Análisis de Covarianza

ANCOVA

Raúl Frugone Zaror   Diego Rocha Retamal

Universidad Católica del Maule

June 30, 2025

Indice

Introducción

El ANCOVA amplía el ANOVA incorporando covariables, con el objetivo de controlar su efecto y así analizar con mayor precisión el impacto de los factores de interés, reduciendo la varianza residual y aumentando la potencia estadística. En palabras más simples, este modelo ajusta comparaciones cuando otra variable externa también influye.

Supuestos ANCOVA

Los supuestos que debemos comprobar previo a realizar un ANCOVA son:

- Linealidad entre la variable respuesta y la covariable

Supuestos ANCOVA

Los supuestos que debemos comprobar previo a realizar un ANCOVA son:

- Linealidad entre la variable respuesta y la covariable
- Independencia

Supuestos ANCOVA

Los supuestos que debemos comprobar previo a realizar un ANCOVA son:

- Linealidad entre la variable respuesta y la covariable
- Independencia
- Homogeneidad en las pendientes de la regresión

Supuestos ANCOVA

Los supuestos que debemos comprobar previo a realizar un ANCOVA son:

- Linealidad entre la variable respuesta y la covariable
- Independencia
- Homogeneidad en las pendientes de la regresión
- Homocedasticidad: los residuos deben tener varianza constante

Supuestos ANCOVA

Los supuestos que debemos comprobar previo a realizar un ANCOVA son:

- Linealidad entre la variable respuesta y la covariable
- Independencia
- Homogeneidad en las pendientes de la regresión
- Homocedasticidad: los residuos deben tener varianza constante
- Normalidad en los residuos

Comprobación de supuestos

Para el supuesto de **Linealidad**, se realiza lo siguiente:

Test

Prueba RESET para comprobar forma funcional de un modelo lineal, propuesta por Ramsey en 1969

Contraste de hipotésis
H_0: Forma funcional correcta entre las variables v/s H_1: Forma funcional incorrecta

Uso en R

A través de la función **resettest()** y librería lmtest

Comprobación de supuestos

Para el supuesto de **Independencia**, se realiza lo siguiente:

Test

Test de Durbin-Watson

Contraste de hipotésis

H_0: No hay autocorrelación de primer orden en los residuos ($\rho = 0$). v/s H_1: Existe autocorrelación ($\rho \neq 0$).

Uso en R

A través de la función **dwtest()** y librería lm

Comprobación de supuestos

Para el supuesto de **homogeneidad de pendientes**, se realiza lo siguiente:

Test

Contraste de interacción entre covariable y factor

Contraste de hipotésis

H_0: $\gamma_1 = \gamma_2 = \cdots = \gamma_k = 0$ (*todas las pendientes son iguales*) v/s H_1: $\exists j$ tal que $\gamma_j \neq 0$ (*al menos una pendiente difiere*)

Uso en R

modelointeracción $< -$ lm(Y trt * X, data = df)

car::Anova(modelo$_i$*nteracción*, *type* = 3)

Comprobación de supuestos

Para el supuesto de **Homocedasticidad**, se realiza lo siguiente:

Test

Breusch-Pagan

Contraste de hipotésis

H_0: $\text{Var}(\varepsilon) = \sigma^2$ constante v/s H_1: $\text{Var}(\varepsilon) \neq \text{Var}(\varepsilon_k)$ con $i \neq k$

Uso en R

A través de la función **bptest()** y librería lm

Comprobación de supuestos

Para el supuesto de **Normalidad**, se realiza lo siguiente:

Test

Shapiro-Wilk

Contraste de hipotésis

H_0: Los residuos siguen una distribución normal v/s H_1: Los residuos no siguen una distribución normal.

Uso en R

A través de la función **shapiro.test()**

Objetivos del ANCOVA

- Controlar el efecto de covariables

Objetivos del ANCOVA

- Controlar el efecto de covariables
- Reducir la variabilidad residual

Objetivos del ANCOVA

- Controlar el efecto de covariables
- Reducir la variabilidad residual
- Incrementar la potencia estadística

Objetivos del ANCOVA

- Controlar el efecto de covariables
- Reducir la variabilidad residual
- Incrementar la potencia estadística
- Ajustar medias de grupo

Objetivos del ANCOVA

- Controlar el efecto de covariables
- Reducir la variabilidad residual
- Incrementar la potencia estadística
- Ajustar medias de grupo
- Corregir sesgos por desbalance en covariables

Uso del ANCOVA

Algunos contextos en los que se puede utilizar el ANCOVA son:

- Experimentos con covariables continuas ajenas que puedan afectar al diseño (edad, temperatura, estatura, etc.)

