

Proyecto final para el curso Ciencia de datos

*¿La consideración de las consecuencias futuras predice el
índice de masa corporal?*

Lucía Álvarez 4.822.283-2

Mentora: Laia Subirats

30/09/2022

1. Definición de las variables

El siguiente trabajo es parte del curso *Data Science* realizado en Barcelona Activa en el año 2022. En este estudio se trabaja con el constructo Consideración de las Consecuencias Futuras (CFC) con la intención de explorar cómo la misma se relaciona con el Índice de Masa Corporal (IMC).

1.1. Consideración de las Consecuencias Futuras

La CFC es un rasgo de personalidad con el cual se puede conocer el grado que las personas consideran y son influenciadas por los resultados inmediatos y distantes de sus acciones (Strathman et al., 1994).

La CFC es evaluada mediante una escala likert de 14 ítems. La escala fue propuesta con una estructura unifactorial (Strathman et al., 1994). Estudios más recientes han propuesto una organización en dos factores correlacionados: Consideración de las Consecuencias Inmediatas (CFC_I) y Consideración de las Consecuencias Futuras (CFC_F) (Joireman et al., 2012). Las personas con elevada CFC_I tienden a realizar acciones que son gratificantes al instante, pero con resultados distantes negativos. Por ejemplo, el consumo de sustancias activas genera un placer hedonista, pero a la misma vez dificulta una adecuada salud futura. Mientras que las personas con elevada CFC_F tienden a realizar acciones con una recompensa futura, realizando ciertos sacrificios su presente. Por ejemplo, un estudiante universitario con elevada CFC_F tenderá a no salir los fines de semana (recompensa instantánea) con el fin de salvar un examen a futuro.

1.1. Consideración de las Consecuencias Futuras y su relación con variables de salud

La mayoría de las acciones saludables requieren de sacrificios prolongados en el tiempo para obtener una recompensa en un futuro. Por ejemplo, salir a correr todos los días produce un beneficio a largo plazo, con un sacrificio inmediato. Por lo cual, se ha evidenciado que las personas con mayor orientación hacia los resultados distantes (alta CFC_F) tienden a reportar mayor realización de acciones saludables, en comparación con las personas con mayor orientación hacia los resultados inmediatos (alta CFC_I). Por ejemplo, CFC-I predice positivamente el índice de masa corporal (Adams, 2012) y negativamente las intenciones de comer saludablemente (Gick, 2014), mientras que CFC-F predice positivamente la realización de actividad física (Joireman et al., 2012; van Beek et al., 2013).

Objetivos:

En este estudio, se pretende evaluar si la CFC predice el IMC. Es importante tener en cuenta que un elevado IMC es considerado un factor de riesgo para las enfermedades crónicas (Camacho-Camargo et al., 2009). Por lo cual, se espera que el IMC se asocie positivamente con la CFC_I y negativamente con la CFC_F.

2. Presentación del conjunto de datos escogido

La muestra está compuesta por 282 sujetos, de los cuales 187 son mujeres y 95 hombres. Del total de la muestra, 32 no completaron el ciclo secundario (liceo), 47 completaron el ciclo secundario, 138 son estudiantes universitarios, 50 tienen estudios universitarios terminados, 4 estudios de maestrías, 11 estudios de doctorado.

El dataframe contiene: (a) los catorce ítems de la escala CFC, (b) la edad, sexo y educación (no completo el ciclo secundario, secundario completo, estudiante universitario, título universitario, título de maestría, Doctorado), (c) frecuencia de consumo de frutas, verduras y bebidas azucaradas (nunca, menos de una vez por semana, una vez por día, dos o tres veces por día), (d) consumo de tabaco (fuma, no fuma), (e) la altura y peso de cada participante, los cuales serán utilizados para calcular el IMC.

La recolección de datos se realizó en Uruguay, mediante un estudio descriptivo y no experimental. Los datos de este estudio se recolectaron para realizar mi tesis de maestría en la Facultad de Psicología de la Universidad de la República.

3. Características generales.

Este trabajo está compuesto por tres partes:

1) Descripción de los datos.

