

Herramientas de Estadística (No)Paramétrica (Parte 5)

Juan C. Correa

Material de uso exclusivo para
INGENIO PANTALEON, S.A.
Diagonal 6, 10-31, Zona 10

Ciudad de Guatemala



- 1 ANOVA Factorial
- 2 Consideraciones Conceptuales
- 3 Contextualización



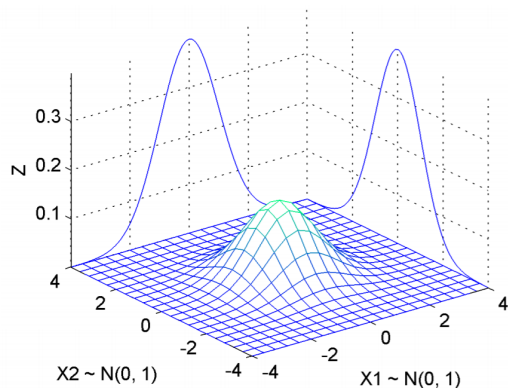
En **Estadística (No)Paramétrica Parte 4** estudiamos al análisis de varianza como una herramienta que nos permitía estimar la relación entre dos variables a través de la comparación entre tres o más grupos.

Ahora, vamos a concentrar nuestra atención hacia lo que se conoce como **Análisis de Varianza Factorial** que abarca la posibilidad de estimar la relación entre dos o más variables independientes y una variable dependiente. En otras palabras, el anova factorial es mucho más versátil que el anova unifactorial que estudiamos anteriormente.



Consideraciones Conceptuales

Como de costumbre, la asociación entre variables, en el contexto de un anova factorial, se mide a través del estadístico F, y se interpreta igual a como hemos visto antes. Sin embargo, ahora conviene entender el concepto de distribución normal multivariada y distribuciones marginales.



```
resultado_TCH <- aov(AOV$TCH ~ AOV$Entrenamiento + AOV$Gerente)
summary(resultado_TCH)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
AOV\$Entrenamiento	2	34302	17151	41.830	<2e-16	***
AOV\$Gerente	1	2262	2262	5.516	0.019	*
Residuals	1526	625677	410			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Ahora tenemos el cálculo de F para Entrenamiento y el cálculo de F para Gerente. Sin embargo, nuestra sintaxis solo modela la acción separada de cada variable independiente, sin considerar la **interacción** entre ellas.



```
resultado_TCH <- aov(AOV$TCH ~ AOV$Entrenamiento * AOV$Gerente)
summary(resultado_TCH)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
AOV\$Entrenamiento	2	34302	17151	41.863	<2e-16	***
AOV\$Gerente	1	2262	2262	5.520	0.0189	*
AOV\$Entrenamiento:AOV\$Gerente	2	1303	652	1.591	0.2041	
Residuals	1524	624373	410			

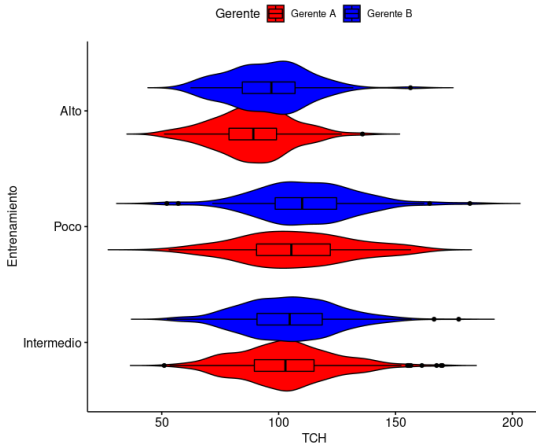
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

En orden de importancia, según su valor F, Entrenamiento es el factor que más incide en TCH, seguido por Gerente, mientras que la interacción entre Gerente y Entrenamiento no resulta significativa.



