

Estadística Descriptiva. Poblaciones y muestras.

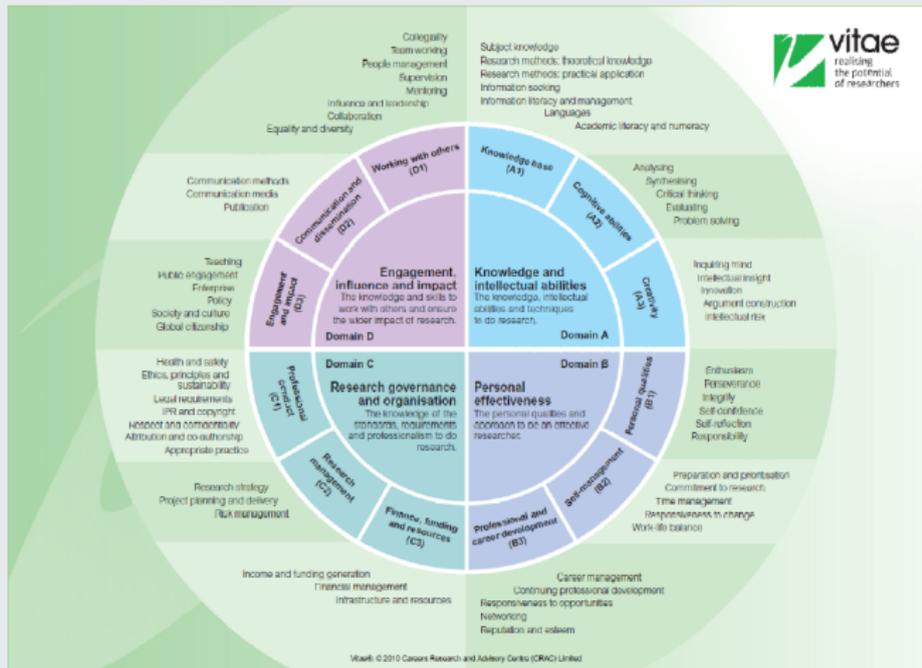
Francisco García García
fgarcia@cipf.es

CIPF's Research Development Programme

Índice

- 1 Introducción
- 2 Análisis descriptivo
- 3 Análisis inferencial

CIPF's Research Development Programme



¿Para qué utilizamos el análisis estadístico?

- Para **responder** a las necesidades del desarrollo científico y tecnológico.
- Para **cuantificar y reducir** la incertidumbre.
- Para **inferir** acerca de toda la población basándonos en una muestra.



