

# Informe de Análisis ANOVA para Calificaciones Estudiantiles

Zenaida Erika Valencia Condori

14 de abril de 2025

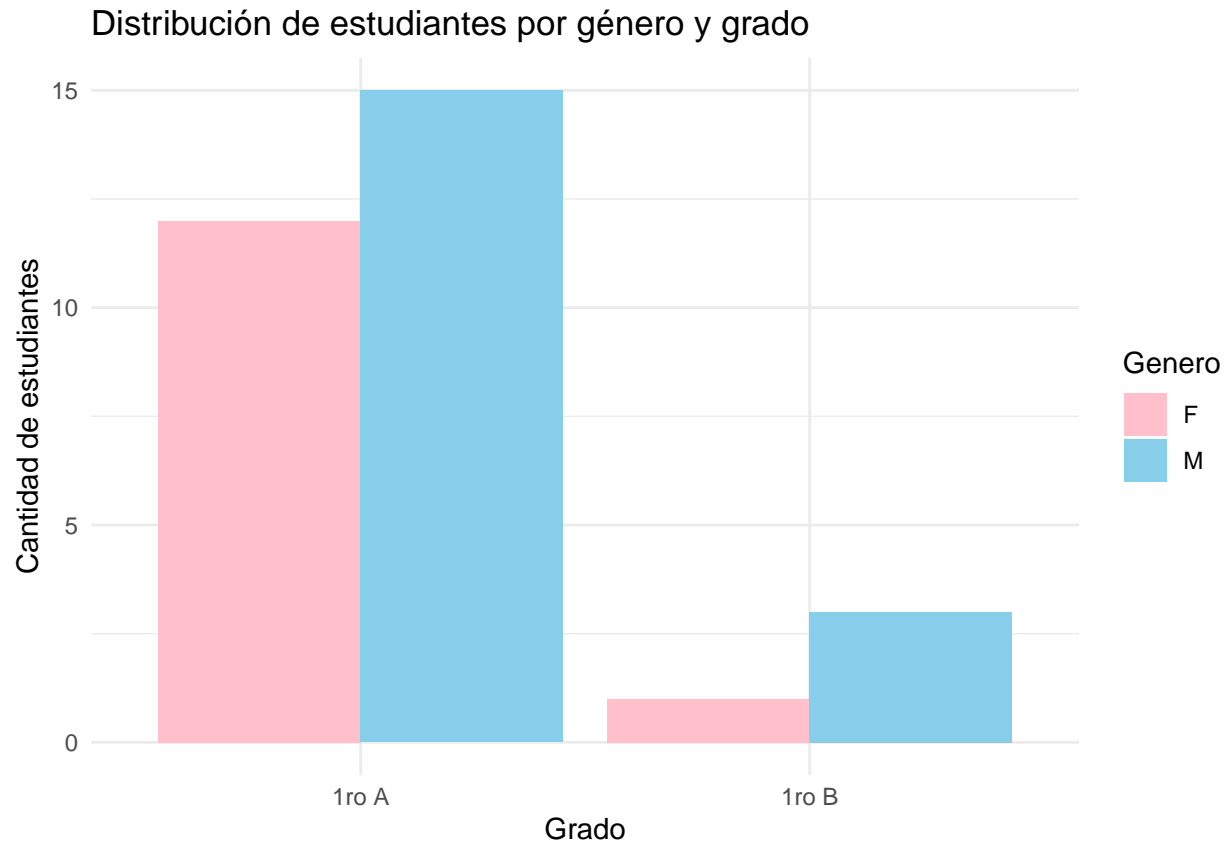
## Contents

```
# Tabla de distribución por género y grado
tabla_distribucion <- table(datos$Genero, datos$Grado)
kable(tabla_distribucion,
      caption = "Distribución de estudiantes por género y grado",
      col.names = c("1ro A", "1ro B"),
      row.names = TRUE)
```

