

Clase 2: Modelos Lineales

Justo Andrés Manrique Urbina

28 de agosto de 2019

1. Residuales

Los residuales se pueden escribir como $e = (I - H)Y$. Los residuales son una combinación lineal de las observaciones. A continuación, ver las propiedades de los residuales:

1.1. Esperanza del residual

$$\begin{aligned} E(e) &= E((I - H)Y). \\ &= (I - H)E(Y). \\ &= (I - H)XB. \\ &= 0. \end{aligned}$$

Tarea: Probar por qué sale 0.

1.2. Covarianza del residual

$$\begin{aligned} cov(e) &= cov((I - H)Y). \\ &= (I - H)cov(Y)(I - H). \\ &= (I - H)\sigma^2 I(I - H). \\ &= \sigma^2(I - H)(I - H). \\ &= \sigma^2(I - H). \end{aligned}$$

1.3. Varianza de un residual

Definamos la matriz H como:

$$H = \begin{pmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{pmatrix}$$

Por lo tanto, la varianza de un residual es definida por $\text{var}(e_i) = \sigma^2(1 - h_{ii})$. La covarianza de dos residuales es $\text{cov}(e_i, e_j) = -\sigma^2 h_{ij}$.

En paralelo, los errores tienen la siguiente distribución:

$$\varepsilon \sim N(0, \sigma^2 I).$$

$$\text{var}(\varepsilon) = \sigma^2.$$

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j.$$

1.4. Teorema B 8.2

Sea lo siguiente:

$$X \sim N(0, I).$$

Sea una matriz B $n \times p$ y R una matriz idempotente simétrica. El rango de R se define como $r(R) = r$. Por lo tanto se puede probar lo siguiente:

- $X^T R X \sim X_r^2$
- $\hat{X} B R = 0 \rightarrow X^T R X$ y $B X$ son independientes.

1.5. Análisis de los residuales

Se tiene la suma de cuadrados de los residuales definida como $\sum_{i=1}^n e_i^2 = e^T e$. Por lo tanto se define:

$$\begin{aligned} e^T e &= [(I - H)Y]^T (I - H)Y. \\ &= Y^T (I - H)Y. \\ &= (XB + \varepsilon)^T (I - H)(XB + \varepsilon). \\ &= \varepsilon^T (I - H)\varepsilon + B^T X^T (I - H)XB + \varepsilon^T (I - H)XB + B^T X^T (I - H)\varepsilon. \end{aligned}$$

Tarea: Probar que la expresión $B^T X^T (I - H)\varepsilon$ es 0.

Por otro lado, tenemos que:

$$\begin{aligned} \frac{e^T e}{\sigma^2} &= \frac{\varepsilon^T}{\sigma} (I - H) \frac{\varepsilon}{\sigma}. \\ &= \varepsilon^{*T} (I - H) \varepsilon^*. \end{aligned}$$

en dónde se define como $\varepsilon^* = \frac{\varepsilon}{\sigma}$. Se tiene, también que $\varepsilon^* \sim N(0, I)$. Por el teorema anteriormente mencionado, se tiene que:

$$\varepsilon^{*T} (I - H) \varepsilon^* \sim X_{\text{rang}(1-H)}^2.$$

$$\varepsilon^{*T} (I - H) \varepsilon^* \sim X_{n-p}^2.$$

$$\frac{e^T e}{\sigma^2} \sim X_{n-p}^2.$$

Hallamos el esperado de dicha distribución, la cual es:

$$E\left(\frac{e^T e}{\sigma^2}\right) = n - p.$$

$$E\left(\frac{e^T e}{n - p}\right) = \sigma^2.$$

Por lo tanto, un estimador insesgado de σ^2 es dado por:

$$\hat{\sigma}^2 = \frac{e^T e}{n - p} = \frac{SS_{res}}{n - p} = MS_{res}.$$

$$\frac{(n - p)\hat{\sigma}^2}{\sigma^2} \sim X_{n-p}^2.$$

1.6. Otra subsección

$$\begin{aligned} \frac{1}{\sigma}(\hat{B} - B) &= \frac{1}{\sigma}((X^T X)^{-1} X^T Y - B). \\ &= \frac{1}{\sigma}((X^T X)^{-1} X^T (XB + \varepsilon) - B). \\ &= \frac{1}{\sigma}((X^T X)^{-1} X^T XB + (X^T X)^{-1} X^T \varepsilon - B). \\ &= (X^T X)^{-1} X^T \frac{\varepsilon}{\sigma}. \\ &= (X^T X)^{-1} X^T \varepsilon^*. \\ &= (X^T X)^{-1} X^T - (X^T X)^{-1} X^T H. \\ &= (X^T X)^{-1} X^T - (X^T X)^{-1} X^T X (X^T X)^{-1} X^T. \\ &= 0. \end{aligned}$$

una matriz de ceros. Bajo el teorema anteriormente mencionado, se tiene entonces que:

$(X^T X)^{-1} X^T \varepsilon^*$ es independiente de $\varepsilon^{*T} (I - H) \varepsilon^*$. $\frac{1}{\sigma}(\hat{B} - B)$ es independiente de $\frac{e^T e}{\sigma^2}$, \hat{B} y $\hat{\sigma}^2$ son independientes.

1.7. Otra subsección

Recordemos que:

- $\hat{\sigma}^2 = \frac{e^T e}{n-p}$ es un estimador insesgado de σ^2
- $\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim X_{n-p}^2$
- $\hat{B}y\hat{\sigma}^2$ son independientes.

1.8. Intervalos de confianza

Se tiene que:

$$\hat{B} \sim N(B, \sigma^2(X^T X)^{-1}).$$

y definimos $C = (X^T X)^{-1}(C_{ij})_{p \times p}$. Se tiene entonces que:

$$\hat{B}_j \sim (B_j, \sigma^2 c_{jj}).$$

$$\frac{\hat{B}_j - B_j}{\sigma \sqrt{C_{jj}}} \sim N(0, 1).$$

Si $Z \sim N(0, 1)$ y $W \sim X_k^2$ son independientes, entonces $T = \frac{Z}{\sqrt{\frac{W}{k}}} \sim X_k^2$.

$$\frac{\frac{\hat{B}_j - B_j}{\sigma \sqrt{C_{jj}}}}{\sqrt{(n-p) \frac{\hat{\sigma}^2}{\sigma^2} \frac{1}{n-p}}}.$$

$$\frac{\hat{B}_j - B_j}{\hat{\sigma} \sqrt{C_{jj}}} \sim t_{(n-p)}.$$

$$P(-t_{1-\frac{\alpha}{2}} \leq T \leq t_{1-\frac{\alpha}{2}}) = 1 - \alpha.$$

$$P(\hat{B}_j - t_{1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{C_{jj}} \leq B_j \leq \hat{B}_j + t_{1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{C_{jj}}) = 1 - \alpha.$$

Por tanto, se tiene que:

$$IC_{(1-\alpha)}(B_j) = (\hat{B}_j \pm t_{1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{C_{jj}}).$$

1.9. Error estándar

$$\sqrt{\text{var}(\hat{B})} = \sqrt{\hat{\sigma}^2 C_{jj}}.$$

Se define entonces el error estándar como $\hat{\sigma} \sqrt{C_{jj}}$.

1.10. Predicción

Definamos lo siguiente:

$$Y_0 = B_0 + B_1 X_{01} + \dots + B_k X_{0k} + \varepsilon_0.$$

$$Y_0 = X_0^T B + \varepsilon_0.$$

El esperado de Y_0 , el cual está definido como $E(Y_0|X_0) = X_0^T B$. Finalmente, definimos el estimado de Y_0 como lo siguiente:

$$\hat{Y}_0 = X_0^T \hat{B}.$$

Y hallamos su esperado.

$$E(\hat{Y}_0) = E(X_0^T \hat{B}) = X_0^T B.$$

Tarea: Desarrollar la varianza de $var(\hat{Y}_0)$, la cual es $\sigma^2 X_0^T (X^T X)^{-1} X_0$.

Si estuviéramos en un modelo de regresión lineal simple, se obtiene lo siguiente:

$$var(\hat{Y}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}} \right).$$

en dónde $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$

Tarea: Probar el siguiente resultado, explicando cada paso.

$$IC_{(1-\alpha)}(\mu_{Y|X_0}) = X_0^T \hat{B} \pm t_{1-\frac{\alpha}{2}} \hat{\sigma} \sqrt{X_0^T (X^T X)^{-1} X_0}.$$

Teniendo que:

$$\frac{X_0^T \hat{B} - \mu_{Y|X_0}}{\sigma \sqrt{X_0^T (X^T X)^{-1} X_0}} \sim N(0, 1).$$

