EL MODELO LINEAL

Clase 4

Propiedades estadísticas

Sergio Camiz

LIMA - Marzo-Mayo 2025

Asuntos de la Clase 4

- Propiedades estadísticas de los estimadores
- Análisis de la varianza
- Inferencia
- El programa `lm`

Propiedades estadísticas de los estimadores

Para poder estudiar propriedades estadísticas de los estimadores, se necesitan unas hipótesis, que llevan a los resultados.

- Se consideran p vectores, incluyendo el vector $\boldsymbol{x}_1 = (1, \ldots, 1)'$, de valores observados sobre n unidades fijadas (no al azar), formando la matriz de variables descriptivas $\boldsymbol{X}$.
- por cada unidad i, se observan valores $y_i,\ i = 1, \ldots, n$ de la variable respuesta, que resulta una variable aleatória.
- siempre se supone un error, así que resulta por cada i

$$y_i = \sum_j \beta_j x_{ij} + \varepsilon_i = \eta_i + \varepsilon_i \tag{1}$$

con ε_i el desvío del valor observado al modelo, o sea, $\varepsilon_i = y_i - \eta_i$ que es también una variable aleatoria.

Se necesitan otras hipótesis:

- Las observiaciones y_i son variables aleatorias independientes.
- Por cada observación, o sea por cada vector $\boldsymbol{x}_i$ (filas de $\boldsymbol{X}$), y_i tiene la misma distribución, con
- $E(y_i|\boldsymbol{x}_i) = \eta_i$, o sea el modelo es correcto;
- $V(y_i|\boldsymbol{x}_i) = \sigma^2$ costante, por cada $i = 1, \ldots, n$.
- un error experimental siempre existe, dependiendo de la influencia sobre y de otros q *factores* $\boldsymbol{x}_{j'}$ no incluídos, así el modelo (1) es un abreviado de

$$y_i = \sum_j \beta_j x_{ij} + \sum_{j'} \beta_{j'} x_{ij'}.$$

Si q es grande el teorema del límite central asegura la normalidad de ε_i.

Estas condiciones se pueden sintetizar ecuivalentemente

$$\begin{cases} y_i = \eta_i + \varepsilon_i \\ E(y_i|\boldsymbol{x}_i) = \eta_i \text{ (correcto)} \\ V(y_i|\boldsymbol{x}_i) = \sigma^2 \text{ (homoscedasticidad)} \\ y_i \text{ y } y_k \text{ independientes } \forall i \neq k \\ y_i \sim N(\eta_i, \sigma^2) \text{ (normalidad)} \end{cases} \qquad \begin{cases} y_i = \eta_i + \varepsilon_i \\ E(\varepsilon_i|\boldsymbol{x}_i) = 0 \\ V(\varepsilon_i|\boldsymbol{x}_i) = \sigma^2 \\ \varepsilon_i \text{ y } \varepsilon_k \text{ independientes } \forall i \neq k \\ \varepsilon_i \sim N(0, \sigma^2) \end{cases}$$

por cada $i \in (1, n)$, y en forma matricial

$$\begin{cases} \boldsymbol{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ E(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{\eta} \\ V(\boldsymbol{y}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I} \\ y_i \text{ y } y_k \text{ independientes } \forall i \neq k \\ \boldsymbol{y} \sim N(\boldsymbol{\eta}, \sigma^2 \boldsymbol{I}) \end{cases} \qquad \begin{cases} \boldsymbol{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ E(\boldsymbol{\varepsilon}) = \boldsymbol{0} \\ V(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I} \\ \varepsilon_i \text{ y } \varepsilon_k \text{ independientes } \forall i \neq k \\ \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}) \end{cases}$$

