

# ¿Cómo uso el paquete ANCOVA?

Jared Milian

## INTRODUCCIÓN

Uno de los problemas con el que frecuentemente se enfrenta el investigador, es el de controlar aquellos factores que no le es posible medir y cuyo efecto no puede justificar, los cuales constituyen el error experimental. Una de las formas de minimizar este error es mediante la aleatorización de los tratamientos y la utilización de material experimental muy homogéneo. Sin embargo, la aleatorización difícilmente cancela la influencia de las variables involucradas en el error y la disponibilidad de material experimental homogéneo no es frecuente en algunos experimentos, principalmente con animales, quedando restringidos a experimentos de laboratorio, invernadero o con animales de bioterio.(Pérez-Rodríguez et al. 2014).

Ronald Fisher en 1932 desarrolló una técnica conocida como Análisis de Covarianza o ANCOVA por sus siglas en inglés, que combina el Análisis de Regresión con el Análisis de Varianza. Covarianza significa variación simultánea de dos variables que se asume están influyendo sobre la variable respuesta. En este caso se tiene la variable independiente tratamientos y otra variable que no es efecto de tratamientos pero que influye en la variable de respuesta, llamada a menudo: covariable(Daniel 2014).

ANCOVA es apropiada para experimentos y estudios de observación que incluyen uno o más portadores numéricos (covariables). La variable en estudio es la que denominamos variable dependiente (Y) y la que influye en ella y no depende de los tratamientos, la variable independiente (X). Las covariables (x) corresponden a influencias molestosas que hacen a las unidades muestrales o experimentales diferentes.

Los usos más importantes del análisis de la covarianza son: 1. Para controlar el error y aumentar la precisión. 2. Ajustar medias de tratamientos de la variable dependiente a las diferencias en conjuntos de valores de variables independientes correspondientes. 3. Interpretación de la naturaleza de los efectos de los tratamientos. 4. Dividir una

covarianza total o suma de productos cruzados en componentes(Fernández Escobar et al. 2010).

El uso de la **covarianza** para controlar el error es un medio de aumentar la precisión con la cual los **efectos de los tratamientos** pueden medirse eliminando por regresión ciertos efectos reconocidos que no pueden ser o no han sido controlados efectivamente por el diseño experimental. Por ejemplo, un experimento de nutrición animal en el que se busca comparar el efecto de varias raciones en la ganancia de peso pero se tiene que los animales asignados a un bloque varían en peso inicial. Ahora, si el peso inicial está correlacionado con la ganancia de peso, una porción del error experimental en la ganancia puede deberse a diferencias en el peso inicial. Mediante el análisis de la covarianza, esta porción, una contribución que puede atribuirse a diferencias en el peso inicial puede calcularse y eliminarse del error experimental para ganancia(Badii Zabeh, Castillo, and Wong 2008).

Como se puede ver, el ANCOVA separa el efecto de las maniobras del efecto de las covariables, es decir, corrige la variable de respuesta eliminando la influencia de las covariables por el hecho de que estas varían conjuntamente con las maniobras o tratamientos, lo cual afecta la variable de desenlace(Pérez-Rodríguez et al. 2014).

El ANCOVA **solo debe ser realizado** si se cumplen tres supuestos: 1) la relación entre la covariable y el desenlace es lineal 2) existe homogeneidad de las pendientes 3) se comprueba la independencia entre la covariable y la maniobra o variable independiente.

VENTAJAS: Reduce el error experimental. Estima datos faltantes. Mejora interpretación de datos.

DESVENTAJAS: Puede quedar en desventaja sin el uso de bloques

Debido a la importancia que tiene en Análisis de covarianza en los diseños experimentales principalmente de índole biológico al momento de controlar el error, se buscó la manera de que, mediante una función llamada “*ANCOVA*” creada en el programador de código: Rstudio(RStudio Team 2021), se automatizaran todos los cálculos necesarios para realizar el análisis de covarianza (específicamente para los casos en el que el diseño experimental fuera mediante bloques completos al azar) de una manera rápida y eficaz.El código fuente del paquete está alojado en la siguiente dirección de GitHub <>.

