r = 0 \quad \text{ó} \quad H_0 : \beta_{(2)} = \mathbf{0}$$

donde $\beta^{'}_{(2)} = [\beta_{q+1}, \beta_{q+2}, ..., \beta_r]$

Pruebas de la razón de verosimilitud para los parámetros de regresión

Particionando

$$\boldsymbol{Z} = \left[\; \boldsymbol{Z_1}_{n\times(q+1)} \; \middle| \; \boldsymbol{Z_2}_{n\times(r-q)} \; \right], \quad \beta = \left[\frac{\beta_{(\mathbf{1})(q+1)\times 1}}{\beta_{(\mathbf{2})(r-q)\times 1}} \right]$$

podemos expresar el modelo lineal general como

$$\boldsymbol{Y} = \boldsymbol{Z}\beta + \varepsilon = \left[\; \boldsymbol{Z_1} \; \middle| \; \boldsymbol{Z_2} \; \right] \left[\frac{\beta_{(\mathbf{1})}}{\beta_{(\mathbf{2})}} \right] + \varepsilon = \boldsymbol{Z}_1\beta_{(1)} + \boldsymbol{Z}_2\beta_{(2)} + \varepsilon$$

Bajo la hipótesis nula $H_0 : \beta_{(2)} = \mathbf{0}$, $\boldsymbol{Y} = \boldsymbol{Z}_1\beta_{(1)} + \varepsilon$. La prueba de la razón de verosimilitud de H_0 está basada sobre

$$\text{Suma extra de cuadrados} = SS_{res}(\boldsymbol{Z_1}) - SS_{res}(\boldsymbol{Z})$$

$$= (\boldsymbol{y} - \boldsymbol{Z_1}\widehat{\beta}_{(1)})'(\boldsymbol{y} - \boldsymbol{Z_1}\widehat{\beta}_{(1)}) - (\boldsymbol{y} - \boldsymbol{Z}\widehat{\beta})'(\boldsymbol{y} - \boldsymbol{Z}\widehat{\beta})$$

donde $\widehat{\beta}_{(1)} = (\boldsymbol{Z_1'Z_1})^{-1}\boldsymbol{Z_1'y}$

Pruebas de la razón de verosimilitud para los parámetros de regresión

Resultado: Sea $\boldsymbol{Z}$ de rango completo $r+1$ y $\varepsilon \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I})$. La prueba del cociente de verosimilitud de $H_0 : \beta_{(2)} = \mathbf{0}$ es equivalente a probar H_0 basada en la *Suma extra de cuadrados* y $s^2 = (\boldsymbol{y} - \boldsymbol{Z}\widehat{\beta})'(\boldsymbol{y} - \boldsymbol{Z}\widehat{\beta})/(n-r-1)$. En particular, la prueba de la razón de verosimilitud rechaza H_0 si

$$\frac{(SS_{res}(\boldsymbol{Z_1}) - SS_{res}(\boldsymbol{Z}))/(r-q)}{s^2} > F_{r-q,n-r-1}(\alpha)$$

donde $F_{r-q,n-r-1}(\alpha)$ es el percentil superior del $(100\alpha)\%$ de una distribucion F con r-q y n-r-1 grados de libertad.

Ejemplo: Pruebas de la razón de verosimilitud para los parámetros de regresión

Las empresas que consideran la compra de una computadora primero deben evaluar sus necesidades futuras para determinar el equipo apropiado a comprar. Un informático recopiló datos de siete sites de empresas similares para que se pudiera desarrollar una ecuación que pronostique los requerimientos de hardware para la gestión de inventario

Los datos están dados en la siguiente tabla

z_1 (Orders)	z_2 (Add–delete items)	Y (CPU time)
123.5	2.108	141.5
146.1	9.213	168.9
133.9	1.905	154.8
128.5	.815	146.5
151.5	1.061	172.8
136.2	8.603	160.1
92.0	1.125	108.5

Source: Data taken from H. P. Artis, *Forecasting Computer Requirements: A Forecaster's Dilemma* (Piscataway, NJ: Bell Laboratories, 1979).

Ejemplo: Pruebas de la razón de verosimilitud para los parámetros de regresión

$$z_1 = \text{número de pedidos de clientes (en miles)}$$
$$z_2 = \text{agregar o quitar numero de articulos(en miles)}$$
$$Y = \text{Tiempo de procesamiento, CPU (en horas)}$$

- Se quiere probar la hipótesis de que la variable z_2 no tiene efectos en el tiempo de procesamiento
- O equivalentemente se quiere probar que $H_0 : \beta_2 = 0$ con un nivel de significancia de $\alpha = .05$

Regresión Lineal Multivariada

- Consideremos la extension multivariada de la regresión lineal multiple, que modela la relación entre m respuestas $Y_1, \ldots, Y_m$ y un conjunto de r variables predictoras $z_1, \ldots, z_r$.
- Se asume que cada una de las respuestas sigue su propio modelo de regresión, i.e.,

$$
\begin{aligned}
Y_1 &= B_{01} + B_{11} z_1 + B_{21} z_2 + \cdots + B_{r1} z_r + \varepsilon_1 \\
Y_2 &= B_{02} + B_{12} z_1 + B_{22} z_2 + \cdots + B_{r2} z_r + \varepsilon_2 \\
&\vdots \\
Y_m &= B_{0m} + B_{1m} z_1 + B_{2m} z_2 + \cdots + B_{rm} z_r + \varepsilon_m
\end{aligned}
$$

donde

$$
E(\varepsilon) = E \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} = \mathbf{0}, \; Var(\varepsilon) = \mathbf{\Sigma}
$$