En seguida se utilizarán las relaciones siguientes. Para su prueba, verse por ejemplo Mood *et al.* (1974):

$$E(ax+b) = aE(x) + b \qquad V(ax+b) = a^2 V(x)$$

$$E(x \pm y) = E(x) \pm E(y) \qquad V(x \pm y) = V(x) + V(y) \pm 2cov(x,y)$$

$$E(xy) = E(x)E(y) + cov(x,y) \qquad V(xy) = V(x)V(y) + E(x)E(y)$$

$$E(x^2) = (E(x))^2 + V(x) \qquad V(x^2) = (V(x))^2 + (E(x))^2$$

Esperanza

Sabiendo que $E(\boldsymbol{y}) = \boldsymbol{\eta} = \boldsymbol{X\beta}$, y que la matriz $\boldsymbol{X}$ es fijada, se resulta que

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{y}) = \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\eta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta} \\ E(\hat{\boldsymbol{\eta}}) &= E(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{X}E(\hat{\boldsymbol{\beta}}) = \boldsymbol{X\beta} = \boldsymbol{\eta} \\ E(\boldsymbol{e}) &= E(\hat{\boldsymbol{\varepsilon}}) = E(\boldsymbol{y} - \hat{\boldsymbol{\eta}}) = E(\boldsymbol{y}) - E(\hat{\boldsymbol{\eta}}) = \boldsymbol{\eta} - \boldsymbol{\eta} = \boldsymbol{0} \end{aligned}$$

Varianza

Considerando también que $V(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}$, resulta

$$\begin{aligned} V(\hat{\boldsymbol{\beta}}) &= V((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'V(\boldsymbol{y})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ V(\hat{\boldsymbol{\eta}}) &= V(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{X}V(\hat{\boldsymbol{\beta}})\boldsymbol{X}' = \sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \sigma^2\mathscr{P} \\ V(\boldsymbol{e}) &= V(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = V(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}) = \\ &= \sigma^2(\boldsymbol{I}_n - \mathscr{P}) = \sigma^2\mathscr{E} \end{aligned}$$

Hay que observar que los estimadores de los parámetros tienen una covarianza entre ellos, dependiendo de las relaciones lineales entre regresores.

El teorema de Gauss-Markov

Dados

1. n conjuntos de observaciones, dispuestas en forma de matriz $(\boldsymbol{X}, \boldsymbol{y})$, con
2. $\boldsymbol{X}$ matriz de valores previamente elegidos, y
3. $\boldsymbol{y}$ valores al azar correspondientes y independientes para los cuales $E(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}$, $V(\boldsymbol{y}|\boldsymbol{X}) = \sigma^2 I$;
4. Sea $\hat{\boldsymbol{\beta}}$ la estimación de mínimos cuadrados de $\boldsymbol{\beta}$.
5. $\boldsymbol{\tau} = \boldsymbol{a}'\boldsymbol{\beta}$, con $\boldsymbol{a}' = (a_1, \ldots, a_n)$ un vector de constantes.

Bajo 1.,...,5., entre todos los estimadores *insesgados* y *lineal* en $\boldsymbol{y}$ de $\boldsymbol{\tau}$, el de mínimos cuadrados $\hat{\boldsymbol{\tau}} = \boldsymbol{a}'\hat{\boldsymbol{\beta}}$ es de varianza mínima.

Por lo tanto, cada $\hat{\beta}_j$ y $\hat{\boldsymbol{\eta}}_i$ entre todos los estimadores insesgados y lineales en $\boldsymbol{y}$, son los de varianza mínima.

Prueba.
Es evidente que $E(\hat{\boldsymbol{\tau}}) = \boldsymbol{\tau}$ y que $\hat{\boldsymbol{\tau}}$ es lineal en $\boldsymbol{y}$ pues

$$\hat{\boldsymbol{\tau}} = \boldsymbol{a}'\hat{\boldsymbol{\beta}} = \boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{c}'\boldsymbol{y},$$

entonces su varianza es

$$V(\hat{\boldsymbol{\tau}}) = V(\boldsymbol{a}'\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a}.$$

Supongamos exista otra estimación de $\boldsymbol{\tau}$, insesgada y lineal en $\boldsymbol{y}$, digamos $\boldsymbol{t} = \boldsymbol{d}'\boldsymbol{y}$, con $\boldsymbol{d} \neq \boldsymbol{c}$. Por la condición de insesgamiento, $E(\boldsymbol{t}|\boldsymbol{X}) = \boldsymbol{\tau}$ y por lo tanto

$$E(\boldsymbol{t}) = \boldsymbol{d}'E(\boldsymbol{y}|\boldsymbol{X}) = \boldsymbol{d}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\tau} = \boldsymbol{a}'\boldsymbol{\beta}$$

por cada $\boldsymbol{\beta}$. Por tanto $\boldsymbol{d}'\boldsymbol{X} = \boldsymbol{a}'$.