Fase 1

- Plan
- Diseño

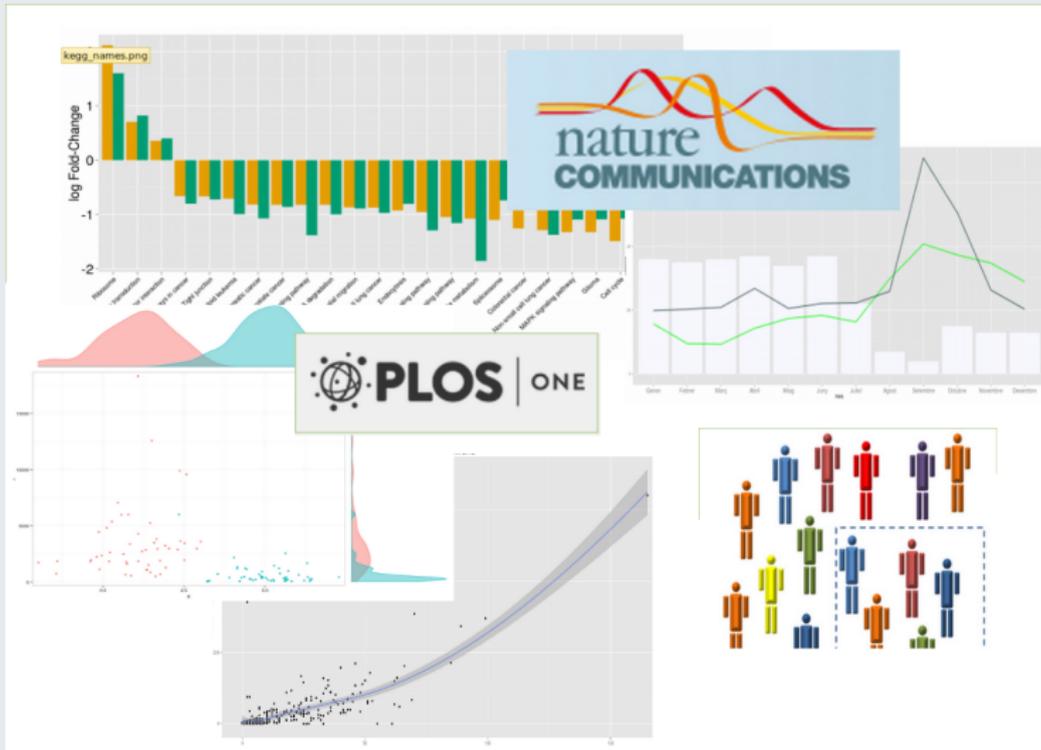
Fase 2

- Recogida de datos
- Procesamiento de los datos
- Análisis de datos

Fase 3

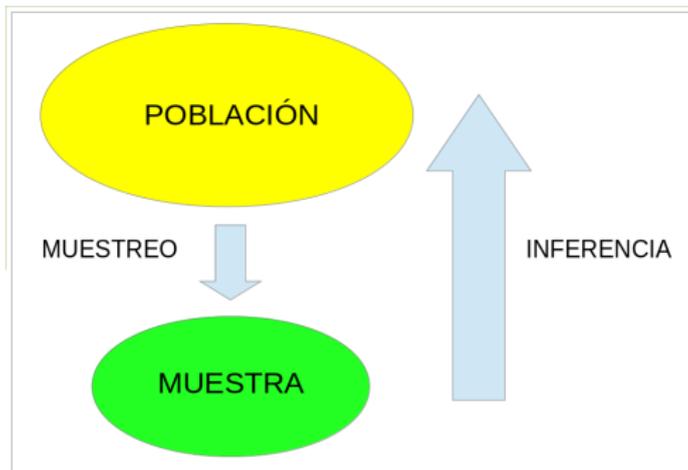
- Presentación
- Interpretación
- Publicación

Fase 3: análisis estadístico de datos en publicaciones, informes, presentaciones...



Análisis estadístico de los datos:

- Poblaciones y muestras
- Análisis Estadístico Descriptivo
- Análisis Estadístico Inferencial



¿Qué es R?

- R es un entorno de programación que permite hacer análisis estadísticos.
- Es una potente herramienta para generar gráficos de cualquier tipo.
- R es de código abierto y pertenece al proyecto GNU de software libre.
- Disponible para las plataformas Linux, Macintosh y Windows.
- R es un software orientado tanto a usuarios principiantes como profesionales e investigadores que necesiten analizar y representar gráficamente sus datos.
- R es gratuito. No necesitamos ningún tipo de licencia.

¿Cómo obtenemos e instalamos R?

- Descargamos **R** en <http://cran.r-project.org/>
- Tras la descarga, ejecutamos el archivo y aparecerá un asistente que nos guiará en el proceso. En unos minutos el software quedará instalado.
- Va bien instalarse **RStudio**, un interfaz que facilita el trabajo con **R**. Está disponible en <http://www.rstudio.com/>
- También hay disponible una versión libre. El modo de instalación es similar: descargamos la herramienta y la instalamos siguiendo el asistente.
- De modo que podemos trabajar directamente desde **R** o bien desde **RStudio**.

Trabajando con R

- Abre estos tres documentos con ejemplos y ejercicios para conocer los objetos habituales en R: **vectores**, **matrices**, **listas**.

¿Cómo leemos los datos en R?

- Si nuestro fichero tiene formato EXCEL, previamente lo guardaremos en formato **csv**, así será sencillo y rápido de importar desde R con la función **read.csv**. (Es conveniente que la primera línea contenga los nombres de las variables cortos e informativos).
- Si nuestro fichero tiene formato **txt**, utilizaremos la función **read.table** desde R para leer estos datos.
- Es posible leer desde R la mayoría de los formatos disponibles mediante otras funciones. Si hay alguna dificultad, siempre está la posibilidad de guardar el fichero original como txt o csv (formatos estándar) y a continuación, utilizaremos las funciones anteriores para su lectura.

Trabajando con R

- Abre este documento con ejemplos y ejercicios: **Lectura de datos**

Operaciones habituales en la gestión de datos con R

- **Ordenar** la base de datos por una o varias variables.
- **Seleccionar o filtrar** registros de interés.
- **Recodificar** una variable.
- **Categorizar** una variable cuantitativa en una cualitativa.

Trabajando con R

- Abre este documento con ejemplos y ejercicios: **Gestión de datos**
- Abre este documento con ejemplos y ejercicios: **Guardar resultados**

Análisis Descriptivo Univariante

Variables cualitativas o categóricas

- **Ejemplos.** Nominales: sexo. Ordinales: nivel de estudios
- **Descripción gráfica:** diagrama de barras, sectores
- **Descripción numérica:** tablas de frecuencias

Variables cuantitativas

- **Ejemplos.** Discretas: número de hijos. Continuas: peso, altura.
- **Descripción gráfica:** diagrama de barras, sectores para discretas y diagrama de cajas e histogramas para continuas.
- **Descripción numérica:** media, mínimo, máximo, varianza, mediana...