1.11. Intervalo de predicción

Se tiene lo siguiente:

- **Nueva observación:** $Y_0 \sim N(X_0^T B, \sigma^2)$
- **Muestra:** $\hat{Y}_0 \sim N(X_0^T B, \sigma^2 X_0^T (X^T X)^{-1} X_0)$

Se tiene que:

$$Y_0 - \hat{Y}_0 \sim N(0, \sigma^2(1 + X_0(X^T X)^{-1} X_0)).$$

1.12. Prueba de Hipótesis

Se tiene una hipótesis lineal general, la cual es:

$$H_0 : CB = d.$$

$$H_1 : CB \neq d.$$

dónde $C_{r \times p}$, $d_{r \times 1}$ y $rango(C) = r$

Tenemos lo siguiente:

- **Modelo Completo:** $Y = XB + \varepsilon$, sin restricción.
- **Modelo Reducido:** $Y = XB + \varepsilon$, sujeto a $H_0 : CB = d$ verdadero.

Tenemos la siguiente estadística de prueba: $\Delta SS_{res} = SS_{res.H_0} - SS_{res}$.

$$\frac{\frac{1}{r} \Delta SS_{res}}{\frac{1}{n-p} SS_{res}} \sim F(r, n-p).$$

Y la prueba es H_0 si $F > F_{1-k}$.

Se define \hat{B}^0 como el estimador de B si $H_0 : CB = d$ es verdadero. Por lo tanto, se debe minimizar $(Y - XB)^T(Y - XB)$ sujeto a $CB = d$. Esto es equivalente a minimizar:

$$L = (Y - XB)^T(Y - XB) - 2\lambda^T(CB - d).$$

$$L = Y^T Y + B^T X^T X B - 2B^T X^T Y - 2\lambda^T(CB - d).$$

Se sacan las derivadas:

$$\frac{\partial L}{\partial B} = -2X^T Y + 2X^T X B - 2C^T \lambda = 0.$$

$$\frac{\partial L}{\partial \lambda} = -2CB + 2d = 0.$$

De la segunda derivada se obtiene que $CB = d$.

En (1), se tiene que:

$$(X^T X)B - X^T Y = C^T \lambda.$$

$$(X^T X)^{-1}(X^T X)B - (X^T X)^{-1}X^T Y = (X^T X)^{-1}C^T \lambda.$$

$$B - \hat{B} = (X^T X)^{-1}C^T \lambda.$$

$$CB - C\hat{B} = C(X^T X)^{-1}C^T \lambda.$$

$$d - C\hat{B} = C(X^T X)^{-1}C^T \lambda.$$

$$(C(X^T X)^{-1}C^T)^{-1}(d - C\hat{B}) = \lambda.$$

Posteriormente se obtiene B :

$$B - \hat{B} = (X^T X)^{-1}C^T \lambda.$$

$$B = \hat{B} - (X^T X)^{-1}C^T (C(X^T X)^{-1}C^T)^{-1}(C\hat{B} - d).$$

$$\hat{B}^0 = \hat{B} - \Delta.$$

$$\hat{Y}^0 = X\hat{B}^0.$$

$$e^0 = Y - \hat{Y}^0.$$

$$= Y - X\hat{B}^0.$$

$$\begin{aligned}
&= Y - X(\hat{B} - \Delta). \\
&= Y - X\hat{B} + X\Delta. \\
&= e + X\Delta.
\end{aligned}$$

Entonces, por otro lado se obtiene:

$$\begin{aligned}
&= e^{0^T} e^0 = (e + X\Delta)^T (e + X\Delta). \\
&= e^T e + \Delta^T X^T X \Delta + \Delta^T X e + e^T X \Delta.
\end{aligned}$$

Tarea: Probar por qué $\Delta^T X e + e^T X \Delta = 0$.