Conceptualmente, podemos dejar

$$[z_{j0}, z_{j1}, \ldots, z_{jr}]$$

que denotan los valores de las variables predictoras para la j-ésima prueba y

$$\boldsymbol{Y}_j = \begin{bmatrix} Y_{j1} \\ Y_{j2} \\ \vdots \\ Y_{jm} \end{bmatrix}, \varepsilon_j = \begin{bmatrix} \varepsilon_{j1} \\ \varepsilon_{j2} \\ \vdots \\ \varepsilon_{jm} \end{bmatrix}$$

son las respuestas de todas las variables y errores para la j-ésima prueba. La matriz de diseño $\boldsymbol{Z}$ tiene la misma forma, $n \times (r+1)$

$$\boldsymbol{Z} = \begin{bmatrix} z_{10} & z_{11} & \cdots & z_{1r} \\ z_{20} & z_{21} & \cdots & z_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n0} & z_{n1} & \cdots & z_{nr} \end{bmatrix}$$

Los componentes matriciales en el caso del modelo de regresión multivariada son

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nm} \end{bmatrix} = \left[\ \boldsymbol{Y}_{(1)} \ \middle| \ \boldsymbol{Y}_{(2)} \ \middle| \ \cdots \ \middle| \ \boldsymbol{Y}_{(m)} \ \right]$$

$$\beta = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rm} \end{bmatrix} = \left[\ \beta_{(1)} \ \middle| \ \beta_{(2)} \ \middle| \ \cdots \ \middle| \ \beta_{(m)} \ \right]$$

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nm} \end{bmatrix} = \left[\ \varepsilon_{(1)} \ \middle| \ \varepsilon_{(2)} \ \middle| \ \cdots \ \middle| \ \varepsilon_{(m)} \ \right] = \begin{bmatrix} \varepsilon'_{(1)} \\ \text{—} \\ \varepsilon'_{(2)} \\ \text{—} \\ \vdots \\ \text{—} \\ \varepsilon'_{(n)} \end{bmatrix}$$

El modelo de regresion lineal multivariado es

$$\boldsymbol{Y} = \boldsymbol{Z}\beta + \varepsilon$$

con

$$E(\varepsilon_{(i)}) = \boldsymbol{0} \text{ y } Cov(\varepsilon_{(i)}, \varepsilon_{(k)}) = \sigma_{ik}\boldsymbol{I} \quad i, k = 1, \ldots, m$$

- Notemos también que las m respuestas observadas sobre la j-esima prueba tiene matriz de covarianzas $\boldsymbol{\Sigma} = \{\sigma_{ik}\}$, pero observaciones de diferentes pruebas no son correlacionadas.
- β y σ_{ik} son los parámetros desconocidos, la matriz $\boldsymbol{Z}$ tiene en la j ésima fila $[z_{j0}, z_{j1}, ..., z_{jr}]$

Regresión Lineal Multivariada: Principios básicos

- Dicho de manera sencilla, la i-ésima respuesta $Y_{(i)}$ sigue el modelo de regresión lineal

$$\boldsymbol{Y}_{(i)} = \boldsymbol{Z}\beta_{\boldsymbol{(i)}} + \varepsilon_{(i)}, \quad i = 1, 2, \ldots, m$$

con $Cov(\varepsilon_{(i)}) = \sigma_{ii}\boldsymbol{I}$.

- Sin embargo, los errores para diferentes respuestas en la misma prueba pueden ser correlacionados.

Dados los valores de la matriz de respuesta $\boldsymbol{Y}$ y los valores de las variables predictoras $\boldsymbol{Z}$, determinamos los estimadores de mínimos cuadrados $\hat{\beta}_{(i)}$ exclusivamente de las observaciones de la i ésima respuesta, $\boldsymbol{Y_i}$. Entonces

$$\hat{\beta}_{\boldsymbol{(i)}} = \boldsymbol{(Z'Z)^{-1}Z'Y_{(i)}}$$

coleccionando los estimadores univariados de mínimos cuadrados obtenemos

$$
\begin{aligned}
\hat{\beta} &= \left[\ \hat{\beta}_{(1)} \ \middle| \ \hat{\beta}_{(2)} \ \middle| \ \cdots \ \middle| \ \hat{\beta}_{(m)} \ \right] \\
&= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}' \left[\ \boldsymbol{Y}_{(1)} \ \middle| \ \boldsymbol{Y}_{(2)} \ \middle| \ \cdots \ \middle| \ \boldsymbol{Y}_{(m)} \ \right] \\
&= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}
\end{aligned}
$$

Ahora para cualquier elección de parámetros de regresión

$$
\boldsymbol{B} = \left[\ \boldsymbol{b}_{(1)} \ \middle| \ \boldsymbol{b}_{(2)} \ \middle| \ \cdots \ \middle| \ \boldsymbol{b}_{(m)} \ \right]
$$

La matriz de errores resultante es

$$
\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{B}
$$

La matriz de suma de errores al cuadrado y productos cruzados resultantes es :

$$(\boldsymbol{Y}-\boldsymbol{ZB})'(\boldsymbol{Y}-\boldsymbol{ZB})=\begin{bmatrix} (Y_{(\mathbf{1})}-Zb_{(\mathbf{1})})'(Y_{(\mathbf{1})}-Zb_{(\mathbf{1})}) & \cdots & (Y_{(\mathbf{1})}-Zb_{(\mathbf{1})})'(Y_{(m)}-Zb_{(m)}) \\ \vdots & \ddots & \vdots \\ (Y_{(m)}-Zb_{(m)})'(Y_{(\mathbf{1})}-Zb_{(\mathbf{1})}) & \cdots & (Y_{(m)}-Zb_{(m)})'(Y_{(m)}-Zb_{(m)}) \end{bmatrix}$$

Podemos demostrar que la selecccion $\boldsymbol{b}_{(i)}=\widehat{\beta}_{(i)}$ minimiza la suma de cuadrados del i esimo elemento de la diagonal, dada por

$$(\boldsymbol{Y}_{(i)}-\boldsymbol{Zb}_{(i)})'(\boldsymbol{Y}_{(i)}-\boldsymbol{Zb}_{(i)})$$

Por tanto

$$tr[(\boldsymbol{Y}-\boldsymbol{ZB})'(\boldsymbol{Y}-\boldsymbol{ZB})]$$

se minimiza cuando se elige $\boldsymbol{B}=\widehat{\beta}$.