## MODELO DEL ANÁLISIS DE COVARIANZA

Para bloques completamente al azar: El ANCOVA tiene como modelo la siguiente ecuación:

$$y_{ij} = \mu + \tau_i + \rho_j + \beta(X_{ij} - \bar{X}) + \epsilon_{ij}$$

Donde:  $y_{ij}$  = Variable de respuesta medida en la  $j$ -ésima repetición y el  $i$ -ésimo tratamiento.

$\mu$  = Media general

$\tau_i$  = Efecto del  $i$ -ésimo tratamiento.

$\rho_j$  = Efecto del  $j$ -ésimo bloque o repetición.

$\beta$  = Coeficiente angular de la regresión.

$X_{ij}$  = Variable independiente o covariable.

$\bar{X}$  = Media general de la covariable.

$\epsilon_{ij}$  = Error experimental.

NO OLVIDAR QUE: Otra suposición necesaria para el análisis correcto de covarianza, es que la variable concomitante  $X$ , no debe ser afectada por los tratamientos.

## OBJETIVO DEL PROYECTO

Como ya se vió, el análisis de covarianza es un método estadístico que permite si una eliminar la heterogeneidad causada en la variable de interés (variable dependiente) por la influencia de una o más variables cuantitativas (covariables). Para la realización del análisis de covarianza se necesitan hacer diferentes cálculos, con ellos se arma una tabla resumen que más tarde permitirá ver que tanto influyó nuestra covariable en el diseño experimental.

**El objetivo de este proyecto** es crear una función la cual automatice en unos cunatos clics todos los cálculos que se hacen (muchas veces a mano o en Excel) para crear una tabla resumen de análisis de covarianza y calcular las  $F$  calculadas y tabuladas que nos permita ver que tan influyente resultó nuestra covariable en el diseño experimental previamente hecho, así como tambien ilustrar mediante ejemplos el uso del paquete.

## ¿CÓMO USAR LA FUNCIÓN ANCOVA? :0

1. Para usar el paquete que alberga a dicha función usted deberá instalarla en su computadora ejecutando en Rstudio el siguiente comando:

```
devtools::install_github("")
```

Tambien deberá asegurarse de que tiene instalado el paquete Knitr el cual le servirá para que los datos sean puestos en tablas y de esta manera se facilite su análisis (ojo, solo hay que instalarla, la función ya la carga por usted):

```
install.packages("knitr")
```

2. Una vez ejecutados los comandos anteriores, usted podrá usar las cualquiera de las dos funciones que estan en el paquete: **ANCOVA** y **ANCOVA\_AJUST**
  - *ANCOVA*: Sirve para saber si su covariable es significativa en su experimento.
  - *ANCOVA\_AJUST*: Si su covariable ha salido significativa, puede usar esta función para ajustar la media de los tratamientos.
3. Ahora deberá llamar la función usando el nombre “ANCOVA” no sin antes haberla cargado:

```
library(ANCOVA)
```

La función le pedirá esto ya una vez cargada:

```
ANCOVA(“y”, “tratamiento”, “bloque”, “covariable”, significancia, data)
```

DONDE:

- **Y es igual** al nombre de su variable respuesta.
- **tratamiento es igual** al nombre que haya puesto para el caso de sus tratamientos. Ejemplo: T, trat, etc.
- **bloque es igual** al nombre que haya puesto para el caso de los bloque. Ejemplo: B, Bloq, etc.
- **covariable es igual** al nombre que haya puesto para el caso de la covariable. Ejemplo: X, cov, ganancia de peso, etc.
- **significancia es igual** al valor de alfa con el que quiera trabajar o error aceptado para el experimento
- **data es igual** al nombre que le asigne a su base de datos.