Análisis Descriptivo Bivariante

A. Cualitativa vs. cualitativa

- **Ejemplos.** Fumar (sí,no) y sexo (hombres, mujeres).
- **Descripción gráfica:** diagrama de barras, sectores.
- **Descripción numérica:** tablas de contingencia.

B. Cualitativa vs. cuantitativa

- **Ejemplo.** Crecimiento celular en 3 grupos de ratones (con diferentes tratamientos).
- **Descripción gráfica:** diagrama de cajas, histogramas para continuas, barras de error.
- **Descripción numérica:** media, mínimo, máximo, varianza, mediana... por cada una de las categorías de la variable cualitativa.

Análisis Descriptivo Bivariante

C. Cuantitativa vs. cuantitativa

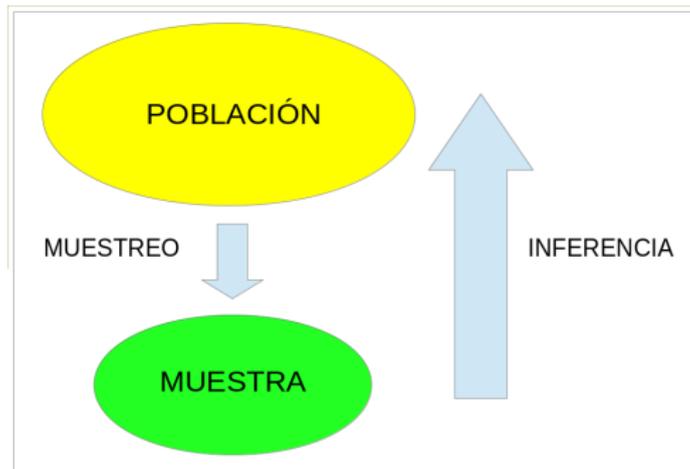
- **Ejemplos.** Altura y peso.
- **Descripción gráfica:** gráfico de dispersión (nube) o de puntos.
- **Descripción numérica:** coeficiente de correlación lineal (Pearson, Spearman...).

Trabajando con R

- Abre este documento con ejemplos y ejercicios: **Análisis descriptivo**

Métodos inferenciales:

- Estimación puntual y por intervalos
- Contrastes de hipótesis



A. Cualitativa vs. cualitativa

- **Ejemplos.** Fumar (sí,no) y sexo (hombres, mujeres).
- **Descripción gráfica:** diagrama de barras, sectores.
- **Descripción numérica:** tablas de contingencia.

Contraste de hipótesis: Chi-cuadrado.

- **H₀:** las variables son independientes.
- **H_a:** las variables son dependientes.

Trabajando con R

- Abre este documento con ejemplos y ejercicios: **Análisis de datos categóricos**

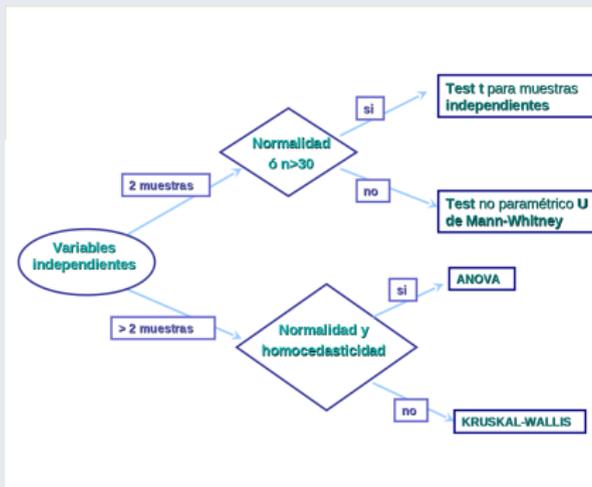
B. Cualitativa vs. cuantitativa

- **Ejemplo.** Crecimiento celular en 3 grupos de ratones (con diferentes tratamientos).
- **Descripción gráfica:** diagrama de cajas, histogramas para continuas, barras de error.
- **Descripción numérica:** media, mínimo, máximo, varianza, mediana... por cada una de las categorías de la variable cualitativa.

Contraste de hipótesis: comparación de medias (medianas).

- **H₀:** todas las medias de los grupos son iguales.
- **H_a:** existen diferencias entre las medias de los grupos.

Contrastes de hipótesis: comparación de medias (medianas)



Trabajando con R

- Abre este documento con ejemplos y ejercicios: **Inferencia de medias**
- Abre este documento con ejemplos y ejercicios: **Análisis de la Varianza (ANOVA)**

C. Cuantitativa vs. cuantitativa

- **Ejemplos.** Altura y peso.
- **Descripción gráfica:** gráfico de dispersión (nube) o de puntos.
- **Descripción numérica:** coeficiente de correlación lineal (Pearson, Spearman...).

Contraste de hipótesis: para evaluar la significación del coeficiente de correlación.

- **H₀:** el coeficiente de correlación es 0.
- **H_a:** el coeficiente de correlación es diferente de 0.