La covarianza entre $\boldsymbol{t}$ y $\hat{\boldsymbol{\tau}}$ vale

$$\begin{aligned}\operatorname{cov}(\hat{\boldsymbol{\tau}}, \boldsymbol{t}) = \operatorname{cov}(\boldsymbol{c}'\boldsymbol{y}, \boldsymbol{d}'\boldsymbol{y}) = E(\boldsymbol{d}'(\boldsymbol{y}-\boldsymbol{\eta})(\boldsymbol{y}-\boldsymbol{\eta})'\boldsymbol{c}) = \sigma^2\boldsymbol{d}'\boldsymbol{c} = \\ = \sigma^2\boldsymbol{d}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a} = \sigma^2\boldsymbol{a}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{a} = V(\hat{\boldsymbol{\tau}})\end{aligned}$$

Se encuentra entonces que

$$0 \leq V(\boldsymbol{t} - \hat{\boldsymbol{\tau}}) = V(\boldsymbol{t}) + V(\hat{\boldsymbol{\tau}}) - 2\operatorname{cov}(\boldsymbol{t}, \hat{\boldsymbol{\tau}}) = V(\boldsymbol{t}) - V(\hat{\boldsymbol{\tau}})$$

y por tanto $V(\boldsymbol{t}) \geq V(\hat{\boldsymbol{\tau}})$.

Análisis de varianza del modelo

Una vez estimados los $\hat{\boldsymbol{\beta}}$ resulta que el cuadrado de la distancia entre $\boldsymbol{y}$ y $\boldsymbol{X}$ es

$$\begin{aligned}\boldsymbol{e}'\boldsymbol{e} &= (\boldsymbol{y} - \hat{\boldsymbol{\eta}})'(\boldsymbol{y} - \hat{\boldsymbol{\eta}}) = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} + \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \\ &= \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}}\end{aligned}$$

pues, según las ecuaciones normales,

$$\boldsymbol{y}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y})' = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}},$$

de donde $\quad \boldsymbol{y}'\boldsymbol{y} = \hat{\boldsymbol{\eta}}'\hat{\boldsymbol{\eta}} + \boldsymbol{e}'\boldsymbol{e} = \boldsymbol{y}'\mathscr{P}\boldsymbol{y} + \boldsymbol{y}'\mathscr{E}\boldsymbol{y}.$

Esto resulta también del teorema de Pitagoras.

Se puede escribir también

$$SS_t = SS_r + SS_e,$$

ya que se ha compartido la suma de cuadrados de las observaciones:

- una parte SS_r debido a la regresión de $\boldsymbol{y}$ sobre X
- la otra, SS_e, debido al error.

  Por tanto:
- $\hat{\boldsymbol{\eta}}$ contiene la información sobre el modelo $\boldsymbol{\eta} = X\boldsymbol{\beta}$,
- $\boldsymbol{e}$ solo contiene la información sobre el error,
- $\boldsymbol{e}'\boldsymbol{e}$ solo debe informar sobre σ^2.

Calculemos ahora las esperanzas SS_r y SS_e:

$$E(SS_r) = E(\boldsymbol{y}'\mathscr{P}\boldsymbol{y}) = \boldsymbol{\beta}'\boldsymbol{X}'\mathscr{P}\boldsymbol{X}\boldsymbol{\beta} + tr(\mathscr{P}\sigma^2 I) = \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + p\sigma^2$$
$$E(SS_e) = E(\boldsymbol{y}'\mathscr{E}\boldsymbol{y}) = \boldsymbol{\beta}'\boldsymbol{X}'\mathscr{E}\boldsymbol{X}\boldsymbol{\beta} + tr(\mathscr{E}\sigma^2 I) = (n-p)\sigma^2$$

resultando el producto $\mathscr{E}\boldsymbol{X} = 0$. Por lo tanto los cuadrados promedios se resultan

$$MS_r = SS_r/p \qquad E(MS_r) = \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}/p + \sigma^2$$
$$MS_e = SS_e/(n-p) \qquad E(MS_e) = \sigma^2$$

así que MS_e es un estimador insesgado de σ^2.

