

Archivo de referencia de los programas de cálculo

Atlas de Riesgos para la Nutrición de la Niñez en México

Octubre de 2024

El presente archivo tiene como finalidad explicar el uso de los archivos de tratamiento en formato .R. Para poder utilizar estos scripts será necesario que cuentes con la última versión de R y es recomendable contar con un IDE para poder interpretar el lenguaje. Ambos se pueden hallar en:

<https://posit.co/download/rstudio-desktop/>

== Estructura básica de los archivos ==

Este directorio contiene cuatro clases de archivos:

- a) Archivos .R , que son los scripts de tratamiento en lenguaje R.
- b) Archivos .xlsx, hojas de cálculo en Excel que contienen los datos originales de las bases de datos del INSP y demás instituciones públicas mexicanas.
- c) Archivos de georreferenciación (extensiones .cpg, .cst, .dbf, .prj, .shp, .shx, .sld), los que se utilizan para trazar el mapa de México e interactuar con él.
- d) Archivos .png, que son los mapas generados por R en resolución media compatible con el diseño editorial.

*** Advertencia de uso de lenguaje técnico ***

En este archivo, debe entenderse “script”, “archivo de tratamiento”, “script de tratamiento” como lo mismo: los archivos con extensión .R que acabas de descargar.

== Funcionamiento general ==

Como regla general y por el diseño de los scripts, todos los archivos mencionados en un script .R deben encontrarse en el mismo directorio que el archivo .R para que este funcione adecuadamente. Antes de comenzar el uso de los scripts, deberás redirigir el directorio de trabajo. En Windows, puedes añadir usar esta línea de código después de la carga de librerías para operar los scripts:

```
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))
```

O bien, en Mac o en Windows puedes redirigir los archivos en este menú:

Session > Set Working Directory > To Source File Location

La descarga que hiciste de los archivos ya contiene la estructura ordenada para que funcione cada uno de los scripts; no es necesario mover ninguno y, de hacerlo, alguno de los scripts podría dejar de funcionar.

También es muy importante mencionar que cada vez que se ejecutan los scripts en R. alguno de los archivos en .xlsx es reemplazado. Esto viene especificado con claridad al interior de cada archivo .R.

*** Advertencia de sobreescritura ***

Lo anterior tiene dos implicaciones: por un lado, esto significa que si haces alguna modificación al archivo .xlsx y ejecutas el script, esta modificación será sobrescrita y perderás esos cambios. Por ello, se recomienda renombrar los archivos en .xlsx si decides trabajarlos manualmente.

Por otro lado, esto también significa que la apertura simultánea de los archivos .xlsx con el archivo .R puede impedir que el último funcione adecuadamente, pues R. no tiene permisos para reescribir archivos abiertos o generar copias automáticamente. Para que los scripts funcionen, los archivos de Excel deben estar cerrados.

== Pipeline de transformación de datos ==

A fin de facilitar la reproducción de los resultados del Atlas, los archivos de tratamiento fueron diseñados modularmente. Esto quiere decir que cada script contiene un número limitado de líneas de código que hacen referencia clara y explícitamente a una parte del proceso, como validación cruzada, generación de mapas y regresiones auxiliares.

Los archivos de tratamiento .R se encuentran numerados; aunque algunos scripts son totalmente independientes, se recomienda ejecutarlos en ese orden para llegar a los resultados del Atlas de forma lineal. A continuación, se describe el funcionamiento de los archivos principales.

= Directorio “Análisis a nivel estatal” =

1. 0_ATLAS_Mapas de prevalencia.R. Este script transforma los datos de la Ensanut 2022 en mapas de acuerdo con los estándares de las normas oficiales mexicanas. Se omite la utilización del IMC y los estándares de la OMS debido a que dicha métrica es una representación pobre de la composición corporal en menores de edad.
2. 1_ATLAS_CincoModelos.R. Este script pone a prueba cinco modelos econométricos diferentes para relacionar las variables dependientes con las independientes: Regresión Multinivel, Modelo de Efectos Fijos, Regresión Cuantílica Aditiva, GLM y Logit Multinivel con el Criterio de Información de Akaike (AIC).

3. 2_ATLAS_ValidaciónCruzadaIEMCPT.R y 2_ATLAS_ValidaciónCruzadaIyDFsEstatales.R. Estos scripts ponen a prueba el modelo XGBoost con el Error Promedio de la Raíz Cuadrada (RMSE).
4. 3_ATLAS_MapasDeRiesgosPE.R, 3_ATLAS_MapasDeRiesgosPT.R, 3_ATLAS_MapasDeRiesgosTE.R. Estos tres scripts actualizan el modelo XGBoost para usarlo con las ratios peso para la edad, peso para la talla y talla para la edad.
5. 4_ATLAS_Análisis de factores.R. Este script realiza un Análisis por Componentes Principales que se utilizó en el Atlas de forma exploratoria.
6. 5_ATLAS_Regresiones.R y 5_ATLAS_RegresionesTE.R. Este script lleva a cabo una serie de regresiones sencillas que se utilizaron en el análisis exploratorio y en la conformación de las primeras intuiciones.
7. 6_ATLAS_Mapas auxiliares.R y 6_ATLAS_Mapas auxiliaresTE.R. Este script está diseñado para elaborar los mapas que asisten la visualización de las primeras intuiciones, así como de los resultados de las regresiones lineales.
8. 7_ATLAS_SelectorVariablesImportantes.R . Este es un breve script auxiliar que no afecta los resultados pues sólo se escribió para elegir variables para hacer pruebas sobre conjuntos reducidos de factores.
9. 8_ATLAS_Tablas.R . Igualmente, este script es auxiliar y sirve para generar elementos de visualización. Se conserva en la versión pública por su posible utilidad para las y los investigadores.

= Directorio “Análisis a nivel estatal>Análisis segmentado por edad” =

10. 31_ATLAS_MapasSegmentadosTE.R. Este script actualiza el modelo XGBoost para utilizarlo en segmentos de 0 a 3, 4 a 6 y 7 a 9 años y generar los mapas correspondientes en función de la ratio de talla para la edad.
11. 32_ATLAS_MapasSegmentadosPE.R. Este script actualiza el modelo XGBoost para utilizarlo en segmentos de 0 a 3, 4 a 6 y 7 a 9 años y generar los mapas correspondientes en función de la ratio de peso para la edad.
12. 33_ATLAS_MapasSegmentadosPT.R. Este script actualiza el modelo XGBoost para utilizarlo en segmentos de 0 a 3, 4 a 6 y 7 a 9 años y generar los mapas correspondientes en función de la ratio de peso para la talla.

= Directorio “Análisis a nivel municipal” =

Este directorio contiene cuatro scripts experimentales que utilizaron datos del Censo de Población y Vivienda 2020 del INEGI para estudiar la capacidad de modelos de covariantes

no teóricos para predecir las alteraciones a la composición física de las niñas y niños mexicanos.

== Otros directorios ==

= Directorio “Análisis a nivel estatal>Mapas color deficiency” =

Esta carpeta contiene los mapas ajustados para que reduzcan las dificultades de visualización que experimentan las personas con alguna deficiencia en la percepción del color.