

# Manual de uso del paquete RLMqual

Libni Henry Ventura Pérez

## Introducción

En la actualidad la mayoría de las investigaciones, utilizan métodos o modelos estadísticos que les permite, calcular o predecir lo que se tiene que hacer para aumentar la producción de algún cultivo, mantener el rendimiento o explicar las relaciones causales entre variables, etc.

Uno de los modelos estadísticos mas utilizados en investigaciones científicas es la regresión lineal múltiple que se utiliza para predecir el valor de una variable de respuesta (Y) utilizando una o más variables predictoras (X) como entrada. Su objetivo es establecer una relación lineal entre las variables predictoras y la variable de respuesta mediante una fórmula matemática (Ovalle, 2021). En este caso se trabajó con variables cualitativas, que, aunque no parece tener ninguna relación, este toma un valor que influye en muchos aspectos, para trabajar este paquete de *regresión lineal múltiple cualitativo*, fue necesario transformar las variables categóricas en variables cuantitativas o *variables dummy*, que en este caso se utilizan cuando son variables binarias, estos tomarían valores de 0 y 1, el hacer esto ya no sería ningún impedimento para que una base de datos no sea analizado (Alfonso, 2019).

Una de las ventajas de utilizar este modelo estadístico es que permite la inclusión y análisis de variables categóricas en el modelo de regresión. Esto proporciona una mayor comprensión de los efectos diferenciales, un mejor control de variables confusas, una mayor flexibilidad en el análisis y una mejora en la precisión del modelo.

En este manual se presentan ejemplos de aplicación de la regresión lineal múltiple cualitativa, en el cual se muestran algunos ejemplos de aplicación mediante un paquete de R construido del cual se denominó RLMqual. El código fuente del paquete está alojado en la siguiente dirección de GitHub <https://github.com/HenrryVentura/RegresionMultipleCualitativa>

La instalación del paquete en cualquier computadora se realiza ejecutando en la consola de R lo siguiente:

```
devtools::install_github("HenryVentura/RegresionMultipleCualitativa")
```

## Objetivos

- Desarrollar un paquete en R, mediante el uso de funciones de regresión múltiple con variables cualitativas, para facilitar el uso de esta función utilizada en temas relacionados con zootecnia.
- Analizar datos, mediante regresión múltiple cualitativa.

## Modelo estadístico

Los modelos lineales múltiples siguen la siguiente ecuación:

$$Y_i = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni}) + e_i$$

Donde

- **$Y_i$ :** Es la variable dependiente o variable de respuesta. Representa la variable que se trata de predecir o explicar en función de las variables independientes.
- **$\beta_0$ :** Es el intercepto o constante. Representa el valor esperado de  $Y_i$  cuando todas las variables independientes son igual a cero.
- **$\beta_1, \beta_2, \dots, \beta_n$ :** Son los coeficientes de regresión. Representan la relación o impacto que tienen las variables independientes ( $X_1, X_2, \dots, X_n$ ) sobre la variable dependiente  $Y_i$ . Cada  $\beta_i$  indica cuánto cambia en promedio  $Y_i$  cuando la variable independiente  $X_i$  aumenta en una unidad, manteniendo constantes las demás variables.
- **$X_1, X_2, \dots, X_n$ :** Son las variables independientes o variables explicativas. Son las variables que se utilizan para explicar o predecir la variable dependiente  $Y_i$ .
- **$e_i$ :** Es el término de error o residual. Representa la diferencia entre el valor observado de  $Y_i$  y el valor estimado o predicho por el modelo. Es una medida del error de la predicción y se asume que sigue una distribución normal con media cero.

## Ejemplo de uso del paquete RLMqual

### Ejemplo 1

Se estudia el efecto que tiene el peso de en dos tipos de razas de bovinos, Holstein y Angus, y se estudia si también si influye el tipo de cría que se les da en este caso, “intensivo

y extensivo”, lo que nos interesa es conocer que tanto influye la raza del animal y el tipo de cría en el peso, por lo cual se analizan los siguientes datos:

```
df <- read.xlsx("./DatosInvestigacion.xlsx")
print(df)
```

Peso	Raza	Cria
500	Holstein	Intensivo
600	Angus	extensivo
550	Angus	Intensivo
700	Holstein	Intensivo
450	Angus	extensivo

Los datos cargados deben de tener una estructura como la anterior, la variable cualitativa debe de estar en la primera columna, y posteriormente se debe de tener las variables cualitativas, después de haber cargado nuestros datos, realizamos nuestro análisis con el paquete como se ejemplifica a continuación:

```
library(RLMqual)
resultados<-RLMqual(df)
print(resultados)
```

proporciona una lista con los siguientes datos:

<pre>\$Minimo [1] -100</pre>	<ul style="list-style-type: none"><li>• <b>\$Minimo:</b> Representa el valor mínimo de los residuos del modelo de regresión.</li></ul>
<pre>\$Maximo [1] 100</pre>	<ul style="list-style-type: none"><li>• <b>\$Maximo:</b> Indica el valor máximo de los residuos del modelo de regresión.</li></ul>
<pre>\$Residual_standard_error [1] 125</pre>	<ul style="list-style-type: none"><li>• <b>\$Residual_Standard_error:</b> Es el error estándar de los residuos del modelo de regresión.</li></ul>
<pre>\$R_cuadrado [1] 0.1554054</pre>	<ul style="list-style-type: none"><li>• <b>\$R_cuadrado:</b> Representa el coeficiente de determinación <math>R^2</math>, que indica la proporción de la varianza total de la variable de respuesta.</li></ul>
<pre>\$Ajustada_R_cuadrado [1] 0.5777027</pre>	<ul style="list-style-type: none"><li>• <b>\$Ajustada_R_cuadrado:</b> Proporciona una medida ajustada de la calidad del ajuste del modelo.</li></ul>

Y Finalmente proporciona los coeficientes.

```
tabla_resultados <- knitr::kable(resultados$resultados)
print(tabla_resultados)
```

	coeficientes	error_estandar	t	p_valor
intercepto	525	88.38835	5.9396970	0.0271938
RazaHolstein	575	176.77670	3.2526912	0.0829299
CriaIntensivo	25	153.09311	0.1632993	0.8852921

En conclusión, en este ejemplo, a un nivel de significancia de **0.25**, lo que implica que el valor inicial de 525 tiene un efecto significativo en el peso. Sin embargo, las variables "*RazaHolstein*" y "*CriaIntensivo*" no tienen un efecto significativo en el peso según el nivel de significancia seleccionado.

## Ejemplo 2

Se estudia el efecto que tiene el rendimiento de cultivos de maíz en dos variedades, A y B. Los datos recolectados son los siguientes:

```
df <- data.frame(
  Rendimiento = c(8, 7, 6, 7.5, 6.5),
  Variedad = c("A", "B", "A", "A", "B")
)
tabla_df <- knitr::kable(df)
print(tabla_df)
```

Rendimiento	Variedad
8.0	A
7.0	B
6.0	A
7.5	A
6.5	B

Posteriormente se lleva a cabo el análisis mediante la regresión múltiple cualitativa.

```
library(RLMqual)
resultados <- RLMqual(df)
tabla_resultados <- knitr::kable(resultados)
print(tabla_resultados)
```

```
$Minimo
[1] -1.166667
```

```
$Maximo
[1] 0.8333333
```

```
$Residual_Standard_error
[1] 0.8740074
```

```
$R_cuadrado
[1] 0.08333333
```

```
$Ajustada_R_cuadrado
[1] 0.3125
```

	coeficientes	error_estandar	t	p_valor
intercepto	7.166667	0.5046084	14.20243	0.0007563
variedadB	6.750000	0.6180165	10.92204	0.0016429

En resumen, a un nivel de significancia de **0.05**, se concluye que tanto el valor inicial del rendimiento como la variable "*VariedadB*" tienen efectos significativos en el rendimiento de los cultivos de maíz dado que el p-valor es menor que un nivel de significancia  $\alpha = 0.05$ .

Otra función mas que incluye en el paquete es la función de **“Predicción”** que ayuda a obtener una predicción de lo que no se conoce, para ejecutar esta función es necesario hacer el análisis de los datos, solo contando con los valores cualitativos, ejemplo:

```
df <- data.frame(Variedad = c("B", "A", "B", "B", "A"))
predicciones <- prediccion(Xn = df, ajuste = resultados)
print(predicciones)
      [,1]
1 13.916667
2  7.166667
3 13.916667
4 13.916667
5  7.166667
```

Estas predicciones representan los valores estimados del rendimiento para cada observación en el dataframe df, utilizando el modelo ajustado previamente.

## Conclusión

La regresión Lineal múltiple cualitativa es un modelo estadístico, que su empleo a temas relacionados con la zootecnia es muy útil debido a que su uso permite analizar los datos de una investigación y proyecta una salida de datos de los cuales se puede interpretar de que tan relacionado esta un valor con otro, al igual que se pueden hacer predicciones de lo que se puede obtener a una base de datos desconocido, es por ello que este paquete desarrollado en R, permite facilitar el análisis de datos cualitativos.

## Bibliografía

Alfonso, Q. J. (2019). *TEMA 5 Variables ficticias Cómo describir información cualitativa*.  
Ovalle, C. T. (2021). *Ejercicio Regresión lineal múltiple*. 22–23.