

Práctica M212

Banjercito — Diplomado en Ciencia de Datos

Instructor: Alan Badillo Salas

1 Introducción

En esta práctica trabajaremos sobre el *dataset* de créditos de clientes para realizar algunas pruebas de hipótesis que determinen el **soporte** de las muestras que caen en la hipótesis y la **probabilidad** de que ocurra.

El *dataset* contiene los siguientes ejes de análisis:

- **ID de Crédito** — Cada crédito
- **Monto Aprobado** — Lo que se prestó en el crédito
- **Total de Abonos** — Número de pagos aportados al crédito
- **Suma de Abonos** — La suma total de los pagos aportados al crédito
- **Promedio de Abonos** — El promedio de cada pago aportado al crédito
- **Desviación de los Abonos** — Qué tanto se aleja del promedio de cada aportado al crédito
- **Porcentaje de Recuperación** — Cuánto del crédito se ha recuperado (pagado)

El objetivo es calcular de cada hipótesis planteada el **soporte** (número de muestras que cumplen la hipótesis) y la **probabilidad** de que ocurra (el número de muestras que cumplen la hipótesis entre el total de muestras).

2 Pruebas de Hipótesis

Una prueba de hipótesis la podemos entender cómo una pregunta o sentencia lógica que conforma un resultado probabilístico, por ejemplo, cuántas muestras o créditos fueron otorgados con un monto mayor a \$20,000 o cuántos créditos recuperaron más del 50% en su suma abonos aportados.

1. H_1 — Se otorga un crédito menor a \$10k (estricto)
2. H_2 — Se otorga un crédito entre \$10k y \$20k (inclusive)
3. H_3 — Se otorga un crédito mayor \$20k (estricto)

4. G_1 — Se recupera menos del 10% del crédito (estricto)
5. G_2 — Se recupera menos entre 10% y 50% del crédito (inclusive)
6. G_3 — Se recupera más del 50% del crédito (estricto)
7. $H_1 \wedge G_1$ — Crédito menor a \$10k con recuperación menor al 10%
8. $H_1 \wedge G_2$ — Crédito menor a \$10k con recuperación del 10% al 50%
9. $H_1 \wedge G_3$ — Crédito menor a \$10k con recuperación mayor al 50%
10. $H_2 \wedge G_1$ — Crédito entre \$10k y \$20k con recuperación menor al 10%
11. $H_2 \wedge G_2$ — Crédito entre \$10k y \$20k con recuperación del 10% al 50%
12. $H_2 \wedge G_3$ — Crédito entre \$10k y \$20k con recuperación mayor al 50%
13. $H_3 \wedge G_1$ — Crédito mayor a \$20k con recuperación menor al 10%
14. $H_3 \wedge G_2$ — Crédito mayor a \$20k con recuperación del 10% al 50%
15. $H_3 \wedge G_3$ — Crédito mayor a \$20k con recuperación mayor al 50%

3 Cálculo de probabilidades condicionales

Las probabilidades condicionales nos permitirán entender verdades ocultas en los datos, como la probabilidad de recuperar un porcentaje del crédito dado que se otorgó un crédito para algún monto aprobado.

1. $P(H_1|G_1)$ — Se otorga un crédito menor a \$10k dado que se recuperó menos del 10%
2. $P(G_1|H_1)$ — Recuperar menos del 10% dado que se otorga un crédito menor a \$10k
3. $P(H_1|G_2)$ — Se otorga un crédito menor a \$10k dado que se recuperó entre el 10% y 50%
4. $P(G_2|H_1)$ — Recuperar entre el 10% y el 50% dado que se otorga un crédito menor a \$10k
5. $P(H_1|G_3)$ — Se otorga un crédito menor a \$10k dado que se recuperó más del 50%
6. $P(G_3|H_1)$ — Recuperar más del 50% dado que se otorga un crédito menor a \$10k
7. $P(H_2|G_1)$ — Se otorga un crédito entre \$10k y \$20k dado que se recuperó menos del 10%