Se suele enrequecer la tabla de análisis de varianza con las esperanzas como sigue:

Fuente	Grados de libertad (DF)	Sumas de cuadrados (SS)	Cuadrados promedios (MS)	Esperanza de los cuadrados promedios E(MS)
Regresión	p	SS_r	$MS_r = SS_r/p$	$\sigma^2 + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}/p = \sigma^2 + \boldsymbol{\eta}'\boldsymbol{\eta}/p$
Error	$n-p$	SS_e	$MS_e = SS_e/(n-p)$	σ^2
Total	n	SS_t		

Los grados de libertad corresponden a las dimensiones de los espacios:

- el espacio de los estimadores X tiene p dimensiones,
- el espacio de los residuos tiene dimensión $n-p$.

Nótese que las esperanzas de los cuadrados promedios son diferentes solo si $\parallel \boldsymbol{\eta} \parallel \neq 0$.

Inferencia

Para hacer inferencia a una poblacíon, y también evaluar si el modelo resulta fiable y es necesario imponer hipótesis sobre la distribución de $\boldsymbol{y}$ condicionada a los $\boldsymbol{X}$.

Si se asume la distribución de los $y_i|\boldsymbol{x}_i$ multinormal, o sea normal por cada valor de los x_{ij}, en seguida se consiguen le distribuciones de los resultados.

Sino hay que conocer resultados parecidos por la distribución que resulta o se hace recurso a métodos de remuestreo (Manly, 2007), construyendo distribuciones empíricas. ´

Se presentan aquí los resultados más importantes relativos a las muestras de una distribución normal multivariada, que servirán a los test estadísticos. Para una lectura más detallada, demostraciones y referencias se puede ver Guttman (1982: pp. 62–95).

Distribuciones estadísticas

Definición. Se dice que el vector aleatorio $\boldsymbol{y} \in \mathbb{R}^n$, tiene una distribución *multi-normal* si su función de densidad $d(\boldsymbol{y})$ es

$$d(\boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^n\,|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}$$

donde la matriz de varianza/covarianza $\boldsymbol{\Sigma}$ es simétrica definida positiva, con $|\boldsymbol{\Sigma}| = \Pi_i \lambda_i$ producto de su autovalores, y el vector de promedios $\boldsymbol{\mu}$ tiene componentes finitas. Entonces se escribe $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, con $E(\boldsymbol{y}) = \boldsymbol{\mu}$ y $V(\boldsymbol{y}) = \boldsymbol{\Sigma}$.

Teorema. Si $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $\boldsymbol{\Sigma}$ es diagonal, las componentes y_i de $\boldsymbol{y}$ son estadísticamente independientes.

Donde la necesidad de normalidad y independencia de los y_i.

Si asumimos normal independiente la distribución de los resíduos $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ se resulta la distribución

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

donde la función de verosimilitud

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{(2\pi)^n \sigma^2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}.$$

Entonces el estimador de máxima verosimilitud de $\boldsymbol{\beta}$ tiene que maximizar la log-verosimilitud

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

y resulta ser $\hat{\boldsymbol{\beta}}$, que por lo tanto es de mínima varianza.