2) Descripción de los ítems de la escala CFC. Posteriormente se realiza un análisis de factores, con el fin de determinar los factores latentes del instrumento. Previo al análisis factorial se evalúa la adecuación de los datos mediante la prueba de Bartlett y Kaiser-Meyer-Olkin (KMO). Para seleccionar la cantidad de factores se evalúa los autovalores, varianza explicada por cada factor y las saturaciones de cada ítem (saturaciones mayores a .30 son consideradas adecuadas). A su vez, se tendrá en cuenta la teoría referida al constructo CFC (es esperado encontrar dos factores latentes, uno referido a los resultados inmediatos y otro referido a las consecuencias futuras).

3) Testeo de diferentes modelos de predicción, con el fin de evaluar si la CFC predice el IMC. Para esto, se testea un análisis de regresión simple con CFC_I como predictor, luego un modelo con CFC_F. Como tercer modelo, se realiza un análisis de regresión múltiple con los factores latentes de la CFC edad, educación e ingreso mensual promedio de los participantes, variables sociodemográficas más usadas en los estudios previos (Adams, 2012; Murphy & Dockray, 2018). Por último, se testea un árbol de decisión con las ocho variables que presentan mayor correlación con el IMC, y un segundo árbol de decisión con las mismas variables que en el modelo de regresión múltiple.

En este estudio, la depuración de datos se realiza mediante el cálculo de puntuaciones Z y rango intercuartílico. A su vez, se trabaja con los datos estandarizados, con la excepción del árbol de decisión.

4. Presentación de los objetivos:

Este estudio se centra en la aplicación de cinco análisis:

- Análisis factorial: con el fin de identificar los factores latentes relacionados a la escala CFC, se seleccionó este tipo de análisis dado que se asume que hay factores latentes subyacentes a los datos.
- Correlaciones: para evaluar la fuerza de las asociaciones de las variables. En este estudio, no se realiza un análisis de regresión simple con la CFC-F como predictora del IMC dado que la correlación entre ambas variables no es estadísticamente significativa ($p > .05$).
- Análisis de regresión simple: Con el fin de evaluar si la CFC-I y CFC-F predicen el IMC.
- Análisis de regresión múltiple: Para evaluar la influencia de los dos factores de la CFC (subescala inmediata y subescala futura) y variables sociodemográficas sobre el IMC.
- Árboles de decisión (regresión): técnica de aprendizaje supervisado aplicada para crear un modelo explicativo y predictivo del IMC, considerando la CFC, variables sociodemográficas, consumo de alimentos saludables y sustancias psicoactivas.

El principal objetivo de este estudio es evaluar si la CFC predice el IMC. Se espera que CFC_I se asocie negativamente con el IMC y que la CFC_F predice positivamente el IMC (Joireman et al., 2006; Kooij et al., 2018; Murphy & Dockray, 2018).

5. Resultados

5.1. Análisis factorial

Al estimar los coeficientes inter-ítems de la escala CFC se observan correlaciones positivas entre los ítems referidos a los resultados inmediatos, mismo patrón entre los ítems referidos a los resultados

futuros. A su vez, los ítems inmediatos presentan una correlación negativa con los ítems futuros. La adecuación de la matriz fue confirmada con valores de la esfericidad de Bartlett de 841.48 con $p < .01$ y un Kaiser-Meyer-Olkin de 0.75. Los valores mencionados previamente indican que se puede continuar con el análisis de la estructura factorial.

En el gráfico de sedimentación, podemos ver que los autovalores descienden por debajo de 1 desde el tercer factor. Por lo cual, se testea un modelo de tres factores y un segundo modelo de dos factores. El primer modelo testeado logra explicar el 35% de la varianza total. Mientras que el segundo modelo logra explicar el 30% de la varianza total. El tercer factor del segundo modelo logra aumentar únicamente un 5% de la varianza total explicada.

En el primer modelo, el tercer factor presenta cargas mayores a .30 únicamente en dos ítems: ítem 7 con una saturación de .54 e ítem 9 con una saturación de -.32. Estos ítems corresponden a la consideración de la consecuencia futura e inmediata respectivamente. En el segundo modelo, en el primer factor cargan todos los ítems referidos a los resultados inmediatos, con cargas positivas y mayores a .30. Mientras que en el segundo factor cargan todos los ítems referidos a los resultados distantes, con cargas positivas y mayores a .30. Los resultados del modelo con dos factores es el esperado según la literatura previa (Joireman et al., 2012). Considerando las cargas factoriales, varianza explicada y la literatura previa, se continuará con el modelo de dos factores.