Regresión Lineal Multivariada: Principios básicos

Entonces usando los estimadores de mínimo cuadrados $\widehat{\beta}$ se puede formar la matriz de valores predichos

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

y la matriz de residuales

$$\hat{\varepsilon} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \left[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\right]\boldsymbol{Y}$$

Notemos que las condiciones de ortogonalidad entre residuales, valores predichos, y columnas de la matriz de diseño $\boldsymbol{Z}$ que se cumplen en el caso univariado también se cumplen en el caso multivariado. Esto debido a que

$$\boldsymbol{Z}'\left[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\right] = \boldsymbol{Z}' - \boldsymbol{Z}' = 0$$

Lo cual significa que los residuales son perpendiculares a las columnas de la matriz de diseño $\boldsymbol{Z}$, es decir,

$$\boldsymbol{Z}'\widehat{\varepsilon} = \boldsymbol{Z}'\left[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\right]\boldsymbol{Y} = 0$$

y son perpendiculares a los valores predichos $\hat{\boldsymbol{Y}}$

$$\hat{\boldsymbol{Y}}'\hat{\varepsilon} = \hat{\beta}'\boldsymbol{Z}'\left[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\right]\boldsymbol{Y} = \boldsymbol{0}$$

Además, debido a que

$$\boldsymbol{Y} = \hat{\boldsymbol{Y}} + \hat{\varepsilon}$$

$$\boldsymbol{Y}'\boldsymbol{Y} = (\hat{\boldsymbol{Y}} + \hat{\varepsilon})'(\hat{\boldsymbol{Y}} + \hat{\varepsilon}) = \widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{Y}} + \widehat{\varepsilon}'\widehat{\varepsilon}$$

$$\left(\begin{array}{c}\text{Suma total de cuadrados}\\ \text{y productos cruzados}\end{array}\right) = \left(\begin{array}{c}\text{Sumas de cuadrados}\\ \text{de predichos y}\\ \text{productos cruzados}\end{array}\right) + \left(\begin{array}{c}\text{Suma de cuadrados}\\ \text{de residuales (errores)}\\ \text{y productos cruzados}\end{array}\right)$$

La suma de cuadrados de residuales y productos cruzados también se puede escribir como

$$\widehat{\varepsilon}'\widehat{\varepsilon} = \boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{Y}} = \boldsymbol{Y}'\boldsymbol{Y} - \widehat{\beta}'\boldsymbol{Z}'\boldsymbol{Z}\widehat{\beta}$$

Ejemplo 1: Regresión Lineal Multivariada

Supongamos que tenemos las siguientes seis observaciones muestrales sobre dos variables independientes (apetitosidad y textura) y dos variables dependientes (calidad total e intención de compra):

Palatability	Texture	Overall Quality	Purchase Intent
65	71	63	67
72	77	70	70
77	73	72	70
68	78	75	72
81	76	89	88
73	87	76	77

Usar estos datos para estimar el modelo de regresión lineal multivariado, para el cual palatability y textura son variables independientes mientras que calidad total e intención de compra son las variables dependientes.

Ejemplo 1: Regresión Lineal Multivariada

Queremos estimar conjuntamente

$$Y_1 = B_{01} + B_{11} z_1 + B_{21} z_2 + \varepsilon_1$$

y

$$Y_2 = B_{02} + B_{12} z_1 + B_{22} z_2 + \varepsilon_2$$

La matriz de diseño es

$$z = \begin{bmatrix} 1 & 65 & 71 \\ 1 & 72 & 77 \\ 1 & 77 & 73 \\ 1 & 68 & 78 \\ 1 & 81 & 76 \\ 1 & 73 & 87 \end{bmatrix}$$

Entonces

$$Z'Z = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 65 & 72 & 77 & 68 & 81 & 73 \\ 71 & 77 & 73 & 78 & 76 & 87 \end{bmatrix} \begin{bmatrix} 1 & 65 & 71 \\ 1 & 72 & 77 \\ 1 & 77 & 73 \\ 1 & 68 & 78 \\ 1 & 81 & 76 \\ 1 & 73 & 87 \end{bmatrix} = \begin{bmatrix} 6 & 436 & 462 \\ 436 & 31852 & 33591 \\ 462 & 33591 & 35728 \end{bmatrix}$$

$$(Z'Z)^{-1} = \begin{bmatrix} 6 & 436 & 462 \\ 436 & 31852 & 33591 \\ 462 & 33591 & 35728 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 62.560597030 & -0.378268027 & -0.453330568 \\ -0.378268027 & 0.005988412 & -0.000738830 \\ -0.453330568 & -0.000738830 & 0.006584661 \end{bmatrix}$$

Ejemplo 1: Regresión Lineal Multivariada

$$Z\,'y_{(1)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 65 & 72 & 77 & 68 & 81 & 73 \\ 71 & 77 & 73 & 78 & 76 & 87 \end{bmatrix} \begin{bmatrix} 63 \\ 70 \\ 72 \\ 75 \\ 89 \\ 76 \end{bmatrix} = \begin{bmatrix} 445 \\ 32536 \\ 34345 \end{bmatrix}$$

Así

$$\begin{aligned} \hat{\beta}_{(1)} &= (Z\,'Z)^{-1} Z\,'y_{(1)} \\ &= \begin{bmatrix} 62.560597030 & -0.378268027 & -0.453330568 \\ -0.378268027 & 0.005988412 & -0.000738830 \\ -0.453330568 & -0.000738830 & 0.006584661 \end{bmatrix} \begin{bmatrix} 445 \\ 32536 \\ 34345 \end{bmatrix} \\ &= \begin{bmatrix} -37.501205460 \\ 1.134583728 \\ 0.379499410 \end{bmatrix} \end{aligned}$$

