

Contenido

| | |
|----------------------------|----------|
| Punto 1 | 3 |
| Solución Punto1 | 3 |
| Punto 2 | 3 |
| Solución Punto 2 | 3 |
| Punto 3 | 5 |
| Punto 4 | 5 |
| Punto 5 | 6 |

Universidad Nacional De Colombia
Sede Medellín



Facultad de Ciencias

Departamento de Estadística

Taller 4

Daniel Felipe Villa Rengifo

Luis David Hernández Pérez

Juan Gabriel Carvajal Negrete

Modelos de Regresión

Febrero, 2025

Se hará el análisis de la base de datos **Credit** de la librería **ISLR** de R. Se tendrá como variable de respuesta **Married** y como variable explicativa **Rating**. Debe codificar las variables de respuesta con 1 para **Yes** y 0 para **No**. Las observaciones a ser incluidas para el análisis en cada grupo aparecen en la Tabla 1. Realice las siguientes actividades:

Punto 1

Realice una descripción de la base de datos. Contextualice el problema y explique cada una de las variables involucradas en el modelo.

Solución Punto1

Descripción de los datos: Es un conjunto de datos simulados que contiene información sobre diez mil clientes. El objetivo es predecir qué clientes dejarán de pagar su tarjeta de crédito.

Descripción de las variables

- **Married:** Variable categórica que indica si la persona está casada (Yes/No).
- **Rating:** Variable numérica que representa el rating de crédito de la persona.

Punto 2

Realice un análisis descriptivo de las variables que se van a tener en cuenta en el modelo. Concluya.

Solución Punto 2

Tabla 1: Distribución de la variable Married en la muestra analizada

| Married | Frecuencia | Porcentaje |
|---------|------------|------------|
| Yes | 42 | 63.63 % |
| No | 24 | 36.36 % |
| Total | 66 | 100 |

