Actividad. Análisis descriptivos bivariantes en R

Objetivos

Por medio de esta actividad, en primer lugar, se pretende comprobar la normalidad de los alimentos y nutrientes seleccionados para asegurar los supuestos estadísticos necesarios. Luego, se calcularán los descriptivos de los alimentos y nutrientes seleccionados en cuatro categorías basadas en el sexo y el IMC recategorizado, a través de la obtención de medias y desviaciones estándar en caso de ser paramétrico, o medianas y rangos intercuartílicos si no lo son. Por último, se aplicará el contraste de hipótesis a cada uno de los alimentos y nutrientes para determinar si existen diferencias significativas entre las categorías de IMC en función del sexo.

Pautas de elaboración

En esta actividad usarás los datos del *dataset* alimentos\_nutrientes.csv para la elaboración de los análisis descriptivos bivariantes. Para esta actividad **se usarán únicamente los alimentos del 1 al 20 y todos los nutrientes del 1 al 19**, además de otras variables que se comentan a continuación. Al final de documento hay un apartado en el que se muestra el formato correcto de entrega de la actividad que hay que seguir. A continuación se muestra las pautas de elaboración que se debe de hacer en el siguiente orden en función de los objetivos:

1. Comprobar la normalidad de los alimentos y nutrientes disponibles y realizar una interpretación de los resultados obtenidos, asegurando que los datos sigan una distribución adecuada para los análisis posteriores:
   * Utilizar pruebas estadísticas de normalidad en R para evaluar la normalidad de los datos de alimentos y nutrientes y **reflejarlas en una tabla modelo 1** adjuntada al final de documento en la sección «Extensión y formato».
2. Calcular y analizar estadísticas descriptivas de los valores sin transformar (media + desviación estándar si son paramétricas, **mediana + rango intercuartílico** (p25-p75) si **no lo son**) de los alimentos y nutrientes seleccionados en cuatro categorías definidas por sexo y rango de IMC (**categoría 1**: sexo == 1 & IMC<=30; **categoría 2**: sexo == 1 & IMC>30; **categoría 3**: sexo == 2 & IMC<=30; **categoría 4**: sexo == 2 & IMC>30), permitiendo comparaciones específicas entre grupos:
   * Filtrar los datos según las categorías definidas (sexo y rango de IMC) utilizando operaciones lógicas en R vistas en clase.
   * Calcular las estadísticas descriptivas (media, desviación estándar, **mediana**, **rango intercuartílico**) de los alimentos y nutrientes para cada grupo utilizando las funciones en R mencionadas en clase y **reflejarlas en una tabla modelo 2** adjuntada al final de documento en la sección «Extensión y formato».
3. Aplicar e interpretar el **contraste de hipótesis** de cada uno de los alimentos y nutrientes seleccionados para ver si hay diferencias entre categorías de IMC en función del sexo:
   * Utilizar las pruebas estadísticas de contraste de hipótesis adecuadas para comparar las diferencias entre los grupos definidos por categorías de IMC y sexo.
   * Realizar las pruebas correspondientes en R y obtener los resultados de los contrastes de hipótesis y **reflejarlas en la tabla modelo 2** adjuntada al final de documento en la sección «Extensión y formato».

Como consejo, y de forma opcional, se puede emplear *loops* para sacar las estadísticas descriptivas (ver más aquí: <https://rpubs.com/Miguel_Tripp/forloops>) o *lapply* (<https://r-coder.com/lapply-en-r/>).

Extensión y formato

Para la resolución de esta actividad, deberás entregar 3 archivos:

* Un único fichero **R Markdown** (*.Rmd*) con todo el código y texto Markdown que hayas generado. No existe un límite de extensión para este fichero. Hay que indicar los siguientes argumentos:
  + title: «Resolución Actividad 1 máster Bioinformática UNIR (2023)»
  + author: «Nombre y 2 apellidos»
  + date: «yyyy-mm-dd»
  + output: html\_document

Para crear código y generarlo en el fichero R Markdown poner siempre:

```{r}

empezar aqui con el código

```

* Un fichero **HTML** generado a partir de dicho archivo R Markdown. Asegúrate de que aparecen todas las figuras, el texto correspondiente y todos los cuadros de código R que hayas introducido en el archivo *.Rmd* anterior.
* Un fichero **Word** únicamente con las 3 tablas modelos. Asegúrate de que aparecen los títulos correspondientes de las tablas, las notas a pie de tabla, indicando en todo caso abreviaturas en orden alfabético, test utilizados para sacar las estadísticas, notas a pie de tabla que se desee reflejar para aclarar cualquier asunto relevante de la tabla, etc. Los valores numéricos se representan con 1 decimal, a excepción de los valores p que se representan con 3 decimales. Los datos descriptivos se representan en una misma celda la media y desviación estándar [por ejemplo: 3.4 (3)] al igual que la **mediana y el RIQ [3.4 (1.3 a 9.8)].** El **formato de tablas que se debe de seguir** es el siguiente:

**Tabla 1.** {Aquí poner título}

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Test utilizado** | **Valor p** | **Interpretación** |
| Variable 1 | XX | XX | XX |
| Variable 2 | XX | XX | XX |
| … | XX | XX | XX |

{Aquí poner pie de tabla}

**Tabla 2**. {Aquí poner título}

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Sexo = 1** | | | **Sexo = 2** | | |
| **Variable** | **IMC≤30** | **IMC>30** | **p-value** | **IMC≤30** | **IMC>30** | **p-value** |
| Variable 1 | XX | XX | XXX | XX | XX | XXX |
| Variable 2 | XX | XX | XXX | XX | XX | XXX |
| … | XX | XX | XXX | XX | XX | XXX |

{Aquí poner pie de tabla}

Diccionario de variables (etiquetas)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Etiqueta** | **Tipo de variable** | **Valores de variables categóricas** |
| id | número de identificación | continua | - |
| altura | altura en cm | continua | - |
| peso | peso en kg | continua | - |
| IMC | índice de masa corporal | continua | - |
| sexo | hombre/mujer | categórica | 1 hombre; 2 mujer |
| edad | edad en años | continua | - |
| estado\_civil | categorías de estado civil | categórica | 0 soltero; 1 casado; 2 viudo; 3 separado; 4 divorciado |
| tabaco | hábito tabáquico | categórica | 0 nunca; 1 fumador; 2 exfumador; 3 *missing*; 4 fumador + exfumadores |
| colesterol | niveles sanguíneos de colesterol | categórica | 0 nunca; 1 <25 años; 2 25-44 años; 3 45-64 años; 4 >=65 años |
| hdl | niveles sanguíneos de HDL-c | categórica | 0 no me hecho análisis; 1 no recuerdo; 2 <35 (bajo); 3 36-60 (normal); 4 >60 (alto) |
| HTA | edad a la que tuvo HTA | categórica | 0 nunca; 1 <25 años; 2 25-44 años; 3 45-64 años; 4 >=65 años |
| hipercolesterolemia | edad a la que tuvo hipercolesterolemia | categórica | 0 nunca; 1 <25 años; 2 25-44 años; 3 45-64 años; 4 >=65 años |
| hipertrigliceridemia | edad a la que tuvo hipertrigliceridemia | categórica | 0 nunca; 1 <25 años; 2 25-44 años; 3 45-64 años; 4 >=65 años |
| ECV\_prev | prevalencia de ECV | categórica | 0 no; 1 si |
| diab\_prev | prevalencia de diabetes | categórica | 0 no; 1 si |
| hta\_prev | prevalencia de HTA | categórica | 0 no; 1 si |
| depres\_prev | prevalencia de depresión | categórica | 0 no; 1 si |
| FA\_prev | prevalencia de fibrilación auricular | categórica | 0 no; 1 si |
| cancer\_prev | prevalencia de cáncer | categórica | 0 no; 1 si |
| METs\_h\_semana | actividad física medida en METs-h a la semana | continua | - |
| cereales | consumo de cereales en gramos al día | continua | - |
| verdura | consumo de verdura en gramos al día | continua | - |
| fruta | consumo de fruta en gramos al día | continua | - |
| pescado | consumo de pescado en gramos al día | continua | - |
| legumbre | consumo de legumbre en gramos al día | continua | - |
| lacteos | consumo de lácteos en gramos al día | continua | - |
| carnes | consumo de carnes en gramos al día | continua | - |
| alimento1-131 | consumo de alimento en **gramos** al día | continua | - |
| nutriente1-19 | consumo de nutriente en **mg** al día | continua | - |

**Rúbrica**

|  |  |  |  |
| --- | --- | --- | --- |
| Actividad. Análisis descriptivos bivariantes en R | Descripción | Puntuación máxima  (puntos) | Peso  % |
| Pruebas | Construcción correcta de las pruebas y descriptivas tal y como se solicitan en el enunciado de la actividad | 3 | 30 % |
| Justificación | Justificación detallada en la respuesta a las preguntas planteadas y en el análisis de los gráficos generados | 3 | 30 % |
| Cuestiones | Resolución de las cuestiones planteadas de forma correcta | 2 | 20 % |
| Originalidad | Generación de un informe R Markdown original, empleando temas o estilos CSS, índices, etc. | 1 | 10 % |
| Orden | Exposición de lo que se pide en el enunciado de forma clara y ordenada | 1 | 10 % |
|  |  | **10** | **100 %** |