## EJEMPLO GENERAL

Siguiendo estas reglas usted debería completar a la función de la siguiente manera suponiendo que su base de datos es de un diseño experimental donde evalúa si la ganancia de peso de animales adquirida poniendo a prueba una cierta dieta, se ve influida por el peso inicial de dichos animales y tiene una base de datos nombrada “datos” y se trabaja con una significancia de 0.05:

DONDE:

- A la variable respuesta (y) la nombro “Ganancia.P”
- A sus tratamientos = “T”
- A sus bloques = “bloque”,
- A su covariable = “Peso.I”
- Ha decidido trabajar con una significancia = 0.5
- y su archivo lleva por nombre “datos”

ENTONCES

`ANCOVA(y = “Ganancia.P”, tratamiento = “T”, bloque = “bloque”, covariable = “Peso.I”, significancia = 0.05, data = datos)`

O TAMBIEN PUEDE HACERLO DE FORMA SINTETICA

`ANCOVA(“Ganancia.P”, “T”, “bloque”, “Peso.I”, 0.05, datos)`

**warning:** Cabe destacar que tanto los bloques como los tratamientos en caso de que lleven un nombre como caracter asignado, deberán ser renombrados manualmente con datos numericos.

## ¿QUÉ ME DEVOLVERÁ LA FUNCIÓN ANCOVA?

La función ANCOVA le devolvera **DOS TABLAS** en las cuales aparecen todos los resultados obtenidos en los calculos, con dichas tablas podrá hacer deducciones mediante pruebas de hipotesis acerca de que tanto impacto causó su covariable en los resultados de su diseño experimental.

La primera tabla resumen alberga...

- FV = FUENTE DE VARIACIÓN: a su vez guarda una fila con la suma de cuadrados y de productos cruzados de x,y, xy del TOTAL, BLOQUES, TRATAMIENTOS, ERROR Y ERROR AJUSTADO.
- Para cada uno de estos datos usted podrá obtener en la misma tabla información como sus grados de libertad (g.l).

#### La segunda tabla resumen:

- la F tabulada de tratamientos
- la F calculada de tratamientos
- la F tabulada de la covariable
- la F calculada de la covariable
- Estas F le servirán para hacer pruebas de hipótesis en donde si la F calculada de los tratamientos  $>$  a la F tabulada de los tratamientos,  $H_0$ : se rechaza, es decir, nuestra covariable si esta siendo significativa en los resultados de nuestro experimento.
- Pero si F calculada de la covariable  $>$  que la F tabulada de la covariable,  $H_0$  no se rechaza por lo que la covariable no esta siendo significativa para nuestros resultados.
- La tabla de las F trae consigo el coeficiente de covarianza el cual le ayudará básicamente al ajuste de medias por tratamiento que se hace cuando se concluye que nuestra covariable ha sido significativa en nuestro experimento. El paquete ANCOVA le puede ayudar hacer este ajuste mediante la función **ANCOVA\_AJUST**, esta función **es similar en cuanto a estructura a la función ANCOVA**, **se argumenta igual** sin embargo, en ella se debe **poner el valor numerico de coeficiente de covarianza** que ANCOVA nos arroje o que se haya conseguido de otra fuente, por lo tanto puede usarla sin haber usado ANCOVA antes:

`ANCOVA(coef_covarianza, "y", "tratamiento", "bloque", "covariable", data)`

**Para usar esta función no hace falta cargarla directamente pues una vez que cargue ANCOVA, ANCOVA\_AJUST estará lista para usarse.**

Para dejar más claro el como debe de usarse el paquete, a continuación se muestran ejemplos específicos que le permitan entender como usar esta función en sus datos para los que requiera hacer análisis de covarianza.

## EJEMPLOS ESPECIFICOS

**Ejemplo 1:** Se hizo un experimento en cerdos donde se busca probar que tan eficiente (en ganancia de peso) es una marca de alimento en 4 distintas raciones. Se trabajó con un diseño en bloques al azar, donde cada uno de los 3 bloques representaba un grupo diferente de cerdos, sin embargo los cerdos con los que se trabajaron eran distintos en su peso inicial. Se requiere saber si el peso inicial de los cerdos es significativa en los resultados obtenidos, haga el análisis de covarianza con un valor de alfa del 0.05

```
ruta <- "C:/Users/Jared/Downloads/proyecto"
datos <- read.csv(file = "datos.csv")
knitr::kable(datos)
```

tratamiento	bloque	PI	GP
1	1	30	165
2	1	24	170
3	1	34	156
4	1	41	201
1	2	27	170
2	2	31	169
3	2	32	179
4	2	32	173
1	3	20	130
2	3	20	171
3	3	35	138
4	3	30	200

