

## MED.EX.20250202

### MED.EX.20250202.1

#### Solución MED.EX.20250202.1

La proposición es **falsa**.

Obsérvese que

$$X_{p+1} = \frac{1}{p} \sum_{j=1}^p X_j.$$

Utilizando la linealidad de la esperanza,

$$\mathbb{E}[X_i X_{p+1}] = \frac{1}{p} \mathbb{E} \left[ X_i \sum_{j=1}^p X_j \right] = \frac{1}{p} \sum_{j=1}^p \mathbb{E}[X_i X_j].$$

Dado que las variables  $X_1, \dots, X_p$  son independientes e idénticamente distribuidas con media 0 y varianza 1, se cumple:

- $\mathbb{E}[X_i X_j] = 0$  si  $i \neq j$ ,
- $\mathbb{E}[X_i^2] = 1$  si  $i = j$ .

Por tanto, en la suma anterior solo contribuye el término correspondiente a  $j = i$ , y se obtiene

$$\mathbb{E}[X_i X_{p+1}] = \frac{1}{p}.$$

Sin embargo,

$$\mathbb{E}[X_i] \mathbb{E}[X_{p+1}] = 0 \cdot 0 = 0,$$

ya que todas las variables tienen esperanza nula.

Como ambos valores no coinciden, la igualdad propuesta es incorrecta.

**Respuesta:** b) Falso.

### MED.EX.20250202.2

#### Solución MED.EX.20250202.2

El script trabaja con el conjunto `seatpos` y ajusta una **regresión lineal múltiple** para explicar `hipcenter` en función del resto de variables. En primer lugar se obtiene el vector de coeficientes del modelo ajustado con

$$\text{hipcenter} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Weight} + \dots + \beta_8 \text{Leg} + \varepsilon.$$

Después se calcula el **VIF** del mismo modelo. Los VIF obtenidos muestran valores muy elevados, en particular para `HtShoes` y `Ht` (del orden de varios

cientos). Esto indica que existe **multicolinealidad severa**: ciertas variables explicativas (especialmente las relacionadas con la altura) están fuertemente relacionadas entre sí, por lo que el modelo tiene dificultad para separar el efecto individual de cada una.

A continuación se genera una variable de ruido **ruido** (normal estándar) y se reajusta el modelo usando como respuesta **hipcenter + ruido**. El objetivo de este paso es comprobar la estabilidad de los coeficientes frente a una perturbación pequeña en la variable respuesta.

Al comparar los coeficientes de ambos ajustes, se observa que algunos cambian de manera apreciable (por ejemplo los asociados a **Ht** y **HtShoes**), pese a que la modificación en la respuesta es solo un ruido aleatorio de media cero. Este comportamiento es característico de la multicolinealidad: **los estimadores de los coeficientes se vuelven inestables** y pueden variar sensiblemente ante cambios pequeños en los datos, lo que dificulta su interpretación.

En resumen, el código pone de manifiesto (i) la presencia de multicolinealidad a través de VIF muy altos y (ii) la inestabilidad de los coeficientes cuando la multicolinealidad es importante.

---

## MED.EX.20250202.3

### Solución MED.EX.20250202.3

La valoración del diálogo con ChatGPT se realiza mediante una rúbrica compuesta por **cinco criterios**, cada uno con un máximo de **2 puntos**, lo que permite obtener una calificación total entre 0 y 10. En este caso, la nota asignada es **7/10**, justificada como sigue:

- **Concepto de multicolinealidad (1/2).**  
ChatGPT describe la multicolinealidad como una alta correlación entre variables explicativas, lo cual es una aproximación válida pero algo simplificada. No se enfatiza suficientemente que el problema surge por dependencias lineales (o casi lineales) entre predictores, más allá de la simple correlación par a par.
- **Métodos de detección (2/2).**  
Se mencionan herramientas estándar y adecuadas, como la matriz de correlación y el **Factor de Inflación de la Varianza (VIF)**, junto con criterios numéricos habituales. Este apartado está correctamente desarrollado.
- **Consecuencias sobre el modelo (1/2).**  
Se alude a problemas relevantes, como la inestabilidad de los coeficientes y el aumento de los errores estándar. No obstante, falta una explicación más clara de por qué la multicolinealidad dificulta la inferencia individual sobre los parámetros, por lo que la puntuación no es completa.

- **Estrategias de tratamiento (2/2).**

Se proponen soluciones variadas y apropiadas: eliminación o combinación de variables, selección automática, regularización (LASSO) y técnicas de reducción de dimensión como PCA. Este criterio está bien cubierto.

- **Ejemplo ilustrativo (1/2).**

Se incluye un ejemplo conceptual sencillo que ayuda a entender el problema, aunque no está específicamente contextualizado en un modelo de regresión logística multinomial ni se apoya en un caso numérico concreto.

Sumando las puntuaciones parciales se obtiene la nota final:

$$1 + 2 + 1 + 2 + 1 = 7/10.$$

En conclusión, la respuesta de ChatGPT es adecuada y demuestra conocimiento general del problema y de sus posibles soluciones, aunque podría mejorar en la precisión conceptual y en la aportación de ejemplos más específicos.