Como consequencia de la multinormalidad, se determinan las distribuciones de formas cuadraticas, o sea las sumas de cuadrados, que resultan en general:

Teorema. Sea un vector aleatorio $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$, con $\boldsymbol{\Sigma} = \boldsymbol{P}'\boldsymbol{P}$ y Q la forma cuadrática centrada

$$Q = (\boldsymbol{y} - \boldsymbol{\mu})'\boldsymbol{G}(\boldsymbol{y} - \boldsymbol{\mu})$$

con $\boldsymbol{G}$ simétrica y real. Entonces la ley de distribución de Q es una combinación lineal de n variables aleatorias independientes de ley chi-cuadrado con 1 grado de libertad

$$Q \sim \sum_i \lambda_i \chi^2_{1,i}$$

donde los λ_i son los autovalores de $\boldsymbol{P}'\boldsymbol{G}\boldsymbol{P}$ (y de $\boldsymbol{\Sigma}\boldsymbol{G}$ y $\boldsymbol{G}\boldsymbol{\Sigma}$).

Por consecuencia, si $\lambda_i = 1$ *o* 0 se deduce:

Teorema. Una condición necesaria y suficiente por que Q tenga una ley de distribución de chi-cuadrado con $p < n$ grados de libertad es que $\boldsymbol{P}'\boldsymbol{G}\boldsymbol{P}$ sea idempotente y de rango p. Si $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$ la condición deviene en que $\boldsymbol{G}$ sea idempotente de rango p.

Teorema (Craig). Sea un vector aleatorio $\boldsymbol{y} = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y las dos formas cuadráticas

$$Q_i = (\boldsymbol{y} - \boldsymbol{\mu})'\boldsymbol{G}_i(\boldsymbol{y} - \boldsymbol{\mu}), i = 1, 2$$

con $\boldsymbol{G}_i$ reales y simétricas. Entonces Q_1 y Q_2 son estadísticamente independientes si y solo si $\boldsymbol{G}_1\boldsymbol{\Sigma}\boldsymbol{G}_2 = \boldsymbol{0}$.

Esto es relevante para los proyectores ortogonales entre si.

Teorema. (Cochran). Sea $y \sim N(0,1)$ una variable aleatoria y sean n observaciones de y independientes, formando un vector aleatorio $\boldsymbol{y} \sim N(\boldsymbol{0}, \boldsymbol{I})$. Sea por otro lado

$$Q = \boldsymbol{y'y} = Q_1 + Q_2 + ... + Q_k$$

donde $Q_i = \boldsymbol{y'A}_i\boldsymbol{y}$ es una forma cuadrática de rango $rg(\boldsymbol{A}_i) = n_i$, y $\boldsymbol{A}_i$ es una matriz simétrica $n \times n, i = 1, ..., k$. Entonces las siguiente condiciones son equivalentes:

- $Q_1, Q_2, ..., Q_k$ son estadísticamente independientes;
- $Q_1, Q_2, ..., Q_k$ tienen individualmente distribuciones de chi-cuadrado, con $n_1, \ldots, n_k$ grados de libertad;
- $n_1 + n_2 + \ldots + n_k = n$.

Esto nos sirve para la partición de SS_y.

Teorema. Sea el vector aleatorio $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y considere su distribución condicional

$$\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{\mu}) = 0$$

donde $\boldsymbol{A}$ es una matriz $p \times n, p < n, \mathrm{rk}(\boldsymbol{A}) = p$. Entonces la distribución condicional de $\boldsymbol{y}$ es tal que

$$Q = (\boldsymbol{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = \chi^2_{n-p}$$

En fin, esto no sirve para SS_e.

Test del modelo

Se supone que el vector $\boldsymbol{y}$ tiene

- en cada punto de observación una distribución normal centrada sobre su esperanza $\boldsymbol{\eta}$
- y de varianza constante, $\boldsymbol{y} \sim N(\boldsymbol{\eta}, \sigma^2\boldsymbol{I})$.
- Bajo la hipótesis nula $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ se tiene
- $E(\boldsymbol{y}) = \boldsymbol{0}$
- y por tanto $\boldsymbol{y} = N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.
- Por el teorema de Craig, SS_r y SS_e son formas cuadráticas de una distribución normal estandarizada estadísticamente independientes, de matrizes $\mathscr{P}$ y $\mathscr{E}$ tales que $\sigma^2\mathscr{P}\mathscr{E} = 0$.