Uso del ANCOVA

Algunos contextos en los que se puede utilizar el ANCOVA son:

- Experimentos con covariables continuas ajenas que puedan afectar al diseño (edad, temperatura, estatura, etc.)
- Ajuste de puntuaciones en estudios de intervención (en psicología o educación, donde los grupos pueden diferir en habilidades iniciales)

Uso del ANCOVA

Algunos contextos en los que se puede utilizar el ANCOVA son:

- Experimentos con covariables continuas ajenas que puedan afectar al diseño (edad, temperatura, estatura, etc.)
- Ajuste de puntuaciones en estudios de intervención (en psicología o educación, donde los grupos pueden diferir en habilidades iniciales)
- Ensayos clínicos al momento de controlar variables continuas (peso, edad, etc.) al comparar dosis o tratamientos

Diseño completamente al ázar

Uso: Cuando no existen fuentes de variación conocidas aparte del tratamiento.

$$Y_{ij} = \mu + \tau_j + \beta\left(X_{ij} - \overline{X}\right) + \varepsilon_{ij},$$

- $i = 1, \ldots, n$ niveles de tratamiento.
- $j = 1, \ldots, t$ réplicas por tratamiento.
- $\varepsilon_{ij} \sim N(0, \sigma^2)$, errores independientes.

Contraste de hipótesis

- Efecto de la covariable β:

$$H_0 : \ \beta = 0 \qquad \text{(la covariable no explica variación en } Y\text{)}$$

$$H_1 : \ \beta \neq 0 \qquad \text{(la covariable tiene efecto significativo)}$$

- Efecto medio de los niveles del tratamiento:

$$H_0 : \ \tau_1 = \tau_2 = \cdots = \tau_k = 0 \qquad \text{(no hay diferencias entre tratamientos, tras ajus}$$

$$H_1 : \ \exists j \text{ tal que } \tau_j \neq 0 \qquad \text{(al menos un tratamiento difiere luego del ajuste)}$$

Es importante confirmar el supuesto de linealidad para confirmar que no existe interacción entre el tratamiento y la covariable, de no ser asi, el modelo ANCOVA clásico no es la mejor opción.

Contraste de hipótesis

Los estadísticos de prueba para los contrastes son respectivamente :

$$F_{\text{cov}} = \frac{\text{SC}_{\text{cov}}/1}{\text{SC}_{\text{error}}/(N-k-1)} \sim F_{1,\, N-k-1},$$

donde SC_{cov} es la suma de cuadrados atribuible a la covariable después de ajustar tratamiento e intercepto.

$$F_{\text{trat}} = \frac{\text{SC}_{\text{trat}}/(k-1)}{\text{SC}_{\text{error}}/(N-k-1)} \sim F_{k-1,\, N-k-1},$$

donde SC_{trat} es la suma de cuadrados debida a los tratamientos bajo el supuesto de pendientes paralelas.

Estimación de parámetros

Modelo

$$Y_{ij} = \mu + \tau_j + \beta\left(X_{ij} - \overline{X}\right) + \varepsilon_{ij},$$

Función de Máxima Verosimilitud

$$L(\mu, \tau, \beta, \sigma^2) = \prod_{i=1}^{n}\prod_{j=1}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})\right)^2}{2\sigma^2}\right).$$

Log-Verosimilitud

$$\ell(\mu, \tau, \beta, \sigma^2) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{t}\left(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})\right)^2$$

Estimación para μ

$$\sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..}))=0$$

- Igualando $\frac{\partial \ell}{\partial \mu}$ a 0

Estimación para μ

$$\sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..}))=0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{t}y_{ij}-N\mu-t\sum_{i=1}^{n}\tau_j-\beta\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})=0$$

- Igualando $\frac{\partial \ell}{\partial \mu}$ a 0
- Separando sumatorias

Estimación para μ

$$\sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..}))=0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{t}y_{ij}-N\mu-t\sum_{i=1}^{n}\tau_j-\beta\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})=0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})=0$$