Para construir la tabla resumen de análisis de covarianza, primero cargamos el paquete ANCOVA mediante la instrucción siguiente.

```
library(ANCOVA)
```

Ahora llamaremos a la función ANCOVA la cuál podrá recibir los argumentos de la siguiente manera para este caso:

```
ANCOVA(y="GP",
       tratamiento = "tratamiento",
       bloque = "bloque",
       covariable = "PI",
```

```
significancia = 0.05,
data = datos)
```

FV	g.l	SCX	SCY	SCXY	SC
:-----	---:	-----:	-----:	-----:	-----:
TOTAL	11	414.66667	4771.000	532.00000	NA
BLOQUES	2	76.16667	459.500	180.25000	NA
TRAT	3	226.66667	2469.667	330.66667	2108.123
ERROR	6	111.83333	1841.833	21.08333	1837.859
ERROR_AJUST	9	338.50000	4311.500	351.75000	3945.981

F_cal_trat	F_tab_trat	F_cal_cov	F_tab_cov	COEF_COVARIANZA
-----:	-----:	-----:	-----:	-----:
1.911756	5.409451	0.0108135	6.607891	0.1885246

o bien:

```
ANCOVA("GP", "tratamiento", "bloque", "PI",0.05, datos)
```

FV	g.l	SCX	SCY	SCXY	SC
:-----	---:	-----:	-----:	-----:	-----:
TOTAL	11	414.66667	4771.000	532.00000	NA
BLOQUES	2	76.16667	459.500	180.25000	NA
TRAT	3	226.66667	2469.667	330.66667	2108.123
ERROR	6	111.83333	1841.833	21.08333	1837.859
ERROR_AJUST	9	338.50000	4311.500	351.75000	3945.981

F_cal_trat	F_tab_trat	F_cal_cov	F_tab_cov	COEF_COVARIANZA
-----:	-----:	-----:	-----:	-----:
1.911756	5.409451	0.0108135	6.607891	0.1885246

PARA AMBOS CASOS RECIBIRÁ LA MISMA RESPUESTA: TABLA<sub>s</sub> RESUMEN DE ANÁLISIS DE COVARIANZA.



Para interpretar los datos solo basta con fijarse en la ultima tabla que nos arroja la función ANCOVA, donde usted podrá ver la F calculada de la covarianza, la F calculada para los tratamientos así como tambien la F tabulada para estos mismos dos. Si su F calculada para los tratamientos es mayor que la F tabulada de los tratamientos, se intuye que su covariable esta haciendo efecto en al menos uno de los bloques. Si por el contrario su F calculada para la covariable es mayor que la F tabulada para la covariable, podriamos decir que su covariable no es significativa.

En cuanto al coeficiente de covarianza, como puede ver esta en las tablas F, este valor se puede pegar y copiar en la función ANCOVA\_AJUST:

```
ANCOVA_AJUST(coef_covarianza = 0.1885246,
             y="GP",
             tratamiento = "tratamiento",
             bloque = "bloque",
             covariable = "PI",
             data = datos)
```

MEDIAS_AJUSTADAS_PARA_TRATAMIENTOS	
1	155.7541
2	170.8798
3	156.9126
4	190.4536

Ejecutar y listo, las medias ajustadas de sus tratamientos :)

Los datos tambien se pueden acomodar en un data frame. Veamos un ejemplo:

**Ejemplo 2:** En un experimento se probaron 3 variedades rendidoras de melón (Hiline, Moonshine y Caravelle), bajo un diseño en bloques completos al azar con 4 bloques. La variable respuesta fue la producción promedio (ton/ha) registrada en cada unidad experimental, la cual consistió en un cuadro de terreno de 10 metros cuadrados. Después de las labores de siembra y cultivo, debido a una helada, se observó que el número de plantas de melón en cada unidad experimental era muy heterogéneo, por lo que esto podria influir de manera negativa los resultados del experimento. Se decidió entonces usar el análisis de covarianza, utilizando como covariable al numero de plantas de melón por cada unidad experimental.