Se tiene entonces que:

$$\begin{aligned}
\Delta SS_{res} &= \Delta^T X^T X \Delta. \\
&= (C\hat{B} - d)^T (C(X^T X)^{-1} C(X^T X)^{-1} X^T X (X^T X)^{-1} C^T (C(X^T X)^{-1} C^T)^{-1} (C\hat{B} - d)). \\
&= (C\hat{B} - d)^T (C(X^T X)^{-1} C^T)^{-1} (C\hat{B} - d).
\end{aligned}$$

1.13. Idea:

$$\begin{aligned}
\hat{B} &\sim N(B, \sigma^2 (X^T X)^{-1}). \\
C\hat{B} &\sim N(CB, \sigma^2 C(X^T X)^{-1} C^T).
\end{aligned}$$

Si H_0 es verdadera, es decir $H_0 : CB = d$ entonces:

$$\begin{aligned}
C\hat{B} &\sim N(d, \sigma^2 C(X^T X)^{-1} C^T). \\
(C\hat{B} - d)^T \frac{(C(X^T X)^{-1} C^T)^{-1}}{\sigma^2} (C\hat{B} - d) &\sim X_r^2.
\end{aligned}$$

Entonces se tiene lo siguiente:

- $\frac{\Delta SS_{res}}{\sigma^2} \sim X_r^2$, si H_0 es verdadera.
- $\frac{SS_{res}}{\sigma^2} \sim X_{n-p}^2$
- ΔSS_{res} depende solamente de \hat{B} ; \hat{B} y SS_{res} son independientes.
- ΔSS_{res} y SS_{res} son independientes.

Recordemos la siguiente regla de la distribución F:

$$\begin{aligned}
w &\sim X_{k1}^2. \\
V &\sim X_{k2}^2.
\end{aligned}$$

Si ambos elementos son independientes, entonces se tiene que:

$$\frac{\frac{w}{k1}}{\frac{V}{k2}} \sim F(k1, k2).$$

Por lo tanto, se tiene que:

$$\frac{\frac{\Delta SS_{res}}{\sigma^2 * r}}{\frac{SS_{res}}{\sigma^2 * (n-p)}} \sim F(r, n-p).$$

Se eliminan los σ^2 y se tiene:

$$\frac{\frac{\Delta SS_{res}}{r}}{\frac{SS_{res}}{(n-p)}} \sim F(r, n-p).$$

1.14. Residuales

Los residuales ordinarios se tienen de la siguiente forma:

$$e = (I - H)Y.$$

$$E(e_i) = 0.$$

$$var(e_i) = \sigma^2(1 - h_{ii}).$$

$$cov(e_i, e_j) = -\sigma^2 h_{ij}, \forall i \neq j.$$

Los residuales estandarizados son:

$$r_i = \frac{e_i}{\hat{\sigma} \sqrt{1 - h_{ii}}}.$$

No obstante, no tiene distribución t porque $\hat{\sigma}^2$ y e_i no son independientes.

Dado ello, creamos los residuales «estudentizados»:

$$e_i \sim N(0, \sigma^2(1 - h_{ii})).$$

$$\frac{e_i}{\sigma \sqrt{1 - h_{ii}}} \sim N(0, 1).$$

1.15. Residuales «estudentizados»

Supongamos que obtenemos la muestra sin la observación i -ésima ($n-1$ observaciones). Por lo tanto, estimamos la varianza de los errores σ^2 obteniendo:

$$\frac{(n-p-1)\hat{\sigma}_i^2}{\sigma^2} \sim X_{n-p-1}^2.$$

Entonces e_i y $\hat{\sigma}_i^2$ son independientes.

Asimismo, para obtener el residual estudentizado se tiene que:

$$\frac{\frac{e_i}{\sigma \sqrt{1 - h_{ii}}}}{\sqrt{\frac{(n-p-1)\hat{\sigma}_i^2}{\sigma^2(n-p-1)}}} = \frac{e_i}{\hat{\sigma}_i \sqrt{1 - h_{ii}}} \sim t_{(n-p-1)}.$$

$$t_i = \frac{e_i}{\hat{\sigma}_i \sqrt{1 - h_{ii}}} = r_i \sqrt{\frac{n-p-1}{n-p-r_i^2}}.$$

1.16. Supuesto de Normalidad de los errores

qq-plot de los residuales estudentizados.

1.17. Valores atípicos

Observaciones con un residual estudentizado alto son posibles valores atípicos.

1.18. Leverage (apalancamiento)

$$\hat{Y} = HY.$$

$$\hat{Y}_i = \sum_{j=1}^n h_{ij} Y_j = h_{ii} Y_i + \sum_{i \neq j} h_{ij} Y_j.$$

$$h_{ii} = X_i^T (X^T X)^{-1} X_i.$$

Si consideramos el modelo lineal simple, tendríamos que:

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{S_{xx}}.$$

Un h_{ii} alto indicaría un punto atípico en las X' s.