- Igualando $\frac{\partial \ell}{\partial \mu}$ a 0
- Separando sumatorias
- Recordemos que: $\sum_{j=1}^{n}(x_{ij}-\bar{x}_{..})=n(\bar{x}_{i.}-\bar{x}_{..})$

Estimación para μ

$$\sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..}))=0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{t}y_{ij}-N\mu-t\sum_{i=1}^{n}\tau_j-\beta\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})=0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})=0$$

$$\hat{\mu}=\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{t}y_{ij}=\bar{y}_{..}$$

- Igualando $\frac{\partial \ell}{\partial \mu}$ a 0
- Separando sumatorias
- Recordemos que: $\sum_{j=1}^{n}(x_{ij}-\bar{x}_{..})=n(\bar{x}_{i.}-\bar{x}_{..})$
- Despejando μ

Estimacion para τ_j

$$\sum_{j=1}^{n}\left(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})\right) = 0$$

Paso a paso

- Igualar $\partial \ell / \partial \tau_j = 0$

Estimacion para τ_j

$$\sum_{j=1}^{n}(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})) = 0$$

$$\sum_{j=1}^{t} y_{ij} - N\mu - N\tau_j - \beta\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = 0$$

Paso a paso

- Igualar $\partial\ell/\partial\tau_j = 0$
- Separar sumatorias

Estimacion para τ_j

$$\sum_{j=1}^{n}(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})) = 0$$

$$\sum_{j=1}^{t} y_{ij} - N\mu - N\tau_j - \beta\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = 0$$

$$\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = N(\bar{x}_{i.} - \bar{x}_{..})$$

Paso a paso

- Igualar $\partial\ell/\partial\tau_j = 0$
- Separar sumatorias
- Uso de identidad en x

Estimacion para τ_j

$$\sum_{j=1}^{n}(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})) = 0$$

$$\sum_{j=1}^{t} y_{ij} - N\mu - N\tau_j - \beta \sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = 0$$

$$\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = N(\bar{x}_{i.} - \bar{x}_{..})$$

$$\tau_j = \frac{\sum_{j=1}^{n} y_{ij} - N\mu - \beta\, N(\bar{x}_{i.} - \bar{x}_{..})}{N}$$

Paso a paso

- Igualar $\partial\ell/\partial\tau_j = 0$
- Separar sumatorias
- Uso de identidad en x
- Despejar τ_j

Estimacion para τ_j

$$\sum_{j=1}^{n}\left(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})\right) = 0$$

$$\sum_{j=1}^{t} y_{ij} - N\mu - N\tau_j - \beta\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = 0$$

$$\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = N(\bar{x}_{i.} - \bar{x}_{..})$$

$$\tau_j = \frac{\sum_{j=1}^{n} y_{ij} - N\mu - \beta\, N(\bar{x}_{i.} - \bar{x}_{..})}{N}$$

$$\tau_j = \bar{y}_{i.} - \bar{y}_{..} - \beta(\bar{x}_{i.} - \bar{x}_{..})$$

Paso a paso

- Igualar $\partial\ell/\partial\tau_j = 0$
- Separar sumatorias
- Uso de identidad en x
- Despejar τ_j
- Reemplazando

Estimacion para β

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{t} \left(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})\right) (x_{ij} - \bar{x}_{..}) = 0$$

- Igualar $\frac{\delta \ell}{\delta \beta}$ a 0

Estimacion para β

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{t}\left(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..})\right)(x_{ij}-\bar{x}_{..}) = 0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{t} y_{ij}(x_{ij}-\bar{x}_{..}) - \mu\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})$$

$$-\sum_{i=1}^{n}\tau_j\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})-\beta\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})^2 = 0$$

- Igualar $\frac{\delta \ell}{\delta \beta}$ a 0
- Separamos las sumatorias

Estimacion para β

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{t} \left(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..})\right)(x_{ij} - \bar{x}_{..}) = 0$$

$$\sum_{i=1}^{n} \sum_{j=1}^{t} y_{ij}(x_{ij} - \bar{x}_{..}) - \mu \sum_{i=1}^{n} \sum_{j=1}^{t} (x_{ij} - \bar{x}_{..})$$

$$- \sum_{i=1}^{n} \tau_j \sum_{j=1}^{t} (x_{ij} - \bar{x}_{..}) - \beta \sum_{i=1}^{n} \sum_{j=1}^{t} (x_{ij} - \bar{x}_{..})^2 = 0$$

$$\beta = \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{t} (x_{ij} - \bar{x}_{..})^2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{t} (y_{ij} - \mu - \tau_j)(x_{ij} - \bar{x}_{..})$$