## **MED.EX.20250202.4**

### **Solución MED.EX.20250202.4**

En el código se ajusta un **modelo de regresión logística binomial** con respuesta **Survived** y una única variable explicativa **Sex**, utilizando datos **agrupados** mediante la opción **weights = frecuencia**.

El tamaño muestral total es

$$n = 233 + 109 + 81 + 468 = 891.$$

La **null deviance** corresponde al modelo nulo, que solo incluye un intercepto y supone una única probabilidad de supervivencia para todos los individuos. En este caso su valor es

$$\text{Null deviance} = 1186.66.$$

Este valor mide el desacuerdo entre los datos observados y un modelo que ignora completamente la variable **Sex**.

La **residual deviance** corresponde al modelo que incluye la variable **Sex** como explicativa. Su valor es

$$\text{Residual deviance} = 917.80.$$

### **Interpretación:**

La diferencia entre ambas deviances es

$$1186.66 - 917.80 = 268.86,$$

que coincide con el estadístico de razón de verosimilitudes, asintóticamente distribuido como una  $\chi^2$  con 1 grado de libertad. Este valor tan elevado indica que la inclusión de **Sex** mejora de forma muy significativa el ajuste del modelo.

En consecuencia, se concluye que la variable **Sex** tiene un efecto claro sobre la probabilidad de supervivencia: el modelo con **Sex** explica mucho mejor los datos que el modelo nulo sin covariables.

---

## MED.EX.20250202.5

### Solución MED.EX.20250202.5

El planteamiento involucra dos variables binarias:

- Variable explicativa (exposición): horario de la ducha (mañana vs noche), que podemos codificar como  $X \in \{0, 1\}$ .
- Variable respuesta: mejora del bienestar (Sí/No), que podemos codificar como  $Y \in \{0, 1\}$ .

Por tanto, el esquema es  $D \leftarrow D$  y el análisis final consistiría en comparar las proporciones de mejora entre grupos, por ejemplo contrastando

$$P(Y = 1 \mid X = 1) \text{ frente a } P(Y = 1 \mid X = 0),$$

y resumiendo con diferencia de proporciones, riesgo relativo u *odds ratio*.

Diseños posibles:

- **Ensayo clínico aleatorizado (prospectivo)**: asignar al azar a los participantes a ducharse por la mañana o por la noche durante un periodo definido, medir el bienestar y registrar si mejora. Es el diseño con mayor evidencia causal, aunque es difícil el cegamiento y puede haber incumplimiento del protocolo. Debe definirse previamente qué se considera “mejora” (cuestionario validado, umbral, etc.).
- **Diseño cruzado (cross-over)**: cada sujeto prueba ambas condiciones en periodos distintos (por ejemplo, una semana mañana y otra semana noche), asignando al azar el orden. Este diseño reduce la variabilidad interindividual, pero obliga a considerar efecto periodo y posible arrastre (carryover). Puede requerirse un periodo de lavado si se sospecha persistencia del efecto.
- **Estudio observacional**: comparar personas que ya se duchan habitualmente por la mañana con las que lo hacen por la noche. Es más sencillo, pero la interpretación causal es limitada por la posible confusión (rutinas

de sueño, trabajo, ejercicio, estrés, etc.), por lo que conviene medir estas variables para controlar/ajustar en el análisis.

En todos los casos es importante contemplar factores adicionales relevantes (hábitos de sueño, actividad física, estrés, consumo de cafeína/alcohol, horario laboral, etc.) y valorar la viabilidad y aspectos éticos. En conclusión, si se busca evidencia causal, es preferible un **diseño aleatorizado**, idealmente **cruzado** si se espera mucha heterogeneidad entre individuos.

## MED.EX.20250202.6

### Solución MED.EX.20250202.6

El código ajusta una **regresión lineal simple en un enfoque bayesiano** usando el paquete **MCMCpack**.

Primero se limpia el entorno (`rm(list = ls())`) y se cargan los datos, con variable respuesta  $y$  y variable explicativa  $x$ . A continuación se ajusta el modelo:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

La función **MCMCregress** no estima los parámetros por mínimos cuadrados, sino que realiza inferencia bayesiana, tratando  $\beta_0, \beta_1$  y  $\sigma^2$  como variables aleatorias con **distribuciones a priori**. En el código se especifican hiperparámetros de dichas priors:

- **b0** y **B0** controlan la distribución a priori de  $\beta$  (vector de coeficientes).
- **c0** y **d0** controlan la distribución a priori de la varianza (o precisión) del error.

La estimación de la distribución **a posteriori** se realiza mediante **Markov Chain Monte Carlo (MCMC)**: se simulan valores de la posteriori generando una cadena de Markov cuya distribución estacionaria es la posteriori del modelo.

Finalmente, `summary(modelo1)` devuelve un resumen de los resultados MCMC, proporcionando estimaciones posteriores (por ejemplo medias posteriores), desviaciones típicas y cuantiles/intervalos creíbles para  $\beta_0, \beta_1$  y el parámetro de dispersión, lo que permite interpretar el efecto de  $x$  sobre  $y$  en un marco bayesiano.