

Tarea 1 - Análisis de datos

Giovanni Gamaliel López Padilla

1. Lee el artículo Data Science vs Statistics (está en Moodle). Piensa en la respuesta que diste en la encuesta; si de nuevo te piden describir Ciencia de Datos en un párrafo ahora después de haber leído el artículo, ¿qué contestarías? (por supuesto no existen repuestas buenas y malas)

La Ciencia de Datos es un conjunto de herramientas de exploración, análisis, modelado y visualización de datos para obtener una interpretación acerca del evento de donde proviene la información. La Ciencia de Datos no puede remplazar a la ciencia, si no que es otra manera de verificar la teoría que se plantea.

2. Lee en las notas (capítulo 1 del pdf en Moodle) lo de los dados no transitivos. Verifica los cálculos.

Las caras de los dados son los siguientes:

- dado A: 4, 4, 4, 4, 0, 0
- dado B: 3, 3, 3, 3, 3, 3
- dado C: 6, 6, 2, 2, 2, 2
- dado D: 5, 5, 5, 1, 1, 1

Para comprobar que $P(A > B) = \frac{2}{3}$, se tiene que las combinaciones donde $A \setminus B$ es 24, esto es porque podemos obtener cuatro veces el número cuatro en el dado A, y en el dado B podemos solamente obtener un 3. El total de combinaciones es 36, por lo tanto:

$$P(A > B) = \frac{24}{36}$$
$$P(A > B) = \frac{2}{3}$$

Comprobando que $P(B > C) = \frac{2}{3}$, se tiene que las combinaciones donde $B \setminus C$ es 24. El número 6 en el dado C se repite dos veces y el 2 se repite cuatro veces, en el dado B contiene unicamente números 3. El total de combinaciones es 36, por lo tanto:

$$P(B > C) = \frac{24}{36}$$
$$P(B > C) = \frac{2}{3}$$

De igual manera, para comprobar $P(C > D) = \frac{2}{3}$, se tiene que el número 6 se repite dos veces en el dado C, como el seis es mayor a cualquier número en el dado D, entonces se tendrían 12 combinaciones donde C es mayor. Por otro lado el número dos se repite cuatro veces y el dado D contiene tres veces el número 1, entonces obtenemos 12 combinaciones donde $C \setminus D$. Dando como resultado que el total de combinaciones donde $C \setminus D$ es 24, por lo tanto:

$$P(C > D) = \frac{24}{36}$$
$$P(C > D) = \frac{2}{3}$$

Comprobando que $P(A > D) = \frac{1}{3}$, el unico caso cuando $A > D$ es cuando obtenemos un 4 en el dado A y un 1 en el dado D, dando asi un total de 12 combinaciones, por lo tanto:

$$P(A > D) = \frac{12}{36}$$

$$P(A > D) = \frac{1}{3}$$

comprobando así, todos los datos que se muestran en el primer capitulo de las notas.

3. **Eliges al azar un número entre 1 y 1200. Calcula la probabilidad de que sea divisible por 4 o por 6.**

La cantidad de números que son divisibles por 4 en el rango 1-1200 es 300. Para el mismo rango, la cantidad de números divisibles por 6 es 200. Existen números que son divisibles por 4 y 6, por ejemplo el número 12, el total de estos números es 100. Condensando esta información obtenemos que:

$$P(A) = \frac{300}{1200} \quad P(B) = \frac{200}{1200}$$

donde $P(A)$ es la probabilidad de que el número elegido sea divisible por 4, $P(B)$ es la probabilidad de que el número elegido sea divisible por 6. Entonces, la probabilidad de obtener un número que sea divisible por 4 y 6 es:

$$P(A \cap B) = \frac{100}{1200}$$

Por lo tanto, la probabilidad de obtener un número que sea divisible por 4 o 6 es:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{300}{1200} + \frac{200}{1200} - \frac{100}{1200} \\ &= \frac{400}{1200} \end{aligned}$$

$$P(A \cup B) = \frac{1}{3}$$

4. **Se van a lanzar dos dados. Se tiene interés en predecir la diferencia entre el valor más grande y el más chico. Si tienes que apostar a un resultado (diferencia) particular ¿cuál tomarías?**

Tomaría el 0, ya que este resultado se obtiene en un total de 6 veces en las combinaciones posibles, que en total son 21. Esto porque podemos obtener 36 combinaciones posibles, pero al solo restar el valor mayor y el chico se reduce este número.