- Igualar $\frac{\delta \ell}{\delta \beta}$ a 0
- Separamos las sumatorias
- usando $\sum (x_{ij} - \bar{x}_{..}) = 0$ y $\sum \tau_j (x_{ij} - \bar{x}_{..}) = 0$

Estimacion para β

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{t}\left(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..})\right)(x_{ij}-\bar{x}_{..}) = 0$$

$$\sum_{i=1}^{n}\sum_{j=1}^{t} y_{ij}(x_{ij}-\bar{x}_{..}) - \mu\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})$$

$$-\sum_{i=1}^{n}\tau_j\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..}) - \beta\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})^2 = 0$$

$$\beta = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij}-\bar{x}_{..})^2}\cdot\sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij}-\mu-\tau_j)(x_{ij}-\bar{x}_{..})$$

- Igualar $\frac{\delta \ell}{\delta \beta}$ a 0
- Separamos las sumatorias
- usando $\sum(x_{ij}-\bar{x}_{..}) = 0$ y $\sum \tau_j\,(x_{ij}-\bar{x}_{..}) = 0$
- Despejando β

Estimacion para σ^2

$$-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{j=1}^{t}\sum_{i=1}^{n}\left(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..})\right)^2=0$$

- Igualar $\frac{\delta\ell}{\delta\sigma^2}$ a cero

Estimacion para σ^2

$$-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{j=1}^{t}\sum_{i=1}^{n}\left(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..})\right)^2=0$$

$$-N\sigma^2+\sum_{j=1}^{t}\sum_{i=1}^{n}(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..}))^2=0$$

- Igualar $\frac{\delta\ell}{\delta\sigma^2}$ a cero
- Multiplicar por $2\sigma^4$

Estimacion para σ^2

$$-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{j=1}^{t}\sum_{i=1}^{n}\left(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..})\right)^2=0$$

$$-N\sigma^2+\sum_{j=1}^{t}\sum_{i=1}^{n}(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..}))^2=0$$

$$\sigma^2=\frac{1}{N}\sum_{j=1}^{t}\sum_{i=1}^{n}\left(y_{ij}-\mu-\tau_j-\beta(x_{ij}-\bar{x}_{..})\right)^2$$

- Igualar $\frac{\delta\ell}{\delta\sigma^2}$ a cero
- Multiplicar por $2\sigma^4$
- Despejar σ^2

Tabla de estimadores para los parámetros

De esta manera, los estimadores para los parámetros del modelo son:

Parámetro	**Estimador**
μ	$\hat{\mu} = \bar{y}_{..}$
τ_j	$\hat{\tau}_j = \bar{y}_{.j} - \bar{y}_{..} - \beta(\bar{x}_{.j} - \bar{x}_{..})$
β	$\hat{\beta} = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij} - \mu - \tau_j)(x_{ij} - \bar{x}_{..})}{\sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij} - \bar{x}_{..})^2}$
σ^2	$\hat{\sigma}^2 = \dfrac{1}{N}\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij} - \mu - \tau_j - \beta(x_{ij} - \bar{x}_{..}))^2$

Tabla ANCOVA

Fuente de variación	**df**	**Sumas de cuadrados y productos**			**Ajustados para la regresión**			F_0	**Valor P**
		x	xy	y	y	df	Cuadrado medio		
Tratamiento	$t-1$	$SCTr_{xx}$	$SCTr_{xy}$	$SCTr_{yy}$	–	–	–	–	–
Error	$t(n-1)$	SCE_{xx}	SCE_{xy}	SCE_{yy}	SCE_m	$t(n-1)-1$	$\frac{SCE_m}{N-t-1}$	–	–
Total	$N-1$	SCT_{xx}	SCT_{xy}	SCT_{yy}	$SCE_m^{'}$	$N-2$	–	–	–
Tratamientos ajustados	2	–	–	–	$SCE_m - SCE_m^{'}$	$t-1$	$\frac{SCE_m - SCE_m^{'}}{t-1}$	$\frac{CMTr}{CME}$	$P(F_0 > F_{1-\alpha,(t-1),(N-t-1)})$