5.2. Correlaciones entre los factores latentes y variables de interés

La subescala inmediata reporta correlaciones positivas con la Edad e IMC, y negativas con CFC_F, educación e ingreso mensual. Mientras que CFC_F presenta una asociación positiva con la edad, educación e IMC, y negativa con ingreso mensual. Las asociaciones de CFC_I con el resto de las variables son las esperadas. Dado que las personas con mayor tendencia a las consecuencias inmediatas de sus acciones tienden a tener menor orientación hacia los resultados distantes, educación e ingreso mensual, pero mayor IMC y edad. Las asociaciones entre CFC_F no son las esperadas, dado que se esperaría que esta subescala reportase una asociación negativa con el IMC y positiva con el ingreso mensual.

El IMC presenta una asociación positiva y pequeña con la CFC_I ($r = .17$ $p < .05$). No se observa una asociación significativa entre IMC y CFC_F.

5.3. Regresiones simples

En el análisis de regresión simple, se observa que por cada unidad de cambio en CFC_I, el cambio en el IMC es de alrededor de 0.05 %. El análisis de regresión simple con la CFC no se realizó, dado que no se observó una correlación estadísticamente significativa entre CFC_F e IMC.

5.4. Regresión múltiple

En el modelo de regresión múltiple, se seleccionó CFC_I, CFC_F, Edad e Ingreso mensual, dado que son las únicas variables que reportan una relación significativa. El modelo tiene un R^2 de .87, esto indica que es capaz de explicar el 87% de la variabilidad observada en el IMC. Al interpretar los coeficientes del modelo se observa que cada aumento de una unidad en: (a) CFC_I se asocia con un aumento de .24 en el IMC; (b) CFC_F se asocia con un aumento de .31 en el IMC; (c) Edad se asocia con un aumento de .29 en el IMC; (d) Ingreso mensual hay un aumento de 0.11 en el IMC. No se esperaba una predicción positiva entre CFC_F e IMC, dado que esto indica que las personas con mayor preocupación por los resultados distantes tienden a reportar elevado IMC.

5.5. Árbol de decisión

A continuación, se presenta la interpretación de los dos árboles de decisión realizados.

En el primer árbol de decisión se observa que:	En el segundo árbol de decisión se observa que:
<ul style="list-style-type: none"> • Si la edad es \leq a 22.5 años: <ul style="list-style-type: none"> - y su educación es \leq a 2.5, predice 23.83 (N = 6) - y su educación es $>$ a 2.5, predice 21.41 (N = 42) • Si la edad es $>$ a 22.5 <ul style="list-style-type: none"> - y su CFC-F es \leq a 6.4, se predice 23.01 (N = 68) - y su CFC-F es $>$ a 6.4, se predice 27.8 (N = 3) • Si su edad es $>$ a 32.5 y su ingreso es \leq a 2.5: <ul style="list-style-type: none"> - si su consumo de verduras es \leq a 2.5, se predice 35.7 (N = 3) - si su consumo de verduras es $>$ a 2.5, se predice 28.4 (N = 11) • Si su edad es $>$ a 32.5, su ingreso es $>$ a 2.5: <ul style="list-style-type: none"> - si su edad es \leq a 36.5 se predice 28.7 (N = 12) - si su edad es $>$ a 36.5 se predice 25.4 (N = 50) 	<ul style="list-style-type: none"> • Si la edad es \leq a 22.5 años: <ul style="list-style-type: none"> - y su CFC_F es \leq a 5.1, predice 21.7 (N = 31) - y su CFC_F es $>$ a 5.1, predice 22.53 (N = 17) • Si la edad es $>$ a 22.5 <ul style="list-style-type: none"> - y su CFC-F es \leq a 6.4, se predice 23.01 (N = 68) - y su CFC-F es $>$ a 6.4, se predice 24.7 (N = 91) • Si su edad es $>$ a 32.5 y su ingreso es \leq a 2.5: <ul style="list-style-type: none"> - si su CFC_I es \leq a 3.6, se predice 37.7 (N = 8) - si su CFC_I es $>$ a 3.6, se predice 28.0 (N = 6) • Si su edad es $>$ a 32.5, su ingreso es $>$ a 2.5: <ul style="list-style-type: none"> - si su edad es \leq a 36.5 se predice 28.7 (N = 12) - si su edad es $>$ a 36.5 se predice 25.4 (N = 50)