Contextualización

```
library(ggpubr)
ggviolin(AOV, x = "Entrenamiento", y = "TCH", merge = TRUE, fill = "Gerente",
  orientation="horiz",
  palette = c("red", "blue"),
  add = "boxplot", color = "black")
```

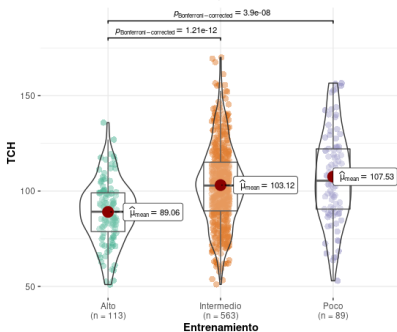


Contextualización

```
library(ggstatsplot)
grouped_ggbetweenstats(data = AOV,
  x = Entrenamiento, y = TCH,
  grouping.var = Gerente,
  p.adjust.method = "bonferroni")
```

Gerente: Gerente A

$F_{\text{Welch}}(2,176.31) = 35.86$, $p = 8.55\text{e-}14$, $\hat{\omega}_p^2 = 0.28$, $CI_{95\%} [0.17, 0.38]$, $n_{\text{obs}} = 765$

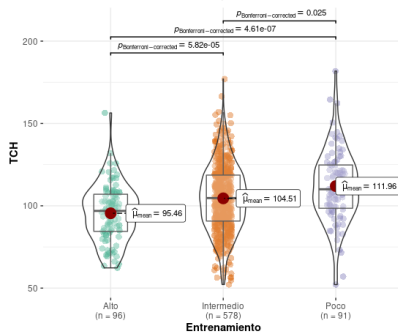


$\log_0(BF_{01}) = -20.33$, $\hat{R}^2_{\text{posterior Bayesian}} = 0.06$, $C_{\text{HDI}}^{95\%} [0.03, 0.10]$, $r_{\text{Cauchy}}^{0.25} = 0.71$

Pairwise test: **Games-Howell test**: Comparisons shown: **only significant**

Gerente: Gerente B

$F_{\text{Welch}}(2,161.88) = 17.97$, $p = 8.96\text{e-}08$, $\hat{\omega}_p^2 = 0.17$, $CI_{95\%} [0.07, 0.27]$, $n_{\text{obs}} = 765$



$\log_0(BF_{01}) = -10.81$, $\hat{R}^2_{\text{posterior Bayesian}} = 0.04$, $C_{\text{HDI}}^{95\%} [0.02, 0.07]$, $r_{\text{Cauchy}}^{0.25} = 0.71$

Pairwise test: **Games-Howell test**: Comparisons shown: **only significant**



En el anova factorial es fundamental conocer el tipo de diseño que está mostrándose con los datos. Se dice que hay un **Diseño Balanceado** cuando el número de observaciones por cada combinación de variables independientes es exactamente igual.

En cambio, se dice que hay un **Diseño no balanceado** cuando el número de observaciones no es igual entre todas las condiciones de las variables independientes.



```
library(car)  
Anova(resultado_TCH, type = "III")
```

Anova Table (Type III tests)

Response: AOV\$TCH

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	896298	1	2187.7267	< 2.2e-16	***
AOV\$Entrenamiento	22195	2	27.0873	2.757e-12	***
AOV\$Gerente	2127	1	5.1919	0.02283	*
AOV\$Entrenamiento:AOV\$Gerente	1303	2	1.5906	0.20415	
Residuals	624373	1524			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Chequeo de Supuestos

Un chequeo de supuestos básicos en el análisis de varianza es el cálculo de la prueba de heteroscedasticidad.

```
library(lmtest)
bptest(resultado_TCH)
library(car)
ncvTest(lm(AOV$TCH ~ AOV$Entrenamiento))
```

studentized Breusch-Pagan test

data: resultado_TCH

BP = 18.32, df = 5, p-value = 0.002571

Non-constant Variance Score Test

Variance formula: ~ fitted.values

Chisquare = 14.62692, Df = 1, p = 0.00013103



Siempre debe evaluarse la violación al supuesto de homoscedasticidad. Si la prueba bp o la prueba ncvTest nos arroja un estadístico considerablemente alejado de cero con p-value significativo, debemos desechar el modelo anova y correr algún alternativo robusto o un alternativo no-paramétrico.



```
library(WRS2)
t1way(AOV$TCH ~ AOV$Entrenamiento*AOV$Gerente)
```

Call:

```
t1way(formula = AOV$TCH ~ AOV$Entrenamiento * AOV$Gerente)
```

Test statistic: $F = 44.5781$

Degrees of freedom 1: 2

Degrees of freedom 2: 204.91

p-value: 0

Explanatory measure of effect size: 0.42

Bootstrap CI: [0.34; 0.5]