Ejemplo 1: Regresión Lineal Multivariada

$$Z'y_{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 65 & 72 & 77 & 68 & 81 & 73 \\ 71 & 77 & 73 & 78 & 76 & 87 \end{bmatrix} \begin{bmatrix} 67 \\ 70 \\ 70 \\ 72 \\ 88 \\ 77 \end{bmatrix} = \begin{bmatrix} 444 \\ 32430 \\ 34260 \end{bmatrix}$$

Así

$$\begin{aligned} \hat{\beta}_{(2)} &= (Z'Z)^{-1} Z' y_{(2)} \\ &= \begin{bmatrix} 62.560597030 & -0.378268027 & -0.453330568 \\ -0.378268027 & 0.005988412 & -0.000738830 \\ -0.453330568 & -0.000738830 & 0.006584661 \end{bmatrix} \begin{bmatrix} 444 \\ 32430 \\ 34260 \end{bmatrix} \\ &= \begin{bmatrix} -21.432293350 \\ 0.940880634 \\ 0.351449792 \end{bmatrix} \end{aligned}$$

Ejemplo 1: Regresión Lineal Multivariada

Así

$$\hat{\beta} = \left[\begin{array}{c|c} \hat{\beta}_{(1)} & \hat{\beta_{(2)}} \end{array} \right] = \begin{bmatrix} -37.501205460 & -21.432293350 \\ 1.134583728 & 0.940880634 \\ 0.379499410 & 0.351449792 \end{bmatrix}$$

Esto nos da la matriz de valores estimados

$$\hat{\boldsymbol{Y}} = \boldsymbol{Z}\widehat{\beta} = \begin{bmatrix} 1 & 65 & 71 \\ 1 & 72 & 77 \\ 1 & 77 & 73 \\ 1 & 68 & 78 \\ 1 & 81 & 76 \\ 1 & 73 & 87 \end{bmatrix} \begin{bmatrix} -37.501205460 & -21.432293350 \\ 1.134583728 & 0.940880634 \\ 0.379499410 & 0.351449792 \end{bmatrix} = \begin{bmatrix} 63.19119 & 64.67788 \\ 73.41028 & 73.37275 \\ 77.56520 & 76.67135 \\ 69.25144 & 69.96067 \\ 83.24203 & 81.48922 \\ 78.33986 & 77.82812 \end{bmatrix}$$

Y la matriz de residuales

$$\hat{\varepsilon} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \begin{bmatrix} 63 & 67 \\ 70 & 70 \\ 72 & 70 \\ 75 & 72 \\ 89 & 88 \\ 76 & 77 \end{bmatrix} - \begin{bmatrix} 63.19119 & 64.67788 \\ 73.41028 & 73.37275 \\ 77.56520 & 76.67135 \\ 69.25144 & 69.96067 \\ 83.24203 & 81.48922 \\ 78.33986 & 77.82812 \end{bmatrix} = \underbrace{\begin{bmatrix} 0.191194690 & -2.322116943 \\ 3.410277515 & 3.372746244 \\ 5.565198512 & 6.6711350244 \\ -5.748557985 & -2.039326498 \\ -5.757968347 & -6.510777845 \\ 2.339855345 & 0.828124797 \end{bmatrix}}_{\text{Notemos que cada columna suma cero!}}$$

Los estimadores de mínimos cuadrados

$$\widehat{\beta} = \left[\ \beta_{\mathbf{(1)}} \ \middle|\ \beta_{\mathbf{(2)}} \ \middle|\ \cdots \ \middle|\ \beta_{\boldsymbol{(m)}}\ \right]$$

del modelo de regresión multivariada, con $\boldsymbol{Z}$ de rango completo $r+1<n$, tienen las siguientes propiedades:

- $E(\hat{\beta}_{\boldsymbol{(i)}}) = \beta_{(i)}$ ó $E(\hat{\beta}) = \beta$
- $Cov(\hat{\beta}_{\boldsymbol{(i)}}, \hat{\beta}_{\boldsymbol{(k)}}) = \sigma_{ik}(\boldsymbol{Z}'\boldsymbol{Z})^{-1} \quad i,k = 1,\ldots,m$

Los residuales

$$\hat{\varepsilon} = \left[\ \widehat{\varepsilon}_{(1)} \ \middle|\ \widehat{\varepsilon}_{(2)} \ \middle|\ \cdots \ \middle|\ \widehat{\varepsilon}_{(m)}\ \right] = \boldsymbol{Y} - \boldsymbol{Z}\widehat{\beta}$$

satisfacen

- $E(\widehat{\varepsilon}_{\boldsymbol{(i)}}) = \mathbf{0}$ y $E(\hat{\varepsilon_{\boldsymbol{(i)}}}' \hat{\varepsilon_{\boldsymbol{(k)}}}) = (n-r-1)\sigma_{ik}$,y así

$$E(\widehat{\varepsilon}) = \mathbf{0} \text{ y } E\left(\frac{1}{n-r-1}\widehat{\varepsilon}'\widehat{\varepsilon}\right) = \boldsymbol{\Sigma}$$

Notemos que $\widehat{\varepsilon}$ y $\widehat{\beta}$ no están correlacionados.