Table 1: Distribución de estudiantes por género y grado

	1ro A	1ro B
F	12	1
M	15	3

```
# Visualización de la distribución
ggplot(datos, aes(x = Grado, fill = Genero)) +
  geom_bar(position = "dodge") +
  theme_minimal() +
  labs(title = "Distribución de estudiantes por género y grado",
       x = "Grado",
       y = "Cantidad de estudiantes") +
  scale_fill_manual(values = c("M" = "skyblue", "F" = "pink"))
```



```
# Creamos una función para calcular estadísticas por grupo de manera segura
calcular_estadisticas <- function(datos, variable, grupo_var, grupo_valor) {
  subset_datos <- datos[datos[[grupo_var]] == grupo_valor, ]
  return(c(
    n = nrow(subset_datos),
    Media = mean(subset_datos[[variable]], na.rm = TRUE),
    Mediana = median(subset_datos[[variable]], na.rm = TRUE),
    DesvEst = sd(subset_datos[[variable]], na.rm = TRUE),
    Min = min(subset_datos[[variable]], na.rm = TRUE),
    Max = max(subset_datos[[variable]], na.rm = TRUE)
  ))
}

# Estadísticas por género
generos <- levels(datos$Genero)
estadisticas_genero <- data.frame(
  Genero = generos,
  n = numeric(length(generos)),
  Media_4_1 = numeric(length(generos)),
  Mediana_4_1 = numeric(length(generos)),
  DesvEst_4_1 = numeric(length(generos)),
  Min_4_1 = numeric(length(generos)),
  Max_4_1 = numeric(length(generos))
)

for (i in 1:length(generos)) {
```

```

stats <- calcular_estadisticas(datos, "4.1", "Genero", generos[i])
estadisticas_genero$n[i] <- stats["n"]
estadisticas_genero$Media_4_1[i] <- stats["Media"]
estadisticas_genero$Mediana_4_1[i] <- stats["Mediana"]
estadisticas_genero$DesvEst_4_1[i] <- stats["DesvEst"]
estadisticas_genero$Min_4_1[i] <- stats["Min"]
estadisticas_genero$Max_4_1[i] <- stats["Max"]
}

kable(estadisticas_genero,
      caption = "Estadísticas descriptivas de la calificación 4.1 por género",
      digits = 2)

