

# Análisis ANOVA y primeras conclusiones

María Isabel Ruiz Martínez

February 1, 2023

## List of Tables

1	Muestra del dataset . . . . .	3
2	Resumen del dataset que tenemos . . . . .	3
3	Resultados del test two way ANOVA . . . . .	4
4	Coefficientes del modelo . . . . .	4

## List of Figures

1	Diagramas de caja y bigotes de las variables <i>map</i> y <i>action</i> . . . .	2
2	Distribuciones para diferentes niveles de los factores . . . . .	2
3	Distribuciones de las variables <i>map</i> y <i>action</i> dependiendo del valor de <i>year</i> . . . . .	3
4	Interacción entre las variables <i>map</i> y <i>action</i> . . . . .	5
5	Gráficas diagnósticas del modelo ANOVA . . . . .	5

## 1 Descripción del dataset

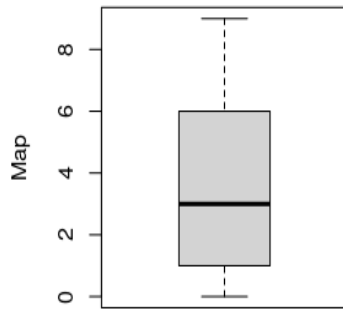
Los datos que se van a usar han sido recopilados por Luis Castillo Vidal y corresponden a la actividad de sus alumnos en la asignatura [Desarrollo Basado en Agentes](#).

El dataset, tras haber sido filtrados los registros erróneos, consta de 47828 filas correspondientes a los diferentes acciones de unos drones en una serie de mundos virtuales. En cada registro se detallan los siguientes atributos:

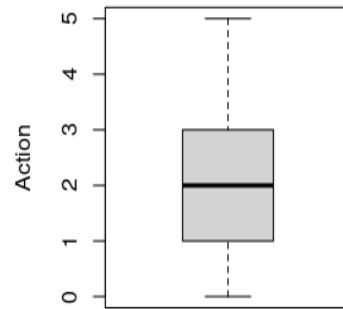
- *year*: identifica el curso académico en el que se realizó dicha acción.
- *group*: grupo de prácticas que ha progradado al dron que acomete la acción.
- *date*: fecha en la que se lleva a cabo la acción.
- *map*: mundo virtual en el que se ha realizado la acción.
- *action*: indica el tipo de acción realizada.

En la Tabla 1 se presentan los primeros seis registros del dataset. Además, en la Tabla 2 puede apreciarse un resumen de los datos que tenemos.

La Figuras 1a y 1b muestran, respectivamente, los gráficos de caja y bigotes de las variables *map* y *action*.

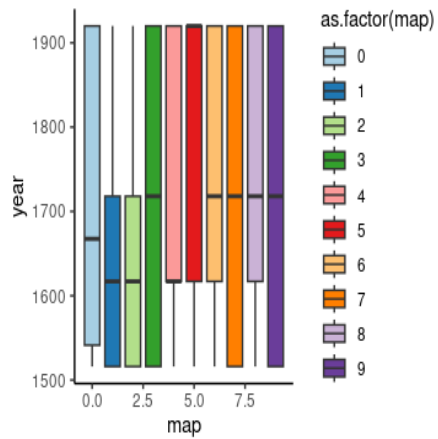


(a) Gráfico de caja y bigotes de la variable *map*

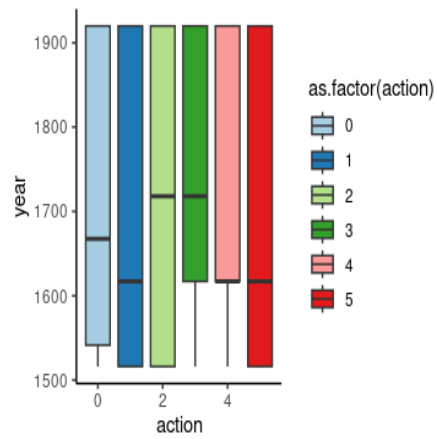


(b) Gráfico de caja y bigotes de la variable *action*

Figure 1: Diagramas de caja y bigotes de las variables *map* y *action*



(a) Distribución de los diferentes niveles del factor *map*



(b) Distribución de los diferentes niveles del factor *action*

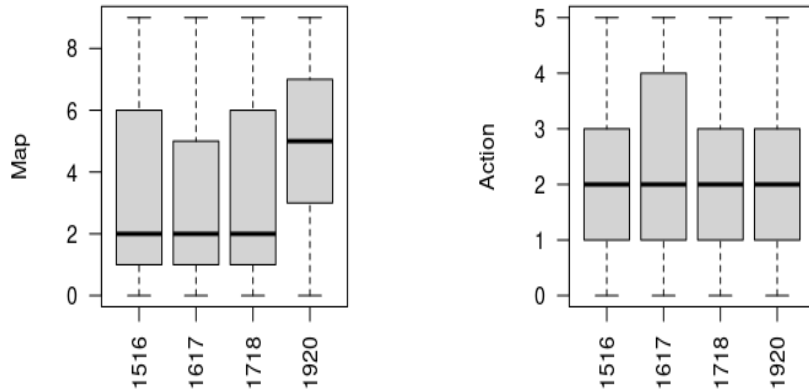
Figure 2: Distribuciones para diferentes niveles de los factores

	year	group	date	map	action
1	1516	Achernar	17/10/2015 19:41:45	0	0
2	1516	Bellatrix	17/10/2015 19:41:45	0	0
3	1516	Cerastes	17/10/2015 19:41:45	0	0
4	1516	Denebola	17/10/2015 19:41:45	0	0
5	1516	Elnath	17/10/2015 19:41:45	0	0
6	1516	Furud	17/10/2015 19:41:45	0	0