- Lo anterior significa que cuando las variables predictoras tienen los valores $\boldsymbol{z_0} = [1, z_{01}, ..., z_{0r}]$, la media de la respuesta de la i-esima variable es $\boldsymbol{z}_0^{'}\boldsymbol{\beta_{(i)}}$, y esta se estima por $\boldsymbol{z}_0^{'}\widehat{\boldsymbol{\beta}}_{\boldsymbol{(i)}}$.
- De manera colectiva,

$$\boldsymbol{z}_0^{'}\hat{\beta} = \boldsymbol{z_0}^{'}\left[\begin{array}{c|c|c|c} \hat{\beta}_{(1)} & \hat{\beta}_{(2)} & \cdots & \hat{\beta}_{\boldsymbol{(m)}} \end{array}\right] = \left[\begin{array}{c|c|c|c} \boldsymbol{z_0}{}'\hat{\beta}_{(1)} & \boldsymbol{z_0}{}'\hat{\beta}_{(2)} & \cdots & \boldsymbol{z_0}{}'\hat{\beta}_{\boldsymbol{(m)}} \end{array}\right]$$

es un estimador insesgado de $\boldsymbol{z_0}^{'}\beta$, i.e.,

$$E(\boldsymbol{z}_0^{'}\hat{\beta}) = \boldsymbol{z}_0^{'}\beta$$

- Esto debido a que

$$E(\boldsymbol{z}_0^{'}\hat{\beta}_{\boldsymbol{(i)}}) = \boldsymbol{z}_0^{'}E(\widehat{\beta}_{(i)}) = \boldsymbol{z}_0^{'}\beta_{\boldsymbol{(i)}}$$

para cada componente

Estimadores de máxima verosimilitud y sus distribuciones se pueden obtener cuando los errores ε tienen una distribución normal multivariada

Resultado: Sea $\boldsymbol{Y} = \boldsymbol{Z}\beta + \varepsilon$ el modelo de regresión multivariada, con $\boldsymbol{Z}$ de rango completo $r+1$, $n \geq (r+1)+m$, y ε sigue una distribución normal multivariada. Entonces

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

es el estimador de máxima verosimilitud de β y

$$\hat{\beta} \sim \boldsymbol{N}(\beta, \boldsymbol{\Sigma})$$

donde los elementos de $\boldsymbol{\Sigma}$ son

$$Cov(\hat{\beta}_{\boldsymbol{(i)}}, \hat{\beta}_{\boldsymbol{(k)}}) = \sigma_{ik}(\boldsymbol{Z}\,'\boldsymbol{Z})^{-1}, \quad i,k = 1,\ldots,m$$

También, el estimador de máxima verosimilitud $\hat{\beta}$ es independiente del estimador de máxima verosimilitud de la matriz positiva definida $\boldsymbol{\Sigma}$, el cual está dada por

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\widehat{\varepsilon}\,'\widehat{\varepsilon} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{Z}\hat{\beta})'(\boldsymbol{Y} - \boldsymbol{Z}\hat{\beta})$$

y

$$n\hat{\boldsymbol{\Sigma}} \sim W_{n-r-1}(\boldsymbol{\Sigma})$$

El maximo de la verosimilitud es

$$L(\hat{\beta}, \hat{\boldsymbol{\Sigma}}) = (2\pi)^{-nm/2}|\widehat{\boldsymbol{\Sigma}}|^{-n/2}e^{nm/2}$$

- El modelo multivariado de regresión múltiple no plantea nuevos problemas computacionales. Los estimadores de mínimos cuadrados (o máxima verosimilitud) $\hat{\beta}_{(i)} = (\boldsymbol{Z}'\boldsymbol{Z})^{-\mathbf{1}}\boldsymbol{Z}'\boldsymbol{y}_{(\boldsymbol{i})}$ se calculan individualmente para cada variable de respuesta.
- Obsérvese, sin embargo, que el modelo requiere que se usen *las mismas variables predictoras* para todas las respuestas.

Estos resultados se pueden usar para desarrollar pruebas de razón de verosimilitud para probar hipotesis sobre los parámetros de la regresión multivariada.

La hipótesis de que las respuestas no dependen de las variables predictoras $z_{q+1}, z_{q+2}, \ldots, z_r$ es

$$H_0 : \beta_{(2)} = \mathbf{0} \qquad donde\, \beta = \begin{bmatrix} \boldsymbol{\beta}_{\mathbf{(1)}\,(q+1)\times m} \\ - \\ \boldsymbol{\beta}_{\mathbf{(2)}\,(r-q)\times m} \end{bmatrix}$$

Si particionamos $\boldsymbol{Z}$ de una manera similar

$$\boldsymbol{Z} = \left[\; \boldsymbol{Z}_{\mathbf{(1)}\,n\times(q+1)} \quad | \quad \boldsymbol{Z}_{\mathbf{(2)}\,n\times(r-q)} \; \right]$$

podemos escribir el modelo general como

$$\boldsymbol{Y} = \boldsymbol{Z}\beta + \varepsilon = \left[\; \boldsymbol{Z}_{\mathbf{1}} \quad | \quad \boldsymbol{Z}_2 \; \right] \begin{bmatrix} \beta_{(1)} \\ - \\ \beta_{(2)} \end{bmatrix} + \varepsilon = \boldsymbol{Z}_{\mathbf{1}}\boldsymbol{\beta}_{\mathbf{(1)}+}\boldsymbol{Z}_{\mathbf{2}}\boldsymbol{\beta}_{\mathbf{(2)}} + \varepsilon$$

Bajo $H_0 : \beta_{(2)} = \mathbf{0}$, $\boldsymbol{Y} = \boldsymbol{Z_1}\beta_{\mathbf{(1)}} + \varepsilon$ y la prueba de la razón de verosimilitud de H_0 se basa en la suma de cuadrados y productos cruzados extra dada por

$$\left(\boldsymbol{Y} - \boldsymbol{Z_1}\hat{\beta}_{\mathbf{(1)}}\right)'\left(\boldsymbol{Y} - \boldsymbol{Z_1}\hat{\beta}_{\mathbf{(1)}}\right) - \left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\beta}\right)'\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\beta}\right) = n\left(\hat{\boldsymbol{\Sigma}}_{\mathbf{1}} - \hat{\boldsymbol{\Sigma}}\right)$$

donde

$$\hat{\beta}_{\mathbf{(1)}} = \left(\boldsymbol{Z_1}'\boldsymbol{Z_1}\right)^{-\mathbf{1}}\boldsymbol{Z_1}'\boldsymbol{Y}$$

y

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{1}} = n^{-1}\left(\boldsymbol{Y} - \boldsymbol{Z_1}\hat{\beta}_{\mathbf{(1)}}\right)'\left(\boldsymbol{Y} - \boldsymbol{Z_1}\hat{\beta}_{\mathbf{(1)}}\right)$$