En consecuencia (teorema de Cochran, con $k = 2$) se resulta

$$SS_r = \sigma^2\chi^2_p \;\text{ y }\; SS_e = \sigma^2\chi^2_{n-p}.$$

Se puede estimar σ^2 también con el máximo de verosimilitud. Derivando $l(\boldsymbol{\beta}, \sigma^2)$ por respecto a σ^2 y igualando a cero

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) = 0,$$

se consigue, sostituyendo $\hat{\boldsymbol{\beta}}$ *a* $\boldsymbol{\beta}$,

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\boldsymbol{e'e} = 0,$$

donde

$$\frac{n}{2\sigma^2} = \frac{1}{2\sigma^4}\boldsymbol{e'e} \;\text{ y }\; \sigma^2_{ml} = \frac{\boldsymbol{e'e}}{n},$$

de varianza minima, pero sesgado, ya que $E(MS_e) = E(SS_e/(n-p)) = \sigma^2$. Se corrige con $\sigma^2 = n/(n-p)\sigma^2_{ml}$.

Como resulta

$$SS_r = \sigma^2 \chi^2_p \quad \text{y} \quad SS_e = \sigma^2 \chi^2_{n-p}$$

con las leyes χ^2 independientes, donde

$$\text{bajo } H_0 : \boldsymbol{\beta} = \mathbf{0}, \quad F = \frac{MS_r}{MS_e} = F_{p,n-p}$$

una F de Fisher con p y $n-p$ grados de libertad. Por tanto, fijado un nivel de probabilidad π el test resulta ser

$$\begin{cases} \text{no aceptar } H_0 : \boldsymbol{\beta} = \mathbf{0}, \text{ si } \dfrac{MS_r}{MS_e} > F_{p,n-p;\pi} \\ \text{aceptar en otro caso.} \end{cases}$$

Por otro lado, es fácil de ver que en la descomposición

$$(\boldsymbol{y}-X\boldsymbol{\beta})'(\boldsymbol{y}-X\boldsymbol{\beta}) = (\boldsymbol{y}-X\hat{\boldsymbol{\beta}})'(\boldsymbol{y}-X\hat{\boldsymbol{\beta}})+(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'X'X(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$$

el primero término a la derecha vale $\boldsymbol{e'e}$ y se puede escribir

$$\frac{1}{\sigma^2}(\boldsymbol{y} - X\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}) = \frac{1}{\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})'\mathscr{E}(\boldsymbol{y} - X\boldsymbol{\beta}),$$
$$\frac{1}{\sigma^2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'X'X(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \frac{1}{\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})'\mathscr{P}(\boldsymbol{y} - X\boldsymbol{\beta}),$$

dos formas cuadráticas independientes. Entonces resulta

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X'X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/p}{\boldsymbol{e'e}/(n-p)} = \frac{\sigma^2 \chi^2{}_p/p}{\sigma^2 \chi^2{}_{n-p}/(n-p)} = F_{p,n-p}$$

donde se construye la región de confianza de $\boldsymbol{\beta}$ al nivel de $1-\pi$.

$$C_{1-\pi} = \left\{\boldsymbol{\beta} \,\middle|\, (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\boldsymbol{X'X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq pMS_e F_{p,n-p;\pi})\right\}$$

Estas estadísticas permiten de enrequecer la tabla de análisis de la varianza con nuevas columnas.

Fuente	*Grados de libertad (DF)*	*Sumas de cuadrados (SS)*	*Cuadrados promedios (MS)*	*Esperanza de los cuadrados promedios E(MS)*	*F*	*Prob*
Regresión	p	SS_r	$MS_r = SS_r/p$	$\sigma^2 + \boldsymbol{\eta'\eta}/p$	$\frac{MS_r}{MS_e}$	π
Error	$n-p$	SS_e	$MS_e = SS_e/(n-p)$	σ^2		
Total	n	SS_t				

De esta manera, toda la información para evaluar el modelo se encuentra en esta tabla. El valor π es la probabilidad asociata al valor de la F calculada, con p y $n-p$ grados de libertad.