```

Table 2: Estadísticas descriptivas de la calificación 4.1 por género

Genero	n	Media_4_1	Mediana_4_1	DesvEst_4_1	Min_4_1	Max_4_1
F	13	2.15	2	1.41	1	5
M	18	2.11	2	1.18	1	5

```

# Estadísticas por grado
grados <- levels(datos$Grado)
estadisticas_grado <- data.frame(
  Grado = grados,
  n = numeric(length(grados)),
  Media_4_1 = numeric(length(grados)),
  Mediana_4_1 = numeric(length(grados)),
  DesvEst_4_1 = numeric(length(grados)),
  Min_4_1 = numeric(length(grados)),
  Max_4_1 = numeric(length(grados))
)

for (i in 1:length(grados)) {
  stats <- calcular_estadisticas(datos, "4.1", "Grado", grados[i])
  estadisticas_grado$n[i] <- stats["n"]
  estadisticas_grado$Media_4_1[i] <- stats["Media"]
  estadisticas_grado$Mediana_4_1[i] <- stats["Mediana"]
  estadisticas_grado$DesvEst_4_1[i] <- stats["DesvEst"]
  estadisticas_grado$Min_4_1[i] <- stats["Min"]
  estadisticas_grado$Max_4_1[i] <- stats["Max"]
}

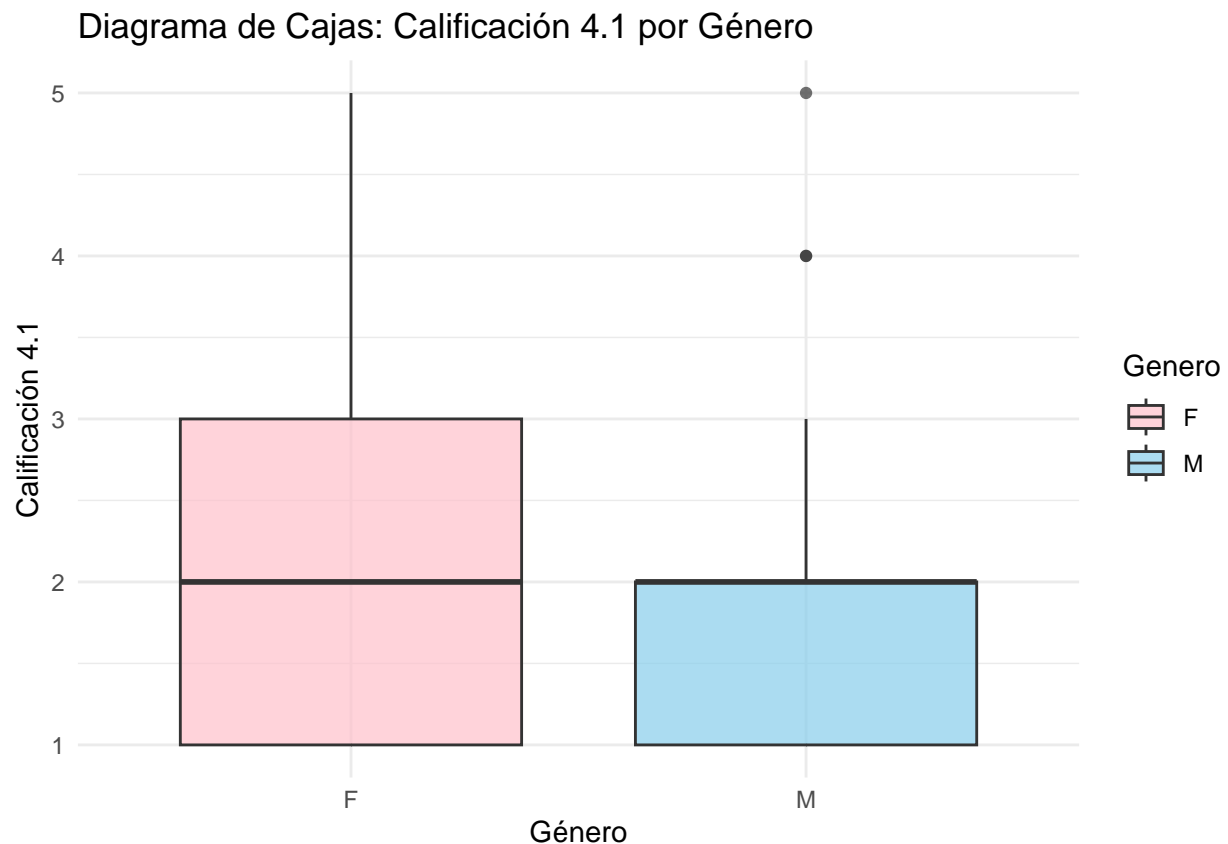
kable(estadisticas_grado,
      caption = "Estadísticas descriptivas de la calificación 4.1 por grado",
      digits = 2)