Sumas de cuadrados y productos cruzados

Donde:

$$SCT_{yy} = \sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij} - \bar{y}_{\bullet\bullet})^2 = \sum_{i=1}^{n}\sum_{j=1}^{t} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{N}$$

$$SCT_{xx} = \sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij} - \bar{x}_{\bullet\bullet})^2 = \sum_{i=1}^{n}\sum_{j=1}^{t} x_{ij}^2 - \frac{x_{\bullet\bullet}^2}{N}$$

$$SCT_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij} - \bar{x}_{\bullet\bullet}) \cdot (y_{ij} - \bar{y}_{\bullet\bullet}) = \sum_{i=1}^{n}\sum_{j=1}^{t} x_{ij} \cdot y_{ij} - \frac{y_{\bullet\bullet} \cdot x_{\bullet\bullet}}{N}$$

Sumas de cuadrados y productos cruzados

Donde:

$$SCTr_{yy} = n\sum_{j=1}^{t}(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2 = \frac{1}{n}\sum_{j=1}^{t} y_{\bullet j}^2 - \frac{y_{\bullet\bullet}^2}{N}$$

$$SCTr_{xx} = n\sum_{j=1}^{t}(\bar{x}_{\bullet j} - \bar{x}_{\bullet\bullet})^2 = \frac{1}{n}\sum_{j=1}^{t} x_{\bullet j}^2 - \frac{x_{\bullet\bullet}^2}{N}$$

$$SCTr_{xy} = n\sum_{j=1}^{t}(\bar{x}_{\bullet j} - \bar{x}_{\bullet\bullet}) \cdot (\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}) = \frac{1}{t}\sum_{j=1}^{t} x_{\bullet j} \cdot y_{\bullet j} - \frac{y_{\bullet\bullet} \cdot x_{\bullet\bullet}}{N}$$

Sumas de cuadrados y productos cruzados

Donde:

$$SCE_{yy} = \sum_{i=1}^{n}\sum_{j=1}^{t}(y_{ij} - \bar{y}_{\bullet j})^2 = SCT_{yy} - SCTr_{yy}$$

$$SCE_{xx} = \sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij} - \bar{x}_{\bullet j})^2 = SCT_{xx} - SCTr_{xx}$$

$$SCE_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{t}(x_{ij} - \bar{x}_{\bullet j}) \cdot (y_{ij} - \bar{y}_{\bullet j}) = SCT_{xy} - SCTr_{xy}$$

$$SCE_m' = SCT_{yy} - \frac{(SCT_{xy})^2}{SCT_{xx}} \qquad SCE_m = SCE_{yy} - \frac{(SCE_{xy})^2}{SCE_{xx}}$$

Enunciado

Como un ejemplo de un experimento en el que puede emplearse el análisis de covarianza, considérese el estudio realizado en INCHALAM S.A., empresa productora y exportadora de alambres y derivados para determinar si existe una diferencia en la resistencia de una fibra de monofilamento producida por tres máquinas diferentes.

Tabla de datos

Table: Datos de la resistencia a la ruptura (y = resistencia en libras, x = diámetro en 10^{-3} pulgadas)

Máquina 1		Máquina 2		Máquina 3	
y	x	y	x	y	x
36	20	40	22	35	21
41	25	48	28	37	23
39	24	39	22	42	26
42	25	45	30	34	21
49	32	44	28	32	15
207	126	216	130	180	106

Modelo

$$y_{ij} = \mu + \tau_j + \beta(x_{ij} - \bar{x}) + \varepsilon_{ij}$$

Donde:

- y_{ij}: Resistencia a la ruptura del i-ésimo material producido por la j-ésima máquina.
- μ: Media general ajustada.
- τ_j: Efecto medio del tratamiento (máquina j), con la restricción $\sum \tau_j = 0$.
- β: Coeficiente de regresión común para la covariable x (diámetro).
- x_{ij}: Diámetro correspondiente a y_{ij}.
- $\bar{x}$: Promedio global del diámetro.
- ε_{ij}: Error aleatorio, $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Contraste de hipótesis

- **Efecto del diámetro (covariable)**:

$$H_0 : \beta = 0 \quad \text{(el diámetro no afecta la resistencia)}$$
$$H_1 : \beta \neq 0 \quad \text{(el diámetro sí afecta la resistencia)}$$