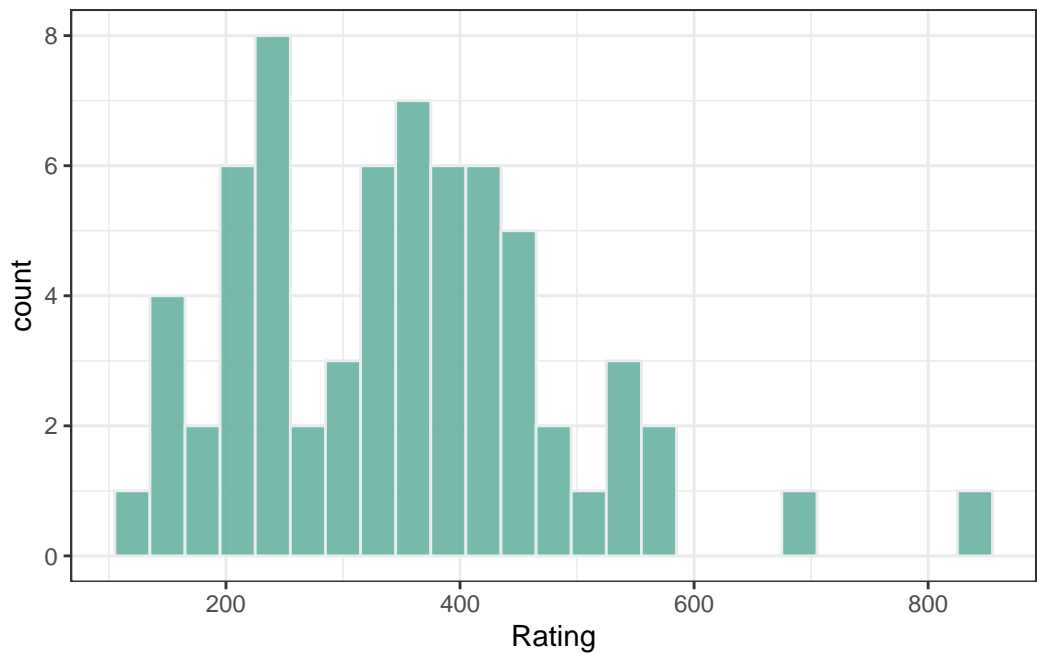


Figura 1: Distribución del Rating

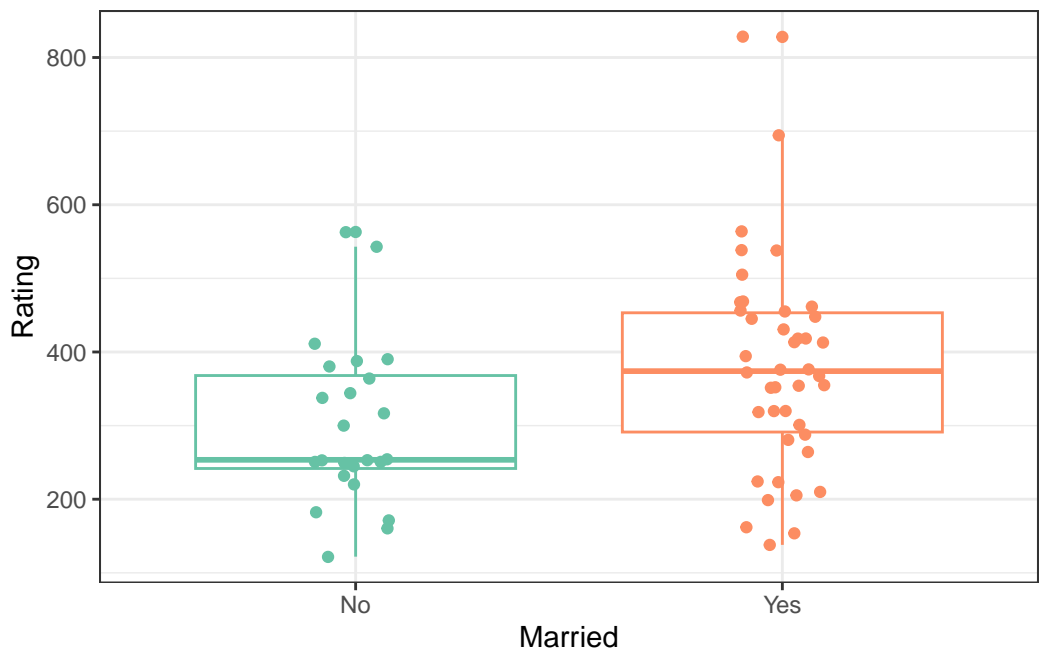


Figura 2: Boxplot Rating vs Married

Punto 3

Ajuste un modelo de regresión logística, muestre la tabla de parámetros ajustados y escriba la ecuación ajustada.

Tabla 2: Coeficientes ajustados

| Coeficiente | Estimación | Error estándar | valor z | Pr(> z) |
|--------------|------------|----------------|---------|-----------|
| (Intercepto) | -1.208379 | 0.812019 | -1.488 | 0.1367 |
| Rating | 0.005268 | 0.002372 | 2.221 | 0.0264 |

$$\hat{Y} = \frac{e^{-1.208379+0.005268 \cdot \text{Rating}}}{1 + e^{-1.208379+0.005268 \cdot \text{Rating}}}$$

Punto 4

Pruebe la significancia individual del parámetro que acompaña a la variable explicativa e interprete el valor de la estimación. También interprete la razón de odds.

Para la significancia de β_1 tenemos la siguiente hipótesis:

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$$

donde le estadístico de prueba es

$$Z_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

De la Tabla 2 tenemos que la variable **Rating**, $\beta_1 = 0.005268$ y

Punto 5

Determine si el modelo ajustado es mejor que el modelo nulo.

Para esta verificación utilizaremos la función `Anova()`

Tabla 3: Anova del modelo

| | Df | Deviance | Resid. Df | Resid. Dev | Pr(>Chi) |
|--------|----|----------|-----------|------------|-----------|
| NULL | NA | NA | 65 | 86.52359 | NA |
| Rating | 1 | 5.862324 | 64 | 80.66127 | 0.0154684 |

De la Tabla 3 podemos concluir que el modelo ajustado (modelo completo) es significativo, por tanto es mejor que el modelo nulo.