

SOL3070 Análisis de Datos Categóricos

Trabajo 1

Información

- Ponderación: 20% de la nota final del curso.
- Bonus: Responder la pregunta *bonus* NO es un requisito necesario para obtener puntaje completo. Responder incorrectamente la pregunta *bonus* no afectará negativamente la nota obtenida, pero responderla correctamente mejorará la nota obtenida en un máximo de 0.5 puntos (o en la cantidad necesaria para obtener nota máxima si la nota original fuera superior a 6.5)

Introducción

En esta tarea usarán el modelo lineal de probabilidad (LPM) y regresión logística para re-analizar los datos utilizados en el artículo “*It’s not just how the game is played, it’s whether you win or lose*” (2019). Este estudio utiliza un experimento online para identificar el efecto causal de las desigualdades de oportunidades y de resultados sobre creencias acerca de las causas de la desigualdad y percepciones de justicia. Para mayor contexto pueden revisar el artículo en el link indicado en el repositorio.

Específicamente, deberán trabajar con un modelo de regresión perteneciente a la sección “Supplementary Materials”. Tanto el artículo como el material suplementario se encuentran disponibles en el repositorio. En ambos documentos encontrarán destacadas las partes relevantes para desarrollar esta tarea.

Datos

Los datos están disponibles en el repositorio del curso para ser descargados. Visualización rápida de la base de datos:

```
path <- "/Users/Mauricio/Library/Mobile Documents/com~apple~CloudDocs/Teaching/ISUC/2023_2"
setwd(path)
```

```
data_paper <- read_csv("data_paper.csv")
data_paper %>% glimpse()
```

Rows: 857

Columns: 17

```
$ pid      <dbl> 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0~
$ wt       <dbl> 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1~
$ t        <chr> "RA2", "RA2", "RA2", "RA2", "RA1", "RA1", "RA1", "PR2", "PR2"~
$ score    <dbl> 551, 569, 606, 484, 566, 595, 528, 575, 569, 558, 565, 601, 5~
$ hs       <dbl> 49.19643, 50.80357, 55.59633, 44.40367, NA, 52.98308, 47.0169~
$ ioo      <dbl> 0.008035714, 0.008035714, 0.055963303, 0.055963303, NA, 0.029~
$ wg       <dbl> 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0~
$ fair     <chr> "Unfair", "Unfair", "Unfair", "Fair", "Unfair", "Fair", "Unfa~
$ most     <chr> "luck", "luck", "talent", "luck", "luck", "luck", "luck", "ru~
$ feelings <chr> "angry", "indifferent", "happy", "indifferent", "sad", "happy~
$ gender   <chr> "Male", "Male", "Male", "Male", "Female", "Female", "Female",~
$ age      <dbl> 21, 29, 25, 42, 30, 27, 53, 33, 25, 48, 45, 42, 32, 38, 46, 3~
$ religion <chr> "none", "catholic", "other", "protestant", "protestant", "non~
$ race     <chr> "white", "white", "white", "white", "white", "white", "white"~
$ inc      <chr> "$50,000-$75,000", "$0-$25,000", "$50,000-$75,000", "$25,000~
$ pol      <dbl> 1, 6, 6, 8, 8, 4, 3, 1, 4, 6, 6, 5, 1, 6, 6, 8, 1, 6, 6, 4, 3~
$ educ     <chr> "Some College", "College Degree", "High School", "Some Colleg~
```

Descripción de variables relevantes

- `wg` indica si el jugador ganó (`wg=1`) o perdió el juego (`wg=0`).
- `pid` indica si el jugador es “player 1” (`pid=0`) o “player 2” (`pid=1`).
- `hs` corresponde a “hand strength”. La variable utilizada acá es equivalente a la usada en el artículo pero multiplicada por 100 para facilitar la interpretación de resultados. Valores cercanos a cero indican que el jugador tenía carta débiles y valores cercanos a 100 indican que el jugador tenía cartas fuertes.
- `wt` indica si el jugador ganó (`wt=1`) o perdió la sesión de entrenamiento (`wt=0`).

Ejercicios

1. Usa un LPM para estimar el modelo implícito en la Figura S3 de la sección “Supplementary Materials” del artículo. Puedes encontrar mayores detalles sobre la especificación del modelo en la ecuación 4 y en la Tabla S3, columna 1. Escribe la ecuación de regresión y presenta un `summary()` de los resultados.

1.1 Interpreta el coeficiente asociado a `wt`.

1.2 De acuerdo al modelo estimado en 1., ¿cuál es el efecto marginal de “hand strength” sobre la probabilidad esperada de ganar el juego?

1.3 En base al modelo usado en 1., calcula las probabilidades esperadas de que un “player 1” que ganó el entrenamiento gane el juego si su “hand strength” es 50 y 60, respectivamente. Expresa formalmente las ecuaciones correspondiente a estas predicciones.