Como la hipótesis nula es $\boldsymbol{\beta} = \mathbf{0}$, si entonces π es más grande del nivel de aceptación que hemos establecido (por ejemplo $5\,\% = 0{,}05$) podemos aceptar que no hay regresión, sino se puede rechazar, en el sentito que hay a lo menos un coeficiente β diferente de cero, y portanto el modelo tiene un sentido.

Test de los $\hat{\beta}_i$

Si el test *ANOVA* es positivo, o sea el modelo tiene sentido, entonces se pueden testar los $\hat{\beta}_i$. Como $\hat{\boldsymbol{\beta}}$ es distribuido multinormalmente con promedio $\boldsymbol{\beta}$ y varianza $\sigma^2(\boldsymbol{X'X})^{-1}$, cada $\hat{\beta}_i$ tiene una distribución normal $\hat{\beta}_i = N(\beta_i, \sigma^2 c_{ii})$ con c_{ii} elemento diagonal de $(\boldsymbol{X'X})^{-1}$. Como su *desvío estándard* es $SD(\hat{\beta}_i) = \sqrt{MS_e c_{ii}}$, el test resulta

$$\begin{cases} \text{no aceptar } H_0 : \hat{\beta}_i = \beta_i, \text{ si} \left|\dfrac{\hat{\beta}_i - \beta_i}{SD(\hat{\beta})}\right| > t_{n-k;\pi/2} \\ \text{aceptar en otro caso.} \end{cases}$$

con intervalo de confianza (con cuidado)

$$C_{1-\pi} = \left\{\beta_i | \hat{\beta}_i - SD(\hat{\beta}_i)\, t_{n-k;\pi/2} \leq \beta_i \leq \hat{\beta}_i + SD(\hat{\beta}_i)\, t_{n-k;\pi/2}\right\}$$

Test de la varianza SS_e

La varianza del modelo σ^2 es estimada para SS_e. Por lo tanto, siendo un chi-cuadrado, su intervalo de confianza es dado por

$$C_{1-\pi} = \left\{ \sigma^2 \middle| \frac{SS_e}{{\chi^2}_{n-k;\pi/2}} \leq \sigma^2 \leq \frac{SS_e}{{\chi^2}_{n-k;1-\pi/2}} \right\}$$

Test de los $\hat{\eta}_i$

Resulta

$$V(\hat{\boldsymbol{\eta}}) = V(\boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{X}V(\hat{\boldsymbol{\beta}})\boldsymbol{X'} = \sigma^2\boldsymbol{X}(\boldsymbol{X'}\boldsymbol{X})^{-1}\boldsymbol{X'},$$

entonces por la estimación de $\hat{\eta}_i = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$, resulta su desvío estándar

$$SD(\hat{\eta}_i) = \sqrt{MS_e\, \boldsymbol{x}_i'(\boldsymbol{X'}\boldsymbol{X})^{-1}\boldsymbol{x}_i}$$

se deriva la condición del test

$$\left| \frac{\hat{\eta}_i - \eta_i}{SD(\hat{\eta}_i)} \right| > t_{n-p;\pi/2}$$

con intervalo de confianza

$$\left\{ \eta \middle| \hat{\eta}_i - SD(\hat{\eta}_1)\, t_{n-p;\pi/2} \leq \eta \leq \hat{\eta}_i + SD(\hat{\eta}_i)\, t_{n-p;\pi/2} \right\}$$

Test del predictor

De manera análoga se calcula el desvío estándar del predictor $\tilde{y}_0$, correspondiente a un nuevo vector $\boldsymbol{x}_0$ que se encuentre dentro de la región ocupada para los $\boldsymbol{x}_{ik}$. Esto vale

$$SD(\tilde{y}_0) = \sqrt{MS_e\,(1 + \boldsymbol{x}_0'(\boldsymbol{X'}\boldsymbol{X})^{-1}\boldsymbol{x}_0)},$$

y por lo tanto su intervalo de confianza resulta ser

$$\left\{ y \middle| \tilde{y}_0 - SD(\tilde{y}_0)\, t_{n-p;\pi/2} \leq y \leq \tilde{y}_0 + SD(\tilde{y}_0)\, t_{n-p;\pi/2} \right\}$$

Nótese que este intervalo es más grande de lo de los η_i, ya que estos son estimaciones del promedio de los y_i ccorrespondientes, mientras aquí es una observación. Entonces la varianza del predictor es la varianza del promedio más la varianza σ^2 alrededor de esto.

