

Lista de Ejercicios 1

1.

Supón que X e Y tienen la siguiente función de masa de probabilidad conjunta:

X\Y	0	1	2
1	10%	10%	20%
2	20%	10%	10%
3	10%	10%	0%

- Encuentra las probabilidades marginales de X e Y.
- Calcula $E[X|Y=y]$, para $y=0,1$ y 2.
- Calcula $E[E[X|Y]]$.
- Calcula $E[X]$ directamente y verifica la Ley de la Esperanza Iterada.
- Calcula $E[Y]$ (cómo prefieres).
- Calcula la Covarianza entre X y Y.

2.

Sean $X \sim U[-1, 1]$ y $Y = X^2$.

- Calcule $E[X|Y = 0.25]$
- Describe la función $E[X|Y=y]$
- ¿Son X y Y independientes?
- ¿Es X independiente en media de Y?
- ¿Es Y independiente en media de X?
- Calcule la covarianza entre X y Y

3.

En R, use los siguientes comandos para cargar una base de datos sobre atletas olímpicos:

```
install.packages("tidytuesdayR") # Solo necesario una vez
```

```
tuesdata <- tidytuesdayR::tt_load(2024, week = 32)  
olympics <- tuesdata$olympics
```

Se puede encontrar más información sobre esos datos en este enlace:

<https://github.com/rfordatascience/tidytuesday/blob/main/data/2024-08-06/readme.md>

- a) Nos interesa conocer cómo la altura, el peso, y la edad de los atletas afecta su desempeño deportivo. La medida de desempeño va a ser si el atleta ganó una medalla (cualquier tipo). Estima tres regresiones con cada una de esas características. Después estima una regresión con las tres al mismo tiempo. Presenta una tabla de resultados e interprétalos. (Antes de empezar, te recomiendo que mantenga apenas las observaciones sin valores faltantes en altura, peso y edad. Y tenga cuidado con valores NA cuando crías la variable de desempeño.)
- b) Estima una regresión de desempeño en el log de la altura, el log del peso, edad, sexo, y temporada. Interpreta cada coeficiente. ¿Cómo se explica el signo del coeficiente en temporada?
- c) A partir del modelo de la parte b, nos interesa saber si el impacto de edad es mayor o menor para mujeres. Testa la hipótesis y explica su conclusión.
- d) Con el modelo en b, haga predicciones sobre la probabilidad de ganar una medalla para cada atleta. Compara la probabilidad estimada para los atletas que ganaron contra los que no ganaron. Presenta un gráfico de la relación entre edad y la predicción.
- e) Repita la parte d, utilizando un modelo logit. ¿Te parece que el modelo es mejor que el lineal?

4.

Pregunta: Exposición a Plaguicidas y Salud Mental

Estás colaborando con un equipo de investigación que estudia el impacto potencial de la exposición a plaguicidas sobre la salud mental en poblaciones rurales. El equipo ha recolectado datos observacionales y está considerando cómo modelar las relaciones causales entre las variables mediante el uso de diagramas causales.

La pregunta de investigación principal es: **¿Cuál es el efecto de la exposición a plaguicidas sobre la salud mental?**

Descripción de variables:

- **ExposiciónPlaguicidas:** Si la persona trabaja en un entorno con alta exposición a plaguicidas.
- **SaludMental:** Diagnóstico de depresión, ansiedad u otros trastornos del estado de ánimo.

- **Enfermedad:** Síntomas físicos o episodios de enfermedad posiblemente relacionados con la exposición.
- **Ocupación:** Tipo de empleo, especialmente si involucra trabajo agrícola.
- **Ingreso:** Nivel de ingresos mensuales.
- **ChequeoMédico:** Si la persona ha visitado a un médico recientemente para una evaluación general.

El equipo propone las siguientes relaciones entre variables:

- **SaludMental** es influida por **ExposiciónPlaguicidas**, **Enfermedad** e **Ingreso**.
- **ExposiciónPlaguicidas** depende de la **Ocupación** (por ejemplo, trabajo agrícola vs. no agrícola).
- **Enfermedad** es causada por la **ExposiciónPlaguicidas**.
- **Ingreso** depende tanto de la **Ocupación** como de la **Enfermedad**.
- **ChequeoMédico** representa si una persona ha tenido un chequeo médico reciente. Depende de la **SaludMental** y de la **Enfermedad**.

Preguntas

a) Dibuje el grafo dirigido acíclico (DAG) que representa las relaciones causales propuestas.

b) Enumere todos los caminos distintos entre **ExposiciónPlaguicidas** y **SaludMental**.

(i) Clasifique cada uno como un camino de **front door** o **backdoor**.

(ii) Indique si el camino está **abierto** o **cerrado**, asumiendo que no se condiciona en ninguna variable.

c) ¿Qué variables se deben controlar para estimar el **efecto causal total** de la exposición a plaguicidas sobre la salud mental? Justifique su respuesta.

d) ¿Qué variables se deben controlar para estimar el **efecto directo** de la exposición a plaguicidas sobre la salud mental?

e) Suponga que el conjunto de datos solo incluye personas que se han hecho un **chequeo médico reciente**.

(a) ¿Qué tipo de sesgo podría introducirse?

(b) ¿Qué caminos se ven afectados por este condicionamiento?

f) Una colega sugiere que la **Ocupación** podría tener un efecto **directo** sobre la salud mental (por ejemplo, debido al estrés laboral o las condiciones del trabajo).

¿Cómo podrías evaluar esta hipótesis utilizando las asociaciones observadas en los datos, suponiendo que el resto del DAG es correcto?