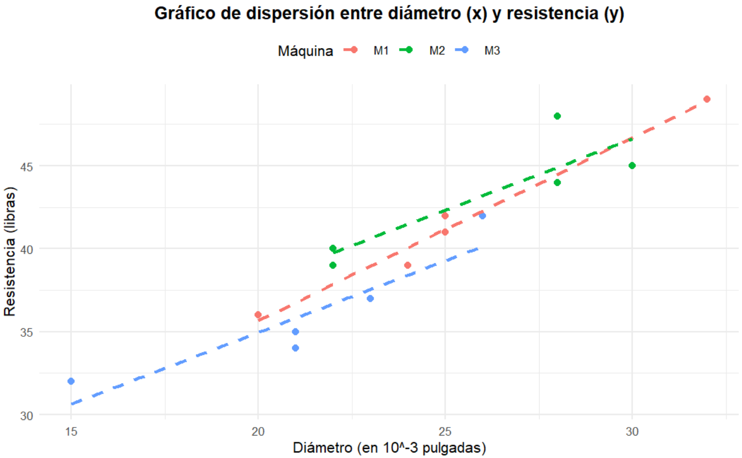
- **Efecto del tratamiento (máquina)**:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0 \quad \text{(no hay diferencias entre máquinas)}$$
$$H_1 : \exists\, l \neq k \text{ tal que } \tau_l \neq \tau_k \quad \text{(al menos una máquina difiere)}$$

Verificación de supuestos

Supuesto	**Valor-p**	**Interpretación**
Linealidad	0.2174	Se comprueba el supuesto, a favor de una forma correcta en las variables.
Independencia	0.219	Se comprueba el supuesto, a favor de que los errores no están autocorrelacionados.
Homocedasticidad	0.3663	Se comprueba el supuesto, a favor de que las varianzas de los residuos son constantes.
Homogeneidad de las pendientes	0.6367	Se comprueba el supuesto, a favor de que las pendientes se mantienen constantes.
Normalidad	0.7201	Se comprueba el supuesto, a favor de que los residuos se distribuyen de manera normal.

Gráfico de dispersión entre variables



Cálculo Sumas Cuadradas y productos cruzados

$$SCT_{xx} = \sum_{j=1}^{3}\sum_{i=1}^{5} x_{ij}^2 - \frac{x_{..}^2}{tn} = (20)^2 + (25)^2 + \cdots + (15)^2 - \frac{(362)^2}{(3)(5)} = 261.73$$

$$SCT_{xy} = \sum_{j=1}^{3}\sum_{i=1}^{5} x_{ij}y_{ij} - \frac{(x_{..})(y_{..})}{tn} = (20)(36)+(25)(41)+\cdots+(15)(32) - \frac{(362)(603)}{(3)(5)} = 282.60$$

$$SCT_{yy} = \sum_{j=1}^{3}\sum_{i=1}^{5} y_{ij}^2 - \frac{y_{..}^2}{tn} = (36)^2 + (41)^2 + \cdots + (32)^2 - \frac{(603)^2}{(3)(5)} = 346.40$$

Cálculo Sumas Cuadradas y productos cruzados

$$SCtr_{xx} = \frac{1}{n}\sum_{j=1}^{3} x_i^2 - \frac{x_{..}^2}{tn} = \frac{1}{5}\left[(126)^2 + (130)^2 + (106)^2\right] - \frac{(362)^2}{(3)(5)} = 66.13$$

$$SCtr_{xy} = \frac{1}{n}\sum_{i=1}^{3} x_i y_i - \frac{(x_{..})(y_{..})}{tn} = \frac{1}{5}\left[(126)(207) + (130)(216) + (106)(180)\right] - \frac{(362)(603)}{(3)(5)} = 96.00$$

$$SCtr_{yy} = \frac{1}{n}\sum_{j=1}^{3} y_i^2 - \frac{y_{..}^2}{tn} = \frac{1}{5}\left[(207)^2 + (216)^2 + (180)^2\right] - \frac{(603)^2}{(3)(5)} = 140.40$$