El cociente de verosimilitud para probar la hipótesis

$$H_0 : \beta_{(2)} = \mathbf{0}$$

está dado por el cociente de varianzas generalizadas

$$\Lambda = \frac{\max\limits_{\beta_{(1)},\Sigma} L(\beta_{\mathbf{(1)}}, \mathbf{\Sigma})}{\max\limits_{\beta,\mathbf{\Sigma}} L(\beta, \Sigma)} = \frac{L(\hat{\beta}_{\mathbf{(1)}}, \hat{\mathbf{\Sigma}}_{\mathbf{1}})}{L(\hat{\beta}, \hat{\mathbf{\Sigma}})} = \left(\frac{\left|\hat{\mathbf{\Sigma}}\right|}{\left|\hat{\mathbf{\Sigma}}_{\mathbf{1}}\right|} \right)^{n/2}$$

la cual es frecuentemente convertida a la lambda de Wilks

$$\Lambda^{2/n} = \frac{\left|\hat{\mathbf{\Sigma}}\right|}{\left|\hat{\mathbf{\Sigma}}_{\mathbf{1}}\right|}$$

Resultado: Para el modelo de regresión multivariada, sea $\boldsymbol{Z}$ de rango completo $r+1$, $n \geq r+1+m$ y ε normalmente distribuidos. Entonces bajo la hipótesis nula $H_0 : \beta_{(2)} = \mathbf{0}$, $n(\hat{\boldsymbol{\Sigma}}) \sim W_{n-r-1}(\boldsymbol{\Sigma})$ independientemente de $n(\hat{\boldsymbol{\Sigma}}_{\mathbf{1}} - \hat{\boldsymbol{\Sigma}})$, la cual a su vez de distribuye como $W_{r-q}(\boldsymbol{\Sigma})$.
La prueba de la razón de verosimilitud de H_0 es equivalente a rechazar H_0 para valores grandes de

$$-2\ln\Lambda = -n\ln\left(\frac{\left|\hat{\boldsymbol{\Sigma}}\right|}{\left|\hat{\boldsymbol{\Sigma}}_{\mathbf{1}}\right|}\right) = -n\ln\frac{\left|n\hat{\boldsymbol{\Sigma}}\right|}{\left|n\hat{\boldsymbol{\Sigma}} + n(\hat{\boldsymbol{\Sigma}}_1 - \hat{\boldsymbol{\Sigma}})\right|}$$

Cuando $n-r$ y $n-m$ son ambos grandes, el estadístico modificado

$$-\left[n-r-1-\frac{1}{2}(m-r+q+1)\right]\ln\left(\frac{\left|\hat{\boldsymbol{\Sigma}}\right|}{\left|\hat{\boldsymbol{\Sigma}}_{\mathbf{1}}\right|}\right) \sim \chi^2_{m(r-q)}$$

Regresando de nuevo a la matriz de la suma de cuadrados de los residuales y productos cruzados que ahora denotamos por

$$\boldsymbol{E} = n\hat{\boldsymbol{\Sigma}}$$

y denotamos la matriz de la suma extra de cuadrados y productos cruzados, que representa la matriz bajo la hipotesis nula, como

$$\boldsymbol{H} = n(\widehat{\boldsymbol{\Sigma}}_{\mathbf{1}} - \widehat{\boldsymbol{\Sigma}})$$

entonces podemos definir la lambda de Wilks como

$$\Lambda^{2/n} = \frac{\left|\hat{\boldsymbol{\Sigma}}\right|}{\left|\hat{\boldsymbol{\Sigma}}_{\mathbf{1}}\right|} = \frac{|\boldsymbol{E}|}{|\boldsymbol{E} + \boldsymbol{H}|} = \prod_{i=1}^{s} \frac{1}{1 + \eta_i}$$

donde $\eta_1 \geq \eta_2 \geq \ldots \geq \eta_s$ son los valores propios ordenados de $\boldsymbol{HE}^{-\mathbf{1}}$ donde $s = \text{mín}(m, r - q)$.

Además de la prueba de la razón de verosimilitud, se han propuesto otros estadísticos para probar $H_0 : \beta_{(2)} = \mathbf{0}$:

- Traza de Pillai

$$\sum_{i=1}^{s} \frac{\eta_i}{1+\eta_i} = tr\left[\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{E})^{-\mathbf{1}}\right]$$

- Traza de Hotelling-Lawley

$$\sum_{i=1}^{s} \eta_i = tr\left[\boldsymbol{H}\boldsymbol{E}^{-\mathbf{1}}\right]$$

- Raíz más grande de Roy

$$\frac{\eta_1}{1+\eta_1}$$

Cada uno de estos estadísticos funciona de manera similar a la lambda de Wilks (particularmente para tamaños de muestra grandes).

Ejemplo: Inferencia en regresión multivariada

Para los datos anteriores, las seis observaciones muestrales de dos variables independientes (apetitosidad y textura) y dos variables dependientes (calidad total e intención de compra)

Palatability	Texture	Overall Quality	Purchase Intent
65	71	63	67
72	77	70	70
77	73	72	70
68	78	75	72
81	76	89	88
73	87	76	77

Probar la hipótesis de que:

1. La apetitosidad no tiene ninguna relación conjunta con intención de compras y calidad total
2. y la textura no tiene ninguna relación conjunta con intención de compras y calidad total.