El programa `lm`

El programa `lm` es el método básico para correr el modelo lineal en R. El comando es:

```
LM = lm(MODELO, data=DATOS)
```

con:

- `LM` la lista donde guardar los resultados,
- `DATOS` el archivo donde se encuentran tanto $\boldsymbol{X}$ como $\boldsymbol{y}$,
- `MODELO` es la descripción del modelo indicada con las reglas siguientes.

Los resultados principales se resultan con el comando:

```
summary(LM)
```

El modelo

- Un modelo se escribe usando una fórmula del tipo *variable de respuesta* $\sim$ *variables predictoras*.
- La tilde se lee $\sim$ como "se modela como una función de".
- Básicamente se emplea el símbolo $\sim$ entre un objeto, p.e. Y, a la izquierda y otros a la derecha, p.e. X, así indicando que se quiere modelar $\boldsymbol{Y}$ con $\boldsymbol{X}$, es decir $Y \sim X$ significa estimar el modelo $y_i = \alpha + \beta x_i + \varepsilon_i$.
- Si se quieren ajuntar otras variables se incluyen interponiendo el símbolo +: así $Y \sim X + Z + T$ significa estimar el modelo $y_i = \alpha + \beta x_i + \gamma z_i + \delta t_i + \varepsilon_i$.

- Si se quieren ajuntar todas las variables contenidas en la matriz indicada en el parámetro `data=`, se escribe brevemente: $Y \sim .$, el punto indicando todas.
- El símbolo $-$ se usa para tirar una variable del conjunto. Así $Y \sim . - T$ indica el modelo con todas las variables en `data` menos la T.
- Así, escribiendo $Y \sim X - 1$ se indica de estimar el modelo sin intercepta, es decir $y_i = \beta x_i + \varepsilon_i$, sin α.
- Se emplea el símbolo : para indicar interacciones entre variables: así con $Y \sim X + Z + X : Z$ significa estimar el modelo $y_i = \alpha + \beta x_i + \gamma z_i + \delta x_i z_i + \varepsilon_i$

- Si se quiere tener variables y interacciones, se puede escribir también $Y \sim X * Z$, igual a $Y \sim X + Z + X : Z$.
- Con tres variables $Y \sim X * Z * T$ se indican todas las interacciones posibles, es decir:
$Y \sim X + Z + T + X : Z + X : T + Z : T + X : Z : T$.
- El símbolo ˆ sirve para indicar las interacciones que se requieren: Si solo se requieren las interacciones hasta el segundo orden. Así $Y \sim (X + Z + T)\hat{}2$ indica el modelo $Y \sim X + Z + T + X : Z + X : T + Z : T$ sin la interacción triple.
- El grupo $\%in\%$ sirve para indicar que los términos a su izquierda están juntados con los a la derecha. Así $Y \sim X + Z + T\,\%in\%T$ se transforma en $Y \sim X : T + Z : T + T$.

- Cualquier fórmula incluyendo nombres de funciones es aceptada: así son aceptados $Y \sim \log(Z) + \cos(T)$.
- Si se puede tener confusión, la expresión matemática se incluye entre $I()$. Así es diferente $Y \sim X+Z$ de $Y \sim I(X+Z)$ ya que se modelan respectivamente $y_i = \alpha + \beta x_i + \gamma z_i + \varepsilon_i$ y $y_i = \alpha + \beta(x_i + z_i) + \varepsilon_i$.

Los siguientes modelos son iguales, aún con fórmulas diferentes:

```
Y ~ X + Z + W + X: Z + X: W + Z: W
Y ~ X * Z * W - X: Z: W
Y ~ (X + Z + W)^2
```

todas corresponde al mismo modelo:
$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 W_i + \beta_4 X_i Z_i + \beta_5 X_i W_i + \beta_6 Z_i W_i + \varepsilon_i$.