Cálculo Sumas Cuadradas y productos cruzados

$$SCE_{xx} = SCT_{xx} - SCtr_{xx} = 261.73 - 66.13 = 195.60$$

$$SCE_{xy} = SCT_{xy} - SCtr_{xy} = 282.60 - 96.00 = 186.60$$

$$SCE_{yy} = SCT_{yy} - SCtr_{yy} = 346.40 - 140.40 = 206.00$$

Cálculo Sumas Cuadradas y productos cruzados

$$SCE_m = SCE_{yy} - \frac{(SCE_{xy})^2}{SCE_{xx}} = 206.00 - \frac{(186.60)^2}{195.60} = 27.99$$

$$SCE_{m'} = SCT_{yy} - \frac{(SCT_{xy})^2}{SCT_{xx}} = 346.40 - \frac{(186.60)^2}{261.73} = 41.27$$

Cálculo Cuadrado Medio

$$CME = \frac{SCE_m}{N-t-1} = \frac{27.99}{11} = 2.54$$

$$CMtr = \frac{SCE_{m'} - SCE_m}{t-1} = \frac{41.27 - 27.99}{2} = 6.64$$

Comparamos se utiliza esta función en RStudio: qf(0.95, 2,11)

$$F_0 = \frac{CMtr}{CME} = \frac{6.64}{2.54} = 2.61; F_{0.05,2,11} = 3.982298$$

$$\text{valor-p} = P(F_0 > F_{\alpha,2,11}) = 0.1181$$

Tabla ANCOVA

Fuente de variación	**df**	**Sumas de cuadrados y productos**			**Ajustados para la regresión**			F_0	**Valor P**
		x	xy	y	y	df	Cuadrado medio		
Tratamiento	2	66.13	96	140.40	–	–	–	–	–
Error	12	195.6	186.6	206	27.99	11	2.54	–	–
Total	14	261.73	282	346.4	41.27	13	–	–	–
Tratamientos ajustados	–	–	–	–	13.28	2	6.64	2.61	0.1181

Estimación de parámetro β

$$\hat{\beta} = \frac{SCE_{xy}}{SCE_{xx}} = \frac{186.6}{195.6} = 0.954$$

Cálculamos el estadístico de prueba para $H_0 : \beta = 0$

$$F_0 = \frac{\frac{E_{xy}^2}{E_{xx}}}{CME} = \frac{178.01}{2.54} = 70.08$$

$$F_{0.95,1,11} = 4.84$$

$$\text{valor-p} = 0.00000423$$

Valores ajustados:

$$\hat{y}_{ij} = \mu + \tau_j + \hat{\beta}(x_{ij} - \bar{x}_{..})$$

$$\hat{y}_{ij} = \bar{y}_{..} + \left[\bar{y}_{.j} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{.j} - \bar{x}_{..})\right] + \hat{\beta}(x_{ij} - \bar{x}_{.j})$$

$$\hat{y}_{ij} = \bar{y}_{.j} + \hat{\beta}(x_{ij} - \bar{x}_{.j})$$

$$\hat{y}_{ij} = \bar{y}_{.j} + 0.954 \cdot (x_{ij} - \bar{x}_{.j})$$

Interpretación

No rechazamos la hipotesis nula para los tratamientos a favor de que los efectos medios de las máquinas son iguales, es decir no existen diferencias significativas entre máquinas de producción
Se rechaza H_0 para la hipotesis de la regresión a favor de que el diámetro del alambre afecta a la resistencia de la fibra de monofilamento.

Conclusión

El análisis de covarianza se presenta como una herramienta estadística especialmente valiosa cuando se desea comparar grupos categóricos (como diferentes máquinas en el contexto anterior) controlando por el efecto de una variables cuantitativa que puede influir en la variable respuesta.
ANCOVA no solo mejora la precisión del análisis al reducir la variabilidad no explicada, sino que también permite realizar comparaciones más justas entre los grupos. La capacidad de "ajustar" los efectos de variables externas lo convierte en una técnica poderosa en contextos experimentales y observacionales donde existen covariables relevantes que no pueden ser ignoradas.

Referencias

- Maxwell, S. E., Delaney, H. D. (2004). Designing Experiments and Analyzing Data: A Model Comparison Perspective (2nd ed.). Lawrence Erlbaum.
- Montgomery, D. C. (2017). Design and Analysis of Experiments (9ª ed.). John Wiley Sons.
- Sosa, S. (s.f.). ANCOVA en R. RPubs. recuperado en 18 de junio de 2025. https://rpubs.com/sebas$_{A}lf$/737954