

# Juego Interactivo Take Away!: La importancia de la calidad y la cantidad de variables en un modelo de regresión lineal

El presente es un resumen sobre un juego didáctico que utiliza el software estadístico R con el fin de potenciar la enseñanza de la regresión lineal. Un juego didáctico basado en el azar y en la computación adquiere un rol importante, ya que este no solo lleva al jugador a que comprenda y refuerce ciertos conceptos importantes de la estadística, sino también logra que el jugador vea cómo el uso de la tecnología puede llegar a facilitar la resolución de problemas. El objetivo de enseñanza principal del juego es comparar los coeficientes de determinación ajustados de los jugadores de acuerdo al modelo resultante del criterio de información de Akaike hacia atrás, el cual a su vez está basado en variables que surgen de una selección aleatoria.

- **Materiales y reglas:**

El juego implementa como materiales al Software R y al paquete Shiny, junto con un mazo de cartas específicamente diseñadas para llevarlo a cabo, además se necesita de un teléfono celular con conexión a internet con el fin de conectarse a la aplicación de Shiny. Cabe mencionar que la aplicación de Shiny utiliza los paquetes RPostgreSQL y DBI para conectarse a la base de datos ubicada en un servidor de SQL. En cuanto a las reglas iniciales del juego, el mínimo de participantes es de 2, mientras que el máximo es de 5. El mazo posee un total de 68 cartas. Dentro de las cuales están las cartas comodines: “¡Cambia Color!”, “Roba +2”, “Back” y “Paso”. Las dos ultimas con la finalidad de devolver la ronda de juego o saltarse el turno respectivamente.

- **Instrucciones del juego**

Al jugador se le brinda el link (<https://takeaway.shinyapps.io/appprueba/>) el cual deberá introducir en el navegador de su teléfono celular. Este link despliega una aplicación Shiny del programa R. Es importante que el jugador se familiarice con la aplicación antes de empezar a jugar, de esta manera podrá emplear este recurso adecuadamente durante el juego. En la portada de la aplicación el jugador puede observar la variable respuesta y las 52 variables predictoras, las cuales puede seleccionar mediante una herramienta de selección múltiple y evaluarlas mediante la aplicación. La primera pestaña llamada “Scatterplot”, despliega, en el caso de las variables numéricas, gráficos de relación lineal entre estas variables y la variable respuesta. En el caso de las variables categóricas despliega un gráfico de cajas donde se puede apreciar visualmente la media, el mínimo y el máximo de cada variable. La pestaña “Summary” despliega un resumen del modelo que surge entre la variable respuesta y las variables predictoras seleccionadas por el jugador al inicio. La pestaña “Data” despliega los datos que posee la base, esto con el fin de que el jugador pueda familiarizarse con los datos de la misma. Finalmente, la pestaña “Selección de Variables” despliega, por un lado: el proceso automático de selección de variables mediante el AIC (hacia atrás) y por otra despliega: el resumen del modelo resultante por el método del

AIC. Como se mencionó anteriormente, se posee un mazo con 68 cartas, de las cuales 52 son cartas de variables y 16 son cartas comodines. En relación a las 52 cartas, cada una de ellas posee un cierto color (ya sea morado, amarillo, verde y rojo), el cual le fue asignado a cada variable de manera aleatoria. Esto significa que, de las 52 cartas de variables, habrán 13 de color morado, 13 de color amarillo, 13 de color verde y 13 de color rojo. En la imagen adjunta se observa la estructura de la carta.