Los resultados del experimento se muestran a continuación:

```
MELON <- data.frame(bloque = c(1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 3, 4),
  tratamiento = c(1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3),
  produccion = c(12, 13, 15, 15.5, 11.5, 12.6, 14,
    13.8, 8.6, 9.8, 11, 12.5),
  N_plantas = c(200, 180, 200, 150, 195, 178,
    160, 150, 220, 150, 180, 190))
```

¿La covariables influye sobre los resultados del experimento?, Responda este cuestionamiento con 95% de confiabilidad

1. Procedemos a cargar(en caso de que aun no se haya hecho),llamar y aurgumentar a la función:

```
library(ANCOVA)
ANCOVA(y = "produccion",
  tratamiento = "tratamiento",
  bloque = "bloque",
  covariable = "N_plantas",
  significancia = 0.05,
  data = MELON)
```

FV	g.l	SCX	SCY	SCXY	SC
:-----	---:	-----:	-----:	-----:	-----:
TOTAL	11	5824.9167	45.409167	-209.60833	NA
BLOQUES	3	3045.5833	19.395833	-184.67500	NA
TRAT	2	463.1667	24.826667	-44.73333	24.772253
ERROR	6	2316.1667	1.186667	19.80000	1.017404
ERROR_AJUST	8	2779.3333	26.013333	-24.93333	25.789657

F_cal_trat	F_tab_trat	F_cal_cov	F_tab_cov	COEF_COVARIANZA
-----:	-----:	-----:	-----:	-----:
60.87121	5.786135	0.8318347	6.607891	0.0085486

- 2.En este caso la covariable no es significativa, pero se ejemplificará la obtención de las medias ajustadas:

```
library(ANCOVA)
ANCOVA_AJUST(coef_covarianza = 0.0085486,
              y = "produccion",
              tratamiento = "tratamiento",
              bloque = "bloque",
              covariable = "N_plantas",
              data = MELON)
```

	MEDIAS_AJUSTADAS_PARA_TRATAMIENTOS
1	13.84864
2	13.04909
3	10.42727

## CONCLUSIÓN

Debido a que este modelo estadístico es ampliamente usado en la rama de zootecnia, pues muchas veces en los experimentos de este indole se trabaja con covariables que pueden afectar los resultados obtenidos haciendo que no se cumpla el objetivo de estos, se decidió programar la función ANCOVA que le permitirá obtener en cuestión de segundos una tabla resumen del análisis de covarianza con la cual podrá hacer sus conclusiones acerca de que tan significativo resulta ser el efecto causado de una covariable en los resultados obtenidos en su diseño experimental.

## LITERATURA CITADA

- Badii Zabeh, Mohammad Hosein, J Castillo, and A Wong. 2008. "Uso de análisis de Covarianza (ANCOVA) En Investigación Científica (Use of Covariance Analysis (ANCOVA) in Scientific Research)." *Innovaciones de Negocios* 5 (9): 25–38.
- Daniel, Prado Campos Carlos. 2014. *Introducción Al Diseño y Análisis de Experimentos*. Editorial Academica Espanola.
- Fernández Escobar, R., Antonio Trapero Casas, Juan Domínguez Giménez, and Andalusia Consejería de Agricultura y Pesca. 2010. *Experimentación En Agricultura*.
- Pérez-Rodríguez, Marcela, Lino Palacios-Cruz, Jorge Moreno, Rodolfo Rivas-Ruiz, and Juan O Talavera. 2014. "Investigación Clínica XIX. Del Juicio Clínico Al análisis de Covarianza." *Revista Médica Del Instituto Mexicano Del Seguro Social* 52 (1): 247–62.
- RStudio Team. 2021. "RStudio: Integrated Development Environment for R." RStudio PBC. <https://www.rstudio.com/>.