Ejemplo-Inferencia en regresión multivariada

Primero probamos la hipótesis de que la apetitosidad no tiene ninguna relación conjunta con intención de compras y calidad total, i.e.,

$$H_0 : \beta_{(1)} = 0$$

La razón de verosimilitud para la prueba de esta hipótesis está dado por la razón de la varianza generalizada

$$\Lambda = \frac{\max\limits_{\beta_{(2)},\Sigma} L(\beta_{\mathbf{(2)}}, \mathbf{\Sigma})}{\max\limits_{\beta,\mathbf{\Sigma}} L(\beta, \Sigma)} = \frac{L(\hat{\beta}_{\mathbf{(2)}}, \hat{\mathbf{\Sigma}}_{\mathbf{2}})}{L(\hat{\beta}, \hat{\mathbf{\Sigma}})} = \left(\frac{\left|\hat{\mathbf{\Sigma}}\right|}{\left|\hat{\mathbf{\Sigma}}_{\mathbf{2}}\right|}\right)^{n/2}$$

Para facilitar el cálculo, usaremos el estadístico de la lambda de Wilks

$$\Lambda^{2/n} = \frac{\left|\hat{\mathbf{\Sigma}}\right|}{\left|\hat{\mathbf{\Sigma}}_{\mathbf{2}}\right|} = \frac{|\boldsymbol{E}|}{|\boldsymbol{E} + \boldsymbol{H}|}$$

La matriz de la suma de cuadrados del error y productos cruzados es

$$\boldsymbol{E} = \begin{bmatrix} 114.31302415 & 99.335143683 \\ 99.335143683 & 108.5094298 \end{bmatrix}$$

y la matriz de la hipótesis nula es

$$\boldsymbol{H} = \begin{bmatrix} 214.96186763 & 178.26225891 \\ 178.26225891 & 147.82823253 \end{bmatrix}$$

Ejemplo: Inferencia en regresión multivariada

así el valor calculado del estadístico de la lambda de Wilks es

$$
\begin{aligned}
\Lambda^{2/n} &= \frac{|\boldsymbol{E}|}{|\boldsymbol{E}+\boldsymbol{H}|} \\
&= \frac{\left|\begin{bmatrix} 114.31302415 & 99.335143683 \\ 99.335143683 & 108.5094298 \end{bmatrix}\right|}{\left|\begin{bmatrix} 114.31302415 & 99.335143683 \\ 99.335143683 & 108.5094298 \end{bmatrix} + \begin{bmatrix} 214.96186763 & 178.26225891 \\ 178.26225891 & 147.82823253 \end{bmatrix}\right|} \\
&= \frac{2536.570229}{7345.238098} = 0.34533534
\end{aligned}
$$

Ejemplo: Inferencia en regresión multivariada

La transformación a un estadistico distribuido Chi-cuadrado (el cual es actualmente válido únicamente cuando n − r y n − m son grandes) es

$$-\left[n-r-1-\frac{1}{2}(m-r+q+1)\right]\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_1|}\right) = -\left[6-2-1-\frac{1}{2}(2-2+1+1)\right]\ln(0.34533534)$$
$$= 0.92351795$$

para α = 0.01 y m(r - q) = 1 gl , el valor crítico es 9.210351, por tanto existe fuerte evidencia para no rechazar la hipótesis nula. También, el *p*-valor aproximado de esta prueba chi-cuadrado es 0.630174. Nótese que esta es una aproximación extremadamente débil (debido a que $n-r=4$ y $n-m=4$).

Ejemplo: Inferencia en regresión multivariada

Ahora probamos la hipótesis de que la textura no tiene relación conjunta con intención de compra y calidad total, i.e.,

$$H_0 : \beta_{(2)} = 0$$

La razón de verosimilitud para la prueba de esta hipótesis está dado por la razón de la varianza generalizada

$$\Lambda = \frac{\max_{\beta_{(1)},\Sigma} L(\beta_{(1)}, \boldsymbol{\Sigma})}{\max_{\beta,\boldsymbol{\Sigma}} L(\beta, \Sigma)} = \frac{L(\hat{\beta}_{(1)}, \hat{\boldsymbol{\Sigma}}_2)}{L(\hat{\beta}, \hat{\boldsymbol{\Sigma}})} = \left(\frac{\left|\hat{\boldsymbol{\Sigma}}\right|}{\left|\hat{\boldsymbol{\Sigma}}_1\right|} \right)^{n/2}$$

Para facilitar el cálculo, usaremos el estadístico de la lambda de Wilks

$$\Lambda^{2/n} = \frac{\left|\hat{\boldsymbol{\Sigma}}\right|}{\left|\hat{\boldsymbol{\Sigma}}_1\right|} = \frac{|\boldsymbol{E}|}{|\boldsymbol{E} + \boldsymbol{H}|}$$

La matriz de la suma de cuadrados del error y productos cruzados es

$$\boldsymbol{E} = \begin{bmatrix} 114.31302415 & 99.335143683 \\ 99.335143683 & 108.5094298 \end{bmatrix}$$

y la matriz de la hipótesis nula es

$$\boldsymbol{H} = \begin{bmatrix} 21.872015222 & 20.255407498 \\ 20.255407498 & 18.758286731 \end{bmatrix}$$

asi el valor calculado del estadístico de la lambda de Wilks es

$$\Lambda^{2/n} = \frac{|\boldsymbol{E}|}{|\boldsymbol{E}+\boldsymbol{H}|}$$
$$= \frac{\left|\begin{bmatrix} 114.31302415 & 99.335143683 \\ 99.335143683 & 108.5094298 \end{bmatrix}\right|}{\left|\begin{bmatrix} 114.31302415 & 99.335143683 \\ 99.335143683 & 108.5094298 \end{bmatrix} + \begin{bmatrix} 21.872015222 & 20.255407498 \\ 20.255407498 & 18.758286731 \end{bmatrix}\right|}$$
$$= \frac{2536.570229}{3030.059055} = 0.837135598$$

La transformación a un estadístico distribuido Chi-cuadrado (el cual es válido únicamente cuando $n-r$ y $n-m$ son grandes) es

$$-\left[n-r-1-\frac{1}{2}(m-r+q+1)\right]\ln\left(\frac{\left|\hat{\boldsymbol{\Sigma}}\right|}{\left|\hat{\boldsymbol{\Sigma}}_1\right|}\right)$$

$$=-\left[6-2-1-\frac{1}{2}(2-2+1+1)\right]\ln(0.837135598)=0.15440838$$

para $\alpha = 0.01$ y $m(r-q)=1$ gl, el valor crítico es 9.210351 que indica fuerte evidencia de no rechazo. También, el p-valor aproximado de esta prueba Chi-cuadrada es 0.925701 (nótese que esta es una aproximación débil (debido a que $n-r=4$ y $n-m=4$).