```

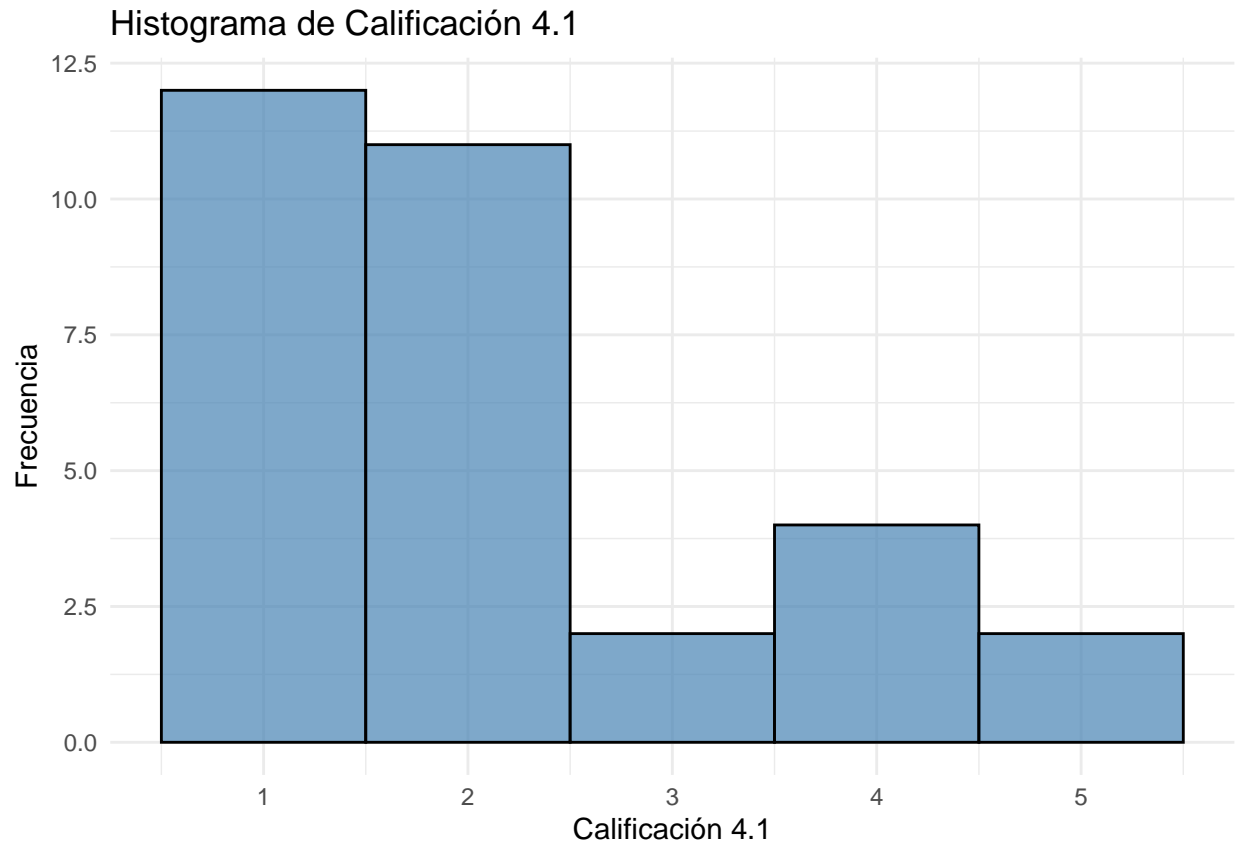
Table 3: Estadísticas descriptivas de la calificación 4.1 por grado