5.5. Evaluación de los modelos testeados

Al comparar todos los modelos se observa que el primer modelo es el que reporta menor MAE, menor MSE y menor RMSE de todos los modelos. Esto indica que es el modelo con menor magnitud promedio del error. Se podría concluir que el modelo de regresión lineal simple es el que presenta menor error. Empero, es importante considerar que es el que tiene menor complejidad.

Al comparar los dos árboles de decisión se observa que el segundo modelo reporta mejores índices de adecuación.

6. Conclusiones

Se logra cumplir parte de las hipótesis de este estudio. Por un lado, la CFC_I logra predecir negativamente el IMC. Por otro lado, la CFC_F no logra predecir el IMC como fue esperado. Primero, no se observó una correlación significativa entre CFC_F e IMC. Segundo, en el análisis de regresión multinivel se observa que la CFC_F se asocia positivamente con el IMC, y en el árbol de decisión no se observan un resultado claro de esta relación

Los resultados obtenidos podrían indicar que los resultados inmediatos son más determinantes en las acciones saludables que los distantes. Es decir, los sacrificios o placeres instantáneos tienen un mayor peso en la toma de decisiones referidas a la salud. Por ejemplo, a las personas que comienzan a correr les acondicionará con mayor fuerza el sacrificio de levantarse temprano, cansarse (consecuencia inmediata), que el tener el cuerpo deseado para el próximo verano (consecuencia futura). A su vez, se podría inferir que la subescala inmediata presenta mayor evidencia de validez que la CFC_F.

7. Referencias bibliográficas

- Adams, J. (2012). Consideration of immediate and future consequences, smoking status, and body mass index. *Health Psychology, 31*(2), 260–263. <https://doi.org/10.1037/a0025790>
- Camacho-Camargo, N., Alvarado, J., Paoli, M., Molina, Z., Cicchetti, R., Santiago, J., Huber, A., & Molina, Y. (2009). Relación entre el índice de masa corporal y las cifras de tensión arterial en adolescentes. *Revista Venezolana de Endocrinología y Metabolismo, 7*(2), 17–24. http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-31102009000200003&lng=es&nrm=iso&tlng=es
- Gick, M. (2014). An exploration of interactions between Conscientiousness and Consideration of Future Consequences on healthy eating. *Personality and Individual Differences, 66*, 181–187. <https://doi.org/10.1016/J.PAID.2014.03.020>
- Joireman, J., Shaffer, M. J., Balliet, D., & Strathman, A. (2012). Promotion Orientation Explains Why Future-Oriented People Exercise and Eat Healthy: Evidence From the Two-Factor Consideration of Future Consequences-14 Scale. *Personality and Social Psychology Bulletin, 38*(10), 1272–1287. <https://doi.org/10.1177/0146167212449362>
- Joireman, J., Strathman, A., & Balliet, D. (2006). Considering Future Consequences: An Integrative Model. *Judgments Over Time: The Interplay of Thoughts, Feelings, and Behaviors*. <https://doi.org/10.1093/ACPROF:OSO/9780195177664.003.0006>
- Kooij, D., Kanfer, R., & Rudolph, C. W. (2018). Future Time Perspective: A Systematic Review and Meta-Analysis. *Article in Journal of Applied Psychology*. <https://doi.org/10.1037/apl0000306>
- Murphy, L., & Dockray, S. (2018). The Consideration of Future Consequences and Health Behaviour : A Meta-Analysis. *Health Psychology Review, 0*(0), 1–55. <https://doi.org/10.1080/17437199.2018.1489298>
- Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology, 66*(4), 742–752. <https://doi.org/10.1037/0022-3514.66.4.742>
- van Beek, J., Antonides, G., & Handgraaf, M. J. J. (2013). Eat now, exercise later: The relation between consideration of immediate and future consequences and healthy behavior. *Personality and Individual Differences, 54*(6), 785–791. <https://doi.org/10.1016/j.paid.2012.12.015>