1.4. Agrega una interacción entre `hs` y `wt` al modelo estimado en 1. Escribe la ecuación de regresión y presenta un `summary()` de los resultados. Interpreta el efecto de “hand strength” estimado en este modelo.

1.5 En base a los resultados del modelo en 1.4. calcula las probabilidades esperadas de que un “player 1” que ganó el entrenamiento gane el juego si su “hand strength” es 60. Expresa formalmente la ecuación correspondiente a esta predicción.

2. Usa una regresión logística para estimar el modelo implícito en la Figura S3 de la sección “Supplementary Materials” del artículo. Puedes encontrar mayores detalles sobre la especificación del modelo en la ecuación 4 y en la Tabla S3, columna 1. Escribe la ecuación de regresión y presenta un `summary()` de los resultados.

2.1 Interpreta el coeficiente asociado a `wt`.

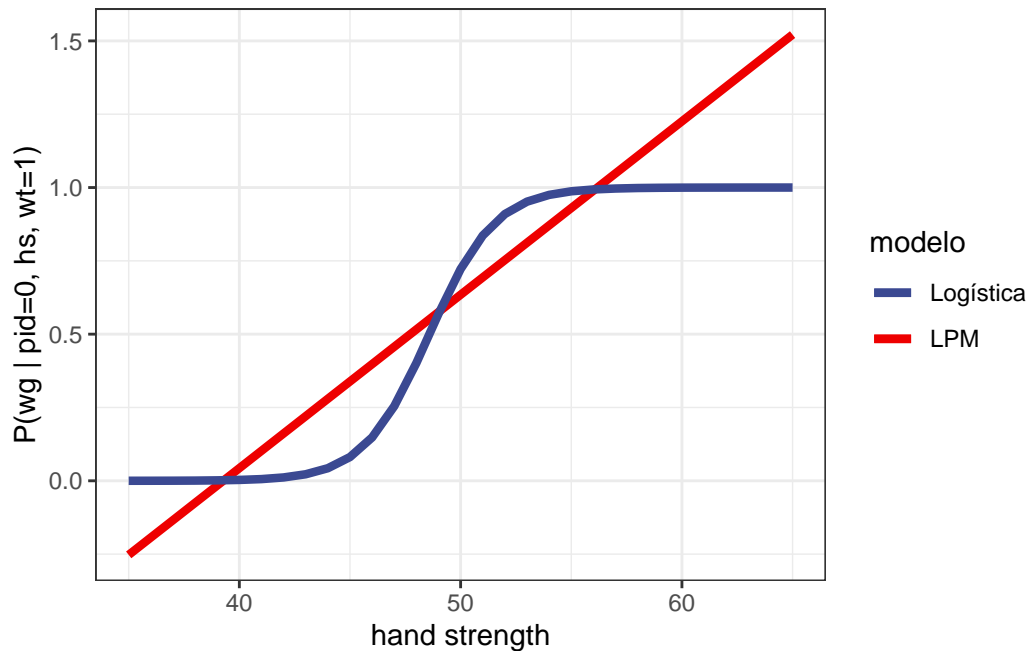
2.2 Transforma e interpreta el coeficiente de interés de 2.1. en términos de odds-ratios. Presenta el desarrollo formal.

2.3 De acuerdo al modelo estimado en 2, calcula las probabilidades esperadas de que un “player 1” que ganó el entrenamiento gane el juego si su “hand strength” es 50 y 60, respectivamente. Expresa formalmente las ecuaciones correspondiente a estas predicciones. Compara este resultado con el obtenido en 1.3.

2.4. De acuerdo al modelo estimado en 2., ¿cual es el efecto marginal de “hand strength” sobre la probabilidad esperada de ganar el juego? Calcula esta cantidad para un “Player 1” que ganó el entrenamiento si su “hand strength” es 50 y si su “hand strength” es 60. Expresa formalmente las ecuaciones correspondientes a estas cantidades. Compara esta respuesta con la respuesta dada en 1.2.

Bonus:

El siguiente gráfico describe las probabilidad predichas por el LMP estimado en 1. y la regresión logística estimada en 2. de que un “player 1” que ganó la sesión de entrenamiento gane el juego, para diferentes valores de “hand strength”.



3.1 Inspecciona visualmente la figura y determina la “hand strength” aproximada en la cual encontramos el mayor *efecto* de “hand strength” sobre la probabilidad de que un “player 1” que ganó el entrenamiento gane el juego.

3.2 ¿Cuál es el mayor efecto posible de “hand strength”?

3.3 Verdadero o falso: “el efecto no lineal de **hs** en el modelo de regresión logística (línea azul) se debe a que la especificación del modelo permite tal no-linealidad (por ejemplo, usando un término cuadrático o cúbico)”. Justifica tu respuesta.