



Facultad de Ingenierías
Inferencia Estadística
Maestría en Inteligencia Artificial y Ciencia de Datos
Prof. Cristian E Garcia.



Desafío 3

Condiciones:

- Subir la tarea en formato pdf en la plataforma UAO-Virtual.
- Es necesario incluir el código de R en formato R. Mostrar los resultados a partir de tablas, gráficos o indicadores que les permita dar respuesta a los planteamientos.
- Deben interpretar los resultados obtenidos en cada situación de acuerdo al contexto.
- Realizar la actividad en grupos máximo de 4 personas.

Situación 1

La viña Chateau Latour es ampliamente reconocida como una de las mejores en el mundo, y posee una rica historia que nace en el año 1638. La calidad de la cosecha tiene un gran impacto en el valor de comercialización del vino, pudiendo llegar a costar miles de dólares. Muchos han investigado el efecto del momento en que se realiza la cosecha sobre la calidad del vino, pero un aspecto menos explorado es el efecto que tienen las lluvias durante el tiempo de la cosecha sobre la misma. Este problema se refiere al último punto. Los datos se encuentran en el archivo **Latour.txt** y corresponden a:

Variable	Descripción
cosecha	: año de la cosecha
calidad	: puntaje asignado a la calidad de la cosecha (1: peor, 5: mejor)
dias	: número de días, a partir del 31 de agosto, en que finalizó la cosecha
lluvia	: 1: lluvia durante la cosecha 0: sin lluvia durante la cosecha

- (a) Obtenga un gráfico de la calidad de la cosecha versus el momento en que ella finalizó, diferenciando con colores o figuras las cosechas en que se produjo algún tipo de lluvia de aquellas en las que no. Comente sobre el efecto del momento de cosecha sobre su calidad, comparando ambos grupos. No olvide etiquetar correctamente la figura.
- (b) Se propone ajustar un modelo que permita tanto diferentes pendientes como interceptos a ambos grupos. Obtenga un intervalo de 95% de confianza para el número de días después del 31 de agosto en que debe finalizar una cosecha para que el valor medio estimado de su calidad disminuya en una unidad, cuando se observa lluvia durante la cosecha. Interpretelo. **Hint:** Note que se le pide inferencia sobre un parámetro de la forma $\frac{1}{x_0^t \beta}$. Puede obtenerlo a partir de un intervalo para $x_0^t \beta$.

Situación 2

Sea

$$Y_1 = \theta + \epsilon_1$$

$$Y_2 = 2\theta - \varphi + \epsilon_2$$

$$Y_3 = \theta + 2\varphi + \epsilon_3$$

donde $E[\epsilon_i] = 0$ para $i = 1, 2, 3$. Encuentre las estimaciones de los parámetros θ y φ por mínimos cuadrados.

Situación 3

Los datos que siguen corresponden al registro de nuevos casos de melanoma informados en Estados Unidos durante 1969–1991 en hombres blancos clasificados por edad y región. La última columna es el tamaño de la población registrada en el censo de EE.UU.

Región	Edad	Casos	Población
NORTE	0–35	61	2880262
	35–44	76	564535
	45–54	98	592983
	55–64	104	450740
	65–74	63	270908
	75+	80	161850
SUR	0–35	64	1074246
	35–44	75	220407
	45–54	68	198119
	55–64	63	134084
	65–74	45	70708
	75+	27	34233

- Grafique el log de la tasa observada (número de casos sobre tamaño de la población) versus la edad. ¿Qué observa en este gráfico?
- Ajuste un modelo de Poisson incorporando un intercepto, una variable dummy para la región, variables dummies para distinguir los grupos de edad (tome como referencia el grupo de menor edad) y el **offset** que le parezca adecuado. ¿El ajuste parece adecuado?
- Verifique si la función de varianza es adecuada. ¿Qué puede estar ocurriendo? ¿Le parece que el problema puede ser solucionado utilizando un parámetro de escala o cambiando a un modelo binomial-negativo?

Situación 4

En un estudio se desean investigar los factores que influyen la elección de alimento primario de los caimanes. El estudio se basa en la información obtenida de 219 caimanes capturados en cuatro lagos de Florida. La variable respuesta de escala nominal es el tipo primario de alimento, en volumen, encontrado al interior del estómago de un caimán. En la tabla a continuación se presenta la clasificación de la elección de alimento primario según lago y tamaño del caimán.

Lago	Tamaño (m)	Peces	Invertebrados	Reptiles	Aves	Otros
Hancock	≤ 2.3	23	4	2	2	8
Hancock	> 2.3	7	0	1	3	5
Ocklawaha	≤ 2.3	5	11	1	0	3
Ocklawaha	> 2.3	13	8	6	1	0
Trafford	≤ 2.3	5	11	2	1	5
Trafford	> 2.3	8	7	6	3	5
George	≤ 2.3	16	19	1	2	3
George	> 2.3	17	1	0	1	3

- Fije la categoría de referencia en: Peces. Ajuste el modelo.
- Calcule las razones de chances para los lagos y el tamaño.
- Presente de manera ordenada dichas razones en una tabla.
- ¿Tiene algún efecto el tamaño del caimán sobre la elección del alimento primario? Justifique e interprete los posibles efectos en términos de razones de chances. Interprete el efecto del tamaño del caimán sobre la elección primaria de alimentos invertebrados.

- (e) ¿Tiene algún efecto el lago sobre la elección del alimento primario? Justifique e interprete los posibles efectos en términos de razones de chances.

Situación 5

Sea el modelo de regresión lineal simple:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (i = 1, 2, \dots, n),$$

donde $E[\varepsilon] = 0$ y $\text{Var}[\varepsilon] = \sigma^2 I_n$. Encuentre los estimadores por mínimos cuadrados de β_0 y β_1 . Demuestre que estos estimadores son no correlacionados si y solo si $\bar{x} = 0$.

Situación 6

Considere la siguiente tabla de $3 \times 4 \times 4$ que resume datos de severidad de los efectos colaterales de una operación, S (1=ninguno, 2=leve, y 3=moderada), tipo de operación, O , y hospital, H . Las cuatro operaciones, que corresponden a tratamientos para úlcera duodenal, son numeradas de acuerdo a su grado de invasividad.

Operación	Severidad											
	Hospital 1			Hospital 2			Hospital 3			Hospital 4		
	1	2	3	1	2	3	1	2	3	1	2	3
1	23	7	2	18	6	1	8	6	3	12	9	1
2	23	10	5	18	6	2	12	4	4	15	3	2
3	20	13	5	13	13	2	11	6	2	14	8	3
4	24	10	6	9	15	2	7	7	4	13	6	4

- (a) Proponga un modelo apropiado para relacionar la severidad de los efectos colaterales y el tipo de operación.
- (b) De acuerdo al modelo previamente ajustado, encuentre una expresión para $P(S = 1 | O)$, $P(S = 2 | O)$ y $P(S = 3 | O)$ para el tipo de operación 1.
- (c) Con una significancia del 5%, ¿tiene algunas de las operaciones 1, 2, o 3 un efecto significativo en comparación con la operación más invasiva, 4? Explique.