5. **Tomas un mazo de cartas. Solamente te fijas en el valor numérico correspondiente (campesino=11,dama=12, rey=13). Eliges una carta al azar, apuntas su valor numérica y la regresas. Eligas otra carta y también apuntas su valor numérico. ¿Cuál es la probabilidad que el primer valor sea mayor que el segundo?**

El total de permutaciones posibles en este caso son 9, de las cuales 3 se obtiene primero un valor mayor al segundo. Esto es representado en la tabla 1. Por lo tanto, la probabilidad que suceda este evento (A) es

$$P(A) = \frac{1}{3}$$

13 13	12 13	11 13
13 12	12 12	11 12
13 11	12 11	11 11

Tabla 1: Total de opciones posibles del problema 5. Pares de números que cumplen la condición del problema (color rojo).

6. **Elige al azar $b, c \in [0, 1]$. ¿Cuál es la probabilidad que su diferencia es menor que 0.1?**

Definimos a $\Omega = [0, 1]^2$, ya que es el espacio total. Entonces, la manera de calcular la probabilidad que la diferencia sea menor a 0.1 (evento A) es:

$$P(A) = \frac{\int_A dx}{\int_{\Omega} dx}$$

Visualmente este problema es representado en la figura 1.

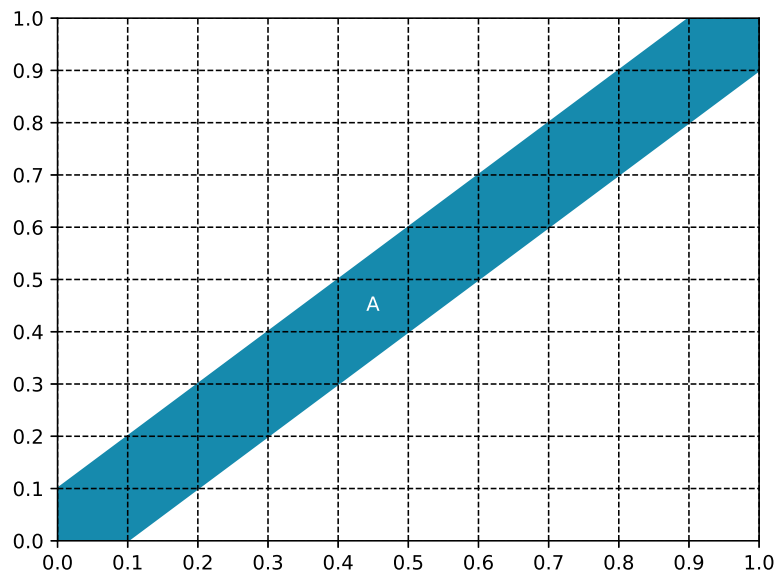


Figura 1: Conjunto de valores validos para el evento A (azul) y el total de posibilidades.

Calculando el área del evento A se obtiene lo siguiente:

$$\begin{aligned}
 \int_A dx &= 2 \left(\frac{1}{2} - \frac{81}{200} \right) \\
 &= 2 \left(\frac{19}{200} \right) \\
 \int_A dx &= \frac{19}{100}
 \end{aligned}$$

Por lo tanto, la probabilidad que suceda algún evento contenido en A es:

$$\begin{aligned} P(A) &= \frac{\int_A dx}{\int_{\Omega} dx} \\ &= \frac{\frac{19}{100}}{1} \\ P(A) &= \frac{19}{100} \end{aligned}$$