

Clase auxiliar # 4

Variables binarias, categóricas e interacciones

Una amiga médica de usted le pide ayuda para estudiar los efectos del consumo de tabaco –representado por las variables *dummies* de fumar $F_n \in \{0,1\}$ y no fumar $NF_n \in \{0,1\}$ respectivamente– y las horas de ejercicio $EJ_n \in \mathbb{R}^+$ sobre la capacidad aeróbica $CA_n \in \mathbb{R}^+$ en humanos. También se tiene información sobre la edad $ED_n \in \mathbb{R}^+$ de las personas estudiadas. Su amiga se ha tomado el tiempo de crear una variable categórica de rango etario $RE_n \in \{1, 2, 3\}$ que toma los valores 1, 2 y 3 dependiendo de si el individuo n es adolescente ($ED_n \in [14-18)$), adulto ($ED_n \in [18-60)$) o adulto mayor ($ED_n \in [60, \infty)$) respectivamente. Como ella sabe que usted tiene un excelente manejo de estadísticas le dice lo siguiente:

1. “En promedio quienes no fuman tienen mejor condición aeróbica que quienes sí”. Luego le pide:
 - a. Hacer un bosquejo de cómo se distribuye la muestra en cuanto a condición aeróbica, separando a fumadores y no fumadores. Mostrar dónde está el promedio muestral.
 - b. Plantear dos ejemplos modelos de regresión lineal **sin colinealidad perfecta** y un contraejemplo **con colinealidad perfecta** diseñados para estimar estos promedios. Explique por qué (no) hay colinealidad perfecta en cada caso y por qué este es un problema en el contraejemplo.
 - c. Para los tres modelos y basándose en el gráfico que usted mismo(a) bosquejó, haga lo siguiente (siempre que se pueda):
 - i. Calcule $\mathbb{E}[CA_n|F_n = 0]$, $\mathbb{E}[CA_n|F_n = 1]$.
 - ii. Calcule la diferencia esperada en condición aeróbica entre un fumador y un no fumador.
 - iii. En cuanto a los parámetros poblacionales de su modelo, especule sus signos (+/-), interprete su significado y muéstrellos gráficamente.

2. “Me interesa estimar la condición aeróbica promedio de fumadores y no fumadores de todos los rangos etarios. Algunos médicos afirman que la diferencia promedio entre fumadores y no fumadores es aproximadamente la misma para todos los grupos etarios. De la misma forma, las diferencias promedio entre rangos etarios es aproximadamente la misma para fumadores y no fumadores. Podríamos asumir estas diferencias como constantes”.
 - a. Hacer un bosquejo de cómo se distribuye la muestra en cuanto a condición aeróbica, clasificando por rango etáreo (RE_n) y consumo de tabaco (F_n) a la vez. Mostrar dónde está el promedio muestral.
 - b. Plantear dos ejemplos modelos de regresión lineal **sin colinealidad perfecta**.
 - c. Para los tres modelos y basándose en el gráfico que usted mismo(a) bosquejó, haga lo siguiente (siempre que se pueda):
 - i. Calcule $\mathbb{E}[CA_n|F_n, RE_n]$ para todas las combinaciones posibles.
 - ii. Calcule la diferencia esperada en condición aeróbica entre un fumador y un no fumador para todos los rangos etarios.
 - iii. Calcule la diferencia esperada en condición aeróbica entre fumadores de distintos rangos etarios y no fumadores de distintos rangos etarios.

- iv. En cuanto a los parámetros poblacionales de su modelo, especule sus signos (+/-), interprete su significado y muéstrellos gráficamente.
3. “Creo que lo que te dije la vez pasada es incorrecto. Sería interesante estimar el promedio de los seis grupos (adolescentes fumadores, adolescentes no fumadores, adultos fumadores, etc.) sin asumir esas diferencias constantes”. Repita el ejercicio de la parte anterior.
4. “Algunos médicos afirman que las horas de ejercicio aeróbico semanal son una variable importante al momento de explicar la capacidad aeróbica de las personas. Además su efecto es el mismo, tanto para fumadores como para no fumadores”. Luego le pide:
- Hacer un bosquejo de cómo se distribuye la muestra en cuanto a condición aeróbica y horas de ejercicio aeróbico semanal, separando a fumadores y no fumadores. En la misma figura grafique un modelo que cumpla con las características descritas por su amiga.
 - Plantear dos ejemplos modelos de regresión lineal **sin colinealidad perfecta** y un contraejemplo **con colinealidad perfecta** diseñados para estimar estos promedios. Explique por qué (no) hay colinealidad perfecta en cada caso y por qué este es un problema en el contraejemplo.
 - Para los dos ejemplos y basándose en el gráfico que usted mismo(a) bosquejó, haga lo siguiente (siempre que se pueda):
 - Calcule $\mathbb{E}[CA_n|F_n = 0]$, $\mathbb{E}[CA_n|F_n = 1]$.
 - Calcule la diferencia esperada en condición aeróbica entre un fumador y un no fumador que hacen la misma cantidad de ejercicio. Obtenga el efecto marginal del ejercicio aeróbico sobre la capacidad aeróbica tanto para fumadores como no fumadores. ¿Son iguales?
 - En cuanto a los parámetros poblacionales de su modelo, especule sus signos (+/-), interprete su significado y muéstrellos gráficamente.
5. “Hay otros médicos que afirman que los efectos de hacer ejercicio aeróbico son más fuertes en fumadores”. Repita el ejercicio de la parte anterior.
6. “Leí en un *paper* que el efecto marginal de hacer ejercicio sobre la condición aeróbica disminuye con la edad”. Proponga un modelo de regresión lineal con interacción de variables continuas que permita estimar dicho efecto. Haga un bosquejo de cómo se comportaría en valor esperado la capacidad aeróbica en función de la edad y la cantidad de ejercicio aeróbico semanal suponiendo que lo que dice su amiga es cierto.