Para iniciar el juego se procede a repartir 7 cartas a cada jugador. Una vez hecho esto, se voltea una de las cartas que quedó en el mazo sin repartir, para tomarla como punto de



inicio (si esta carta corresponde a un comodín se debe retornar al mazo y tomar una nueva carta las veces que sea necesario). Al mismo tiempo, cada jugador deberá utilizar la aplicación para introducir en esta las variables pertenecientes a cada carta obtenida. El jugador procederá a realizar un análisis estadístico con base en sus conocimientos. Este, entonces, deberá observar las pestañas del scatterplot, la del summary y la de selección de variables (la cual se considera de mayor importancia, debido a que esta le indica al jugador que cartas puede eliminar). Antes de que inicie la primera ronda, se debe recalcar que, en todos los turnos, los jugadores deben de desechar una de sus cartas (es decir una de sus variables), la cual no podrá utilizarse nuevamente. Sin embargo, el jugador no tendrá que desechar ninguna de sus variables cuando aplique la carta comodín “Paso...” o “Back!”. En el caso hipotético de que un jugador se quede sin cartas, este deberá tomar, en su respectivo turno, como mínimo una carta del mazo.

Una vez que cada jugador conoce cuales cartas necesita eliminar, inicia el juego. El participante que inicia el juego, deberá eliminar su primera carta basándose en la carta que fue volteada del mazo, siempre y cuando ambas cartas coincidan en color o tipo de variable categórica. Se debe mencionar que cada vez que un jugador elimina una carta/variable de su “mano”, este también debe de removerla de su aplicación. Con base en lo anterior, puede ocurrir dos situaciones posibles: la primera; que el usuario no tenga cartas del color de la ronda, o la segunda; que en el summary del modelo o en la pestaña de selección de variables (el cual utiliza el criterio de Akaike) se determine que las cartas que posee el jugador son tan importantes que no deban ser eliminadas por este. Si cualquiera de estas dos situaciones ocurriera, el usuario deberá empezar a agarrar cartas del mazo y deberá detenerse hasta que obtenga una variable/carta que coincida con el color actual de la ronda o que coincida en tipo de variable categórica. No se debe dejar de lado el hecho de que cada vez que el participante agarra una carta de mazo, este debe de introducirla en su aplicación y debe de ir observando cómo se comporta el modelo con el ingreso de esta nueva variable. Esto debido a que, al incluir esta variable en el modelo, esta podría resultar no importante o podría dar como resultado que una o varias variables que antes eran relevantes, ahora ya no lo sean. Cabe destacar que, el criterio del participante toma un rol importante en este instante del juego ya que, si obtiene cartas del mazo que coincidan con el color de la ronda, pero que sean de suma importancia según el modelo, dependerá de él si decide seguir agarrando cartas o desperdiciar una variable considerada por el modelo como relevante. Independientemente de la decisión que tome el jugador, este deberá proceder a desechar alguna carta elegida. Una vez hecho esto, la carta de la cual se deshizo se convertirá en la carta base de la ronda actual. En caso contrario, si el jugador sí posee una carta que coincida en color o tipo de variable categórica, procederá a colocarla justo encima de la carta actual del mazo y esta, por tanto, se convertirá en la actual de la ronda.

Una vez que el mazo se quede sin cartas para agarrar, termina el juego. Los jugadores deberán introducir todas las variables/cartas, que les quedaron en la mano, en la aplicación (en caso de que aún no lo hayan hecho) y seguidamente deben fijarse en el  $R_a^2$  que aparece en la ventana de “selección de variables” (al final de la ventana). De esta forma aquel participante que posea el mayor  $R_a^2$  será el ganador.

- Situaciones de aprendizaje

Es acá donde se muestra el objetivo del juego, ya que cada jugador no busca tener un mayor o un menor número de variables, sino que, lo que realmente pretende, es obtener variables que en primera instancia tengan buena relación lineal con la variable respuesta. En segunda instancia; que posean un coeficiente de determinación significativo con una buena probabilidad asociada. Y en tercera instancia; que las variables que el jugador tenga en su posesión sean lo suficientemente importantes como para no ser eliminadas por el criterio de información de Akaike. De esta forma se aprecia que tener variables que cumplan estos aspectos se traducirá en un mayor coeficiente de determinación ajustado y por lo tanto en la victoria del jugador.

Finalmente, de lo anteriormente mencionado se puede derivar que el principal aprendizaje al finalizar el juego radica en que no importa la cantidad de variables que un modelo posea, sino su calidad y cuánta variabilidad aporten estas variables predictoras a la respuesta.