

## **Análisis estadístico**

### **Presentación de los datos**

Este estudio considera la toma de muestras de  $n=XXXX$  pacientes o estudiantes o atletas o muestras etc. (incluir especificación = comentar algunas características de estas muestras).

Para cada paciente o estudiante o atleta, se midieron  $YYY$  variables (en caso que no se hayan nombrado, pueden mencionarlas ahora).

Dentro de estas variables, se tienen variables cuantitativas (números) y/o cualitativas (características no numéricas).

### **Descripción del análisis a efectuar**

#### ***(Esto se puede hacer en todos los casos: análisis exploratorio de datos)***

Para todas las variables de interés en este estudio, se realizará un análisis univariado de cada variable por separado a través de un análisis exploratorio de datos utilizando elementos gráficos como gráficos de sectores circulares y gráficos de barra, entre otros para conocer la distribución de frecuencia de cada variable). Además se entregarán estadísticas descriptivas de cada variable, como medidas de tendencia central y posición (usando promedios, mediana y cuantiles) y medidas de dispersión (a través de la varianza y/o la desviación estándar). Con estos análisis se busca conocer las principales características de los datos bajo estudio e indagar acerca de los primeros hallazgos presentes en la muestra de datos bajo estudio.

#### ***(Si existe interés en comparar algún par de variables: análisis bivariado)***

Se realizarán algunos análisis bivariados de los datos entre pares de variables o grupos de variables que presenten especial interés. Se revisará la asociación que existe entre las variables  $X$  e  $Y$  en base a la evidencia recogida en... (estudios bibliográficos, análisis previos, hipótesis de estudio, etc.). Para esto se utilizará como medida el coeficiente de correlación de Spearman que corresponde a una medida de asociación entre dos variables aleatorias tanto continuas como discretas. También se puede emplear correlación de Pearson (otra medida de dependencia lineal entre dos variables aleatorias cuantitativas que entrega un valor entre 0 y 1, mientras más grande mayor relación entre las variables)

El análisis anterior se puede completar con tablas cruzadas y algunas pruebas como el test de independencia de Chi cuadrado o algunas pruebas  $t$  de Student para evaluar algunas hipótesis asociadas a alguna(s) de la(s) variable(s) involucradas en este estudio, como por ejemplo comprobar que alguna de las ideas preliminares sobre los datos se cumplen en el conjunto de datos analizados o no, asociándolo a un nivel de significancia que suele ser del 95%.

#### ***(Si las variables a contrastar pertenecen a individuos independientes: muestra independiente)***

#### ***(Si las variables a contrastar pertenecen a los mismos individuos: muestra pareada)***

- Si los valores de una muestra afectan los valores de la otra muestra, entonces las muestras son dependientes.

- Si los valores de una muestra no revelan información sobre los valores de la otra muestra, entonces las muestras son independientes.

La forma de calcular el estadístico  $t$  es diferente según las muestras sean independientes o apareadas. Las hipótesis nula y alternativa que se plantean son las siguientes:

$H_0$ : media de  $X = (\text{ó } \leq \text{ó } \geq)$  media de  $Y$

$H_1$ : media de  $X \neq (\text{ó } > \text{ó } <)$  media de  $Y$

**(Si se tienen más de 30 datos: verificar supuestos distribucionales)**

Como la muestra de datos a analizar considera  $n=XXXX$  observaciones, se efectuarán distintos análisis estadísticos para saber si podemos asumir que la muestra de datos podría seguir la distribución Gaussiana o conocida más comúnmente como distribución Normal. Para analizar esto, se emplearán análisis gráficos como gráficos de comparación de cuantiles y se mostrará el histograma de los datos. El histograma consiste en representar los datos mediante un curva de su distribución y superponer la curva de una distribución normal con la misma media y desviación estándar que muestran los datos, para así poder compararlas y ver qué tan cercanas se encuentran una de la otra. Otro gráfico para medir la Gaussianidad que es útil consiste en comparar los cuantiles de una distribución Normal (cuantiles teóricos) con los cuantiles obtenidos en los datos a analizar (cuantiles empíricos), este gráfico se llama qqplot o gráfico de cuantiles Gausianos.

El análisis gráfico anterior será acompañado de contrastes o test de hipótesis para chequear si los datos analizados podrían considerarse como Gaussianos.

El test denominado de Shapiro-Wilk se emplea para contrastar normalidad cuando el tamaño de la muestra es menor de 50. Para muestras grandes es equivalente al test de Kolmogorov-Smirnov. En todos ellos, se considera como hipótesis nula que los datos sí proceden de una distribución normal y como hipótesis alternativa que no lo hacen. El  $p\text{-value}$  de estos test indica la probabilidad de obtener una distribución como la observada si los datos proceden realmente de una población con una distribución normal.

Existen otros test propuestos en la literatura para contrastar la hipótesis de Gaussianidad de los datos. Por ejemplo el test de *Jarque-Bera* cuantifica que tanto se desvían los coeficientes de asimetría y curtosis de los esperados en una distribución normal.

El conocer si los datos se comportan como una distribución normal o no nos ayudará a determinar algunos análisis más específicos de los datos como también a las conclusiones y alcances que puede tener el estudio que se está efectuando.

**(Si se tienen menos de 30 datos: estadística no paramétrica)**

En caso de contar con pocas observaciones, no es posible verificar que ellas podrían provenir de una distribución gaussiana y todos los contrastes o pruebas que se realicen deben ser empleando estadísticas no paramétricas. La estadística no paramétrica, no realiza un supuesto distribucional de los datos, de manera que se abordarán pruebas y contrastes específicos.

### **Títulos Bibliografía Digital PUCV (para quienes necesitan sumar bibliografía)**

1. Probabilidad y Estadística para Ingeniería y Ciencias, Devore, Jay L.

<http://bibliografiaobligatoria.pucv.cl.pucv.idm.oclc.org/index.php/bibliotecapucv/catalog/book/95>

2. Introducción a la Probabilidad y Estadística, Mendenhall, William; Beaver, Robert y Beaver, Barbara

<http://bibliografiadigital.pucv.cl.pucv.idm.oclc.org/index.php/bibliotecapucv/catalog/book/526>

3. Exploratory Data Analysis Using R, Pearson, Ronald K

<http://bibliografiadigital.pucv.cl.pucv.idm.oclc.org/index.php/bibliotecapucv/catalog/book/118>

4. Probabilidad y Estadística, Canavos, George C.

<http://bibliografiadigital.pucv.cl.pucv.idm.oclc.org/index.php/bibliotecapucv/catalog/book/358>