- Supongamos que el modelo $\boldsymbol{Y} = \boldsymbol{Z}\beta + \varepsilon$, con los ε distribuidos normalmente, se ajustó a un conjunto de datos. Si el modelo resultó adecuado, se puede usar para fines predictivos.
- Así, un problema es predecir la respuesta media, $E(\boldsymbol{Y}) = \beta' \boldsymbol{z_0}$, correspondiente a un valor fijo de las variable predictoras, $\boldsymbol{z_0}$.
- De los resultados sobre las distribuciones muestrales de los estimadores de maxima verosimilitud tenemos que

$$\hat{\beta}' \boldsymbol{z_0} \sim \boldsymbol{N_m}(\beta' \boldsymbol{z_0}, \boldsymbol{z_0}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z_0}\boldsymbol{\Sigma}) \;\; \text{y} \;\; n\hat{\boldsymbol{\Sigma}} \sim W_{n-r-1}(\boldsymbol{\Sigma})$$

El valor desconocido de la función de regresión en z_0 es $\beta' \boldsymbol{z_0}$, Así, la T^2 se puede escribir como

$$T^2 = \left(\frac{\hat{\beta}' \boldsymbol{z}_0 - \beta' \boldsymbol{z_0}}{\sqrt{\boldsymbol{z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_0}} \right)' \left(\frac{n}{n-r-1} \hat{\boldsymbol{\Sigma}} \right)^{-1} \left(\frac{\hat{\beta}' \boldsymbol{z}_0 - \beta' \boldsymbol{z_0}}{\sqrt{\boldsymbol{z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_0}} \right)$$

- De esta forma el elisoide de confianza del $100(1-\alpha)\%$ para la funcion de regresión $\beta' \boldsymbol{z_0}$ asociado con $\boldsymbol{z_0}$, está dado por la desigualdad

$$\left(\hat{\beta}' \boldsymbol{z}_0 - \beta' \boldsymbol{z_0}\right)' \left(\frac{n}{n-r-1}\hat{\boldsymbol{\Sigma}}\right)^{-1} \left(\hat{\beta}' \boldsymbol{z}_0 - \beta' \boldsymbol{z_0}\right)$$
$$\leq \boldsymbol{z_0}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z_0}\left(\frac{m(n-r-1)}{n-r-m}F_{m,n-r-m}(\alpha)\right)$$

donde $F_{m,n-r-m}(\alpha)$ representan los percentiles superiores del $(1-\alpha)\%$ de una F con m y $n-r-m$ grados de libertad.
- Los intervalos de confianza *simultáneos* del $100(1-\alpha)\%$ para $E(\boldsymbol{Y}_i) = \boldsymbol{z}_0'\beta_{(\boldsymbol{i})}$ asociado con $\boldsymbol{z_0}$, son

$$\boldsymbol{z_0}'\hat{\beta}_{(\boldsymbol{i})} \pm \sqrt{\frac{m(n-r-1)}{n-r-m}F_{m,n-r-m}(\alpha)}\sqrt{z_0'(Z'Z)^{-1}z_0\frac{n}{n-r-1}\hat{\sigma}_{ii}} \quad i=1,.$$

donde $\hat{\beta}_{(i)}$ es la i-esima columna de $\hat{\beta}$ y $\hat{\sigma}_{ii}$ es el i-esimo elemento de la diagonal de $\hat{\boldsymbol{\Sigma}}$

- Finalmente, podemos construir elipsoides e intervalos de confianza para el valor predicho de $\boldsymbol{Y_0}$ asociado con z_0.
- Asumiendo que el error del modelo $\boldsymbol{Y_0} = \beta' \boldsymbol{z_0} + \varepsilon_0$ sigue una distribución normal, entonces

$$\boldsymbol{Y_0} - \widehat{\beta}' \boldsymbol{z_0} = (\beta - \widehat{\beta})' \boldsymbol{z_0} + \varepsilon_0 \sim N_m(\boldsymbol{0}, (1 + \boldsymbol{z_0'}(\boldsymbol{Z'Z})^{-1}\boldsymbol{z_0})\boldsymbol{\Sigma})$$

e independiente de $n\hat{\boldsymbol{\Sigma}} \sim W_{n-r-1}(\boldsymbol{\Sigma})$

- De esta forma, el elipsoide de predicción del $100(1-\alpha)\%$ para $\boldsymbol{Y_0}$ asociado con $\boldsymbol{z_0}$ está dado por

$$\left(\mathbf{Y_0} - \hat{\beta}'\boldsymbol{z_0}\right)' \left(\frac{n}{n-r-1}\hat{\boldsymbol{\Sigma}}\right)^{-1} \left(\mathbf{Y_0} - \hat{\beta}'\boldsymbol{z_0}\right)$$

$$\leq \left(\mathbf{1} + \boldsymbol{z_0}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z_0}\right) \left(\frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}(\alpha)\right)$$

- y los intervalos de predicción simúltaneos del $100(1-\alpha)\%$ para las respuestas individuales Y_{0i} son

$$\mathbf{z_0}'\hat{\beta}_{(\mathbf{i})} \pm \sqrt{\frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}(\alpha)} \sqrt{1+\mathbf{z_0}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z_0} \frac{n}{n-r-1} \hat{\sigma}_{ii}} \quad i=1,\ldots$$