Grado	n	Media_4_1	Mediana_4_1	DesvEst_4_1	Min_4_1	Max_4_1
1ro A	27	2.22	2.0	1.31	1	5
1ro B	4	1.50	1.5	0.58	1	2

```
# Diagrama de cajas por género
p <- ggplot(datos, aes(x = Genero, y = `4.1`, fill = Genero)) +
  geom_boxplot(alpha = 0.7) +
  theme_minimal() +
  labs(title = "Diagrama de Cajas: Calificación 4.1 por Género",
       x = "Género",
       y = "Calificación 4.1") +
  scale_fill_manual(values = c("M" = "skyblue", "F" = "pink"))
print(p)
```



```
# Histograma de calificación 4.1
p <- ggplot(datos, aes(x = `4.1`)) +
  geom_histogram(bins = 5, fill = "steelblue", color = "black", alpha = 0.7) +
  theme_minimal() +
  labs(title = "Histograma de Calificación 4.1",
       x = "Calificación 4.1",
       y = "Frecuencia")
print(p)
```



```
# ANOVA por género
anova_genero <- aov(`4.1` ~ Genero, data = datos)
summary_anova_genero <- summary(anova_genero)
print(summary_anova_genero)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Genero      1   0.01  0.0138    0.008  0.928
## Residuals  29  47.47  1.6369
```

```
# Interpretación automática usando report
report(anova_genero)
```

```
## The ANOVA (formula: `4.1` ~ Genero) suggests that:
##
## - The main effect of Genero is statistically not significant and very small
## (F(1, 29) = 8.42e-03, p = 0.928; Eta2 = 2.90e-04, 95% CI [0.00, 1.00])
##
## Effect sizes were labelled following Field's (2013) recommendations.
```

```
# Prueba de homogeneidad de varianzas
levene_genero <- leveneTest(`4.1` ~ Genero, data = datos)
print(levene_genero)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value Pr(>F)
## group 1    0.889 0.3535
##       29
```

```

# ANOVA por grado
anova_grado <- aov(`4.1` ~ Grado, data = datos)
summary_anova_grado <- summary(anova_grado)
print(summary_anova_grado)

##              Df Sum Sq Mean Sq F value Pr(>F)
## Grado         1   1.82   1.817   1.154  0.292
## Residuals    29  45.67   1.575

# Interpretación automática usando report
report(anova_grado)

## The ANOVA (formula: `4.1` ~ Grado) suggests that:
##
## - The main effect of Grado is statistically not significant and small (F(1, 29)
## = 1.15, p = 0.292; Eta2 = 0.04, 95% CI [0.00, 1.00])
##
## Effect sizes were labelled following Field's (2013) recommendations.

# Prueba de homogeneidad de varianzas
levene_grado <- leveneTest(`4.1` ~ Grado, data = datos)
print(levene_grado)

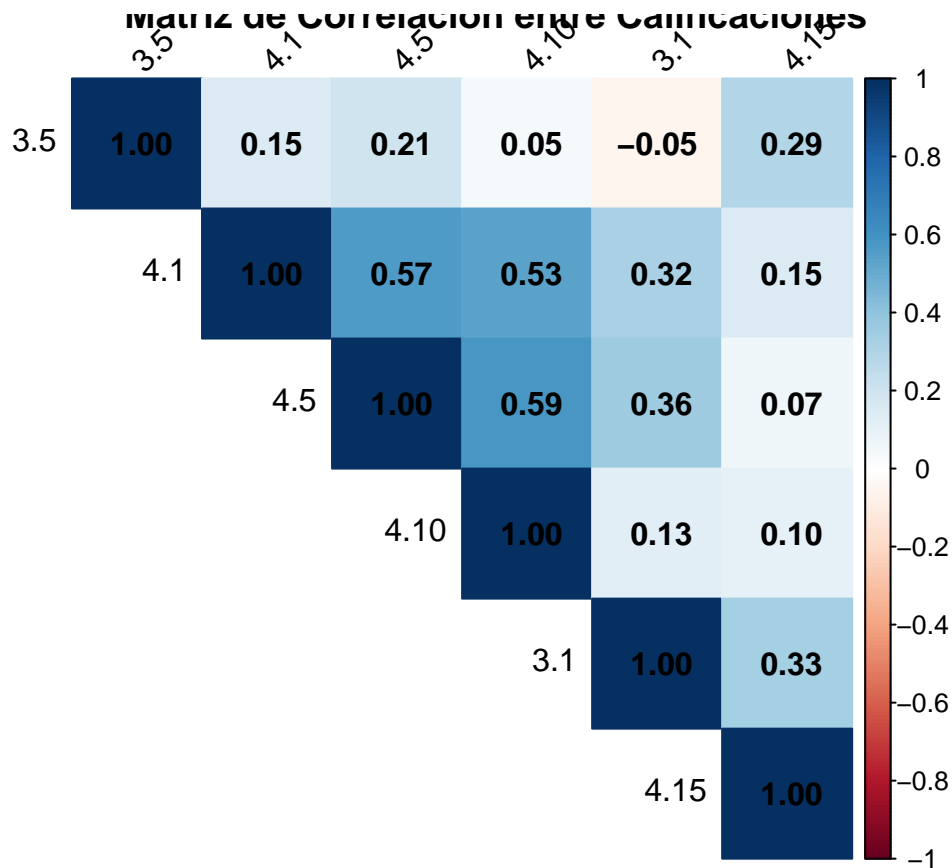
## Levene's Test for Homogeneity of Variance (center = median)
##              Df F value Pr(>F)
## group 1      1.033 0.3179
##          29

# Seleccionamos solo las columnas de calificaciones
columnas_calificaciones <- c("3.1", "3.5", "4.1", "4.5", "4.10", "4.15")
calificaciones <- datos[, columnas_calificaciones]
matriz_correlacion <- cor(calificaciones, use = "pairwise.complete.obs")
print(matriz_correlacion)

##              3.1          3.5          4.1          4.5          4.10          4.15
## 3.1  1.00000000 -0.05306602  0.3204712  0.35837016  0.12784738  0.33356841
## 3.5 -0.05306602  1.00000000  0.1531870  0.20552932  0.04987854  0.29498464
## 4.1  0.32047117  0.15318696  1.00000000  0.56791825  0.53007444  0.15158356
## 4.5  0.35837016  0.20552932  0.5679182  1.00000000  0.58595225  0.06607822
## 4.10 0.12784738  0.04987854  0.5300744  0.58595225  1.00000000  0.10080274
## 4.15 0.33356841  0.29498464  0.1515836  0.06607822  0.10080274  1.00000000

# Visualización de la matriz de correlación
corrplot(matriz_correlacion, method = "color",
          type = "upper", order = "hclust",
          tl.col = "black", tl.srt = 45,
          addCoef.col = "black", # Añadir coeficientes
          title = "Matriz de Correlación entre Calificaciones")