Table 1: Muestra del dataset

year	group	date	map	action
Min. :1516	Length:47828	Length:47828	Min. :0.000	Min. :0.000
1st Qu.:1516	Class :character	Class :character	1st Qu.:1.000	1st Qu.:1.000
Median :1617	Mode :character	Mode :character	Median :3.000	Median :2.000
Mean :1700			Mean :3.834	Mean :2.325
3rd Qu.:1920			3rd Qu.:6.000	3rd Qu.:3.000
Max. :1920			Max. :9.000	Max. :5.000

Table 2: Resumen del dataset que tenemos



(a) Distribución de la variable *map* en función de *year* (b) Distribución de la variable *action* en función de *year*

Figure 3: Distribuciones de las variables *map* y *action* dependiendo del valor de *year*

Además, podemos ver la distribución de los diferentes niveles del factor *map* y de los diferentes niveles del factor *action* en las Figuras 2a y 2b.

Por último, también podemos ver la distribución de las variables *map* y *action* en función del año (variable *year*) en las Figuras 3a y 3b.

## 2 Introducción

En este estudio inicial se desarrollará un modelo estadístico para determinar el efecto de los parámetros *map* y *action* (dos variables explicativas) en la variable respuesta *year*.

La relevancia de cada una de las variables en el modelo se determinará por el test *two way ANOVA* con un 5% de nivel de significancia y se empleará la técnica de los *mínimos cuadrados* para estimar los coeficientes del modelo considerado.

## 3 Two way ANOVA

Un resumen de los resultados obtenidos al realizar el test two way ANOVA se muestra en la Tabla 3. Puede observarse que la variable *map* es significativa al nivel 0, que la variable *action* es significativa al nivel 0.01 y que la variable *map:action* (el término de interacción) no es significativa. Así pues, puede concluirse que el dataset es homogéneo, es decir, las combinaciones *map:action* son estadísticamente iguales en todos los años considerados.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
map	1	62918413.58	62918413.58	2689.15	0.0000
action	1	101244.69	101244.69	4.33	0.0375
map:action	1	8797.03	8797.03	0.38	0.5398
Residuals	47824	1118946173.13	23397.17		

Table 3: Resultados del test two way ANOVA

La notación escalar del modelo ajustado al aplicar el test tiene la siguiente estructura:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1 \cdot x_2 + \epsilon \quad (1)$$

donde  $\beta_0$  es el intercepto,  $\beta_1$  y  $\beta_2$  son los coeficientes de los efectos principales,  $\beta_3$  es el coeficiente del término de interacción,  $x_1$  y  $x_2$  son los parámetros sometidos a investigación (en este caso,  $x_1$  representa el parámetro *mapa* y  $x_2$  representa la acción),  $y$  representa el año y  $\epsilon$  es el *término error*.

La Tabla 4 muestra los valores de los coeficientes de la fórmula que se han obtenido tras ajustar el modelo de regresión a los datos.

	x
(Intercept)	1655.03
map	12.31
action	-1.51
map:action	0.12

Table 4: Coeficientes del modelo

La Figura 4 muestra la interacción entre los parámetros *mapa* y *acción*. Así pues, puede observarse que todas las líneas de la gráfica siguen más o menos el mismo patrón, lo que evidencia que no hay una gran interacción entre ambos.

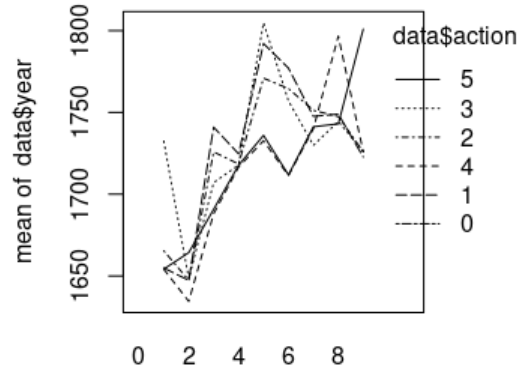


Figure 4: Interacción entre las variables *map* y *action*

La Figura 5 muestra que no se violan las suposiciones que hemos realizado sobre el modelo. La media y la varianza de los residuos no parece que varíe respecto de los valores ajustados. Como consecuencia, concluiré que podemos suponer la homocedasticidad. Además, si nos fijamos en el *Normal Q-Q plot*, puede observarse que los residuos son gaussianos.

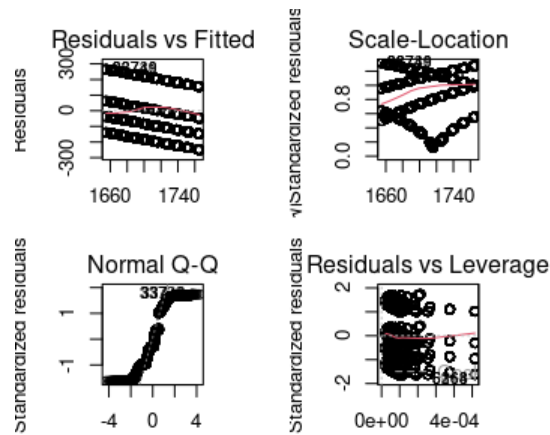


Figure 5: Gráficas diagnósticas del modelo ANOVA