8. $P(G_1|H_2)$ — Recuperar menos del 10% dado que se otorga un crédito entre \$10k y \$20k
9. $P(H_2|G_2)$ — Se otorga un crédito entre \$10k y \$20k dado que se recuperó entre el 10% y 50%
10. $P(G_2|H_2)$ — Recuperar entre el 10% y el 50% dado que se otorga un crédito entre \$10k y \$20k
11. $P(H_2|G_3)$ — Se otorga un crédito entre \$10k y \$20k dado que se recuperó más del 50%
12. $P(G_3|H_2)$ — Recuperar más del 50% dado que se otorga un crédito entre \$10k y \$20k
13. $P(H_3|G_1)$ — Se otorga un crédito mayor a \$20k dado que se recuperó menos del 10%
14. $P(G_1|H_3)$ — Recuperar menos del 10% dado que se otorga un crédito mayor a \$20k
15. $P(H_3|G_2)$ — Se otorga un crédito mayor a \$20k dado que se recuperó entre el 10% y 50%
16. $P(G_2|H_3)$ — Recuperar entre el 10% y el 50% dado que se otorga un crédito mayor a \$20k
17. $P(H_3|G_3)$ — Se otorga un crédito mayor a \$20k dado que se recuperó más del 50%
18. $P(G_3|H_3)$ — Recuperar más del 50% dado que se otorga un crédito mayor a \$20k

Fórmula de la probabilidad condicional $P(A|B) = \frac{P(A \cap B)}{P(B)}$

4 Retos

Resuelve los siguientes retos analizando las probabilidades anteriores e intenta graficar los resultados.

Las siguientes preguntas se pueden resolver con la probabilidad condicional $P(H_i|G_j)$ donde $i = 1, 2, 3$ para algún monto aprobado y $j = 1, 2, 3$ para algún porcentaje de recuperación.

- ¿Cuál es intervalo del monto aprobado de mayor probabilidad para una recuperación del crédito menor al 10%? (H_1 , H_2 o H_3)
- ¿Cuál es intervalo del monto aprobado de mayor probabilidad para una recuperación del crédito entre el 10% y el 50%? (H_1 , H_2 o H_3)

- ¿Cuál es intervalo del monto aprobado de mayor probabilidad para una recuperación del crédito mayor al 50%? (H_1 , H_2 o H_3)

Las siguientes preguntas se pueden resolver con la probabilidad condicional $P(G_j|H_i)$ donde $i = 1, 2, 3$ para algún monto aprobado y $j = 1, 2, 3$ para algún porcentaje de recuperación.

- Dado que el monto aprobado fue menor a \$10,000 ¿Cuál es el porcentaje de recuperación de mayor probabilidad? (G_1 , G_2 o G_3)
- Dado que el monto aprobado fue entre \$10,000 a \$20,000 ¿Cuál es el porcentaje de recuperación de mayor probabilidad? (G_1 , G_2 o G_3)
- Dado que el monto aprobado fue mayor a \$20,000 ¿Cuál es el porcentaje de recuperación de mayor probabilidad? (G_1 , G_2 o G_3)

Genera las siguientes gráficas de barras:

1. La probabilidad $P(H_1|G_j)$ para cada G_j con $j = 1, 2, 3$
2. La probabilidad $P(H_2|G_j)$ para cada G_j con $j = 1, 2, 3$
3. La probabilidad $P(H_3|G_j)$ para cada G_j con $j = 1, 2, 3$
4. La probabilidad $P(G_1|H_i)$ para cada H_i con $i = 1, 2, 3$
5. La probabilidad $P(G_2|H_i)$ para cada H_i con $i = 1, 2, 3$
6. La probabilidad $P(G_3|H_i)$ para cada H_i con $i = 1, 2, 3$

Explica cada gráfica y que significa un porcentaje mayor o menor y qué significa ser el máximo y mínimo o que todos sean iguales.