```



```
# Crear un dataframe para almacenar los resultados de ANOVA para todas las calificaciones
resultados_anova_multiple <- data.frame(
  Calificacion = character(),
  Valor_F = numeric(),
  P_valor = numeric(),
  Significativo = character(),
  stringsAsFactors = FALSE
)

# Lista de calificaciones a analizar
columnas_calificacion <- c("3.1", "3.5", "4.1", "4.5", "4.10", "4.15")

# Realizar ANOVA para cada calificación
for (col in columnas_calificacion) {
  # Crear la fórmula de manera segura
  formula_str <- paste0("`", col, "` ~ Genero")
  formula_obj <- as.formula(formula_str)

  # Ajustar el modelo ANOVA
  modelo <- aov(formula_obj, data = datos)
  resultado <- summary(modelo)[[1]]

  # Extraer el valor F y p-valor
  f_valor <- resultado$`F value`[1]
  p_valor <- resultado$`Pr(>F)`[1]
```

```

# Determinar si es significativo
significativo <- ifelse(p_valor < 0.05, "Sí",
                        ifelse(p_valor < 0.1, "Casi (=0.1)", "No"))

# Agregar al dataframe de resultados
resultados_anova_multiple <- rbind(resultados_anova_multiple,
                                    data.frame(Calificacion = col,
                                                Valor_F = round(f_valor, 3),
                                                P_valor = round(p_valor, 3),
                                                Significativo = significativo))
}

kable(resultados_anova_multiple,
      caption = "Resultados ANOVA para múltiples calificaciones por género")

```

Table 4: Resultados ANOVA para múltiples calificaciones por género

Calificacion	Valor_F	P_valor	Significativo
3.1	1.204	0.282	No
3.5	0.404	0.530	No
4.1	0.008	0.928	No
4.5	0.218	0.644	No
4.10	1.180	0.286	No
4.15	0.007	0.934	No