



# DIPLOMADO

## DISEÑO Y ANÁLISIS DE DATOS PARA LA INVESTIGACIÓN CIENTÍFICA CON R Y R STUDIO

## PROGRAMACIÓN PARA LA ESTADÍSTICA APLICADA



## TEMARIO

## **Unidad 1. Metodología de la Investigación y principios de programación en R y RStudio**

### **1.1 Fundamentos de la investigación e instalación de R y Rstudio.**

- 1.1.1 Tipos de investigaciones (exploratoria, descriptiva, correlacional y explicativa).
- 1.1.2 Diseños de investigaciones (transversales y longitudinales).
- 1.1.3 Tipos de variables (dependientes e independientes).
- 1.1.4 Identificación del problema de investigación.
- 1.1.5 Planteamiento de preguntas de investigación, objetivos e hipótesis.
- 1.1.6 Instalación de R y Rstudio.

### **1.2 Programación en R e importación de bases de datos**

- 1.2.1 Menús y ventanas en Rstudio
- 1.2.2 Instalación de paquetes
- 1.2.3 Desarrollo de scripts: escritura de código y comentarios
- 1.2.4 Estructura de una función en R.
- 1.2.5 Objetos: vectores, data frames, matrices y listas.
- 1.2.6 Lectura de archivos delimitados por comas (.csv).
- 1.2.7 Lectura de archivos de hojas de cálculo (Microsoft Excel).
- 1.2.8 Lectura de bases de datos de otras Fuentes (web scrapping).
- 1.2.9 Validación de bases de datos (estructura, encabezados, tipos de variables).
- 1.2.10 Web scrapping.

## **Unidad 2. Extracción y manipulación de información**

### **2.1 Manipulación de datos.**

- 2.1.1 Paquete Tidyverse: ciencia de datos.
- 2.1.2 Selección de columnas.
- 2.1.3 Filtro de observaciones.
- 2.1.4 Creación de nuevas variables.
- 2.1.5 Ordenamiento de valores.
- 2.1.6 Resumen de columnas y bases de datos.
- 2.1.7 Agrupamientos.
- 2.1.8 Formatos de bases de datos.

### **2.2 Exploración y análisis gráfico de datos.**

- 2.2.1 Presentación de ggplot2.
- 2.2.2 Anatomía de un gráfico.
- 2.2.3 Herramientas disponibles para la creación de gráficos de alta calidad.
- 2.2.4 Gráficos de una variable: histogramas y densidad.
- 2.2.5 Gráficos de dos y tres variables: box-plot, violín, dispersión, barras, densidad.
- 2.2.6 Gráficos de promedios.
- 2.2.7 Asignando títulos, nombres de ejes y etiquetas.
- 2.2.8 Exportación de gráficos.

## **Unidad 3. Conceptos de estadística descriptiva e introducción a la estadística inferencial**

### **3.1 Estadística descriptiva: conceptos básicos.**

- 3.1.1 ¿Qué es la estadística descriptiva?
- 3.1.2 Medidas de tendencia central: media y mediana.
- 3.1.3 Medidas de variabilidad: desviación estándar y varianza.
- 3.1.4 Resúmenes estadísticos.

### **3.2 Estadística inferencial: conceptos básicos.**

- 3.2.1 Tamaño de muestra para estimar la media y proporción poblacional.
- 3.2.2 Intervalos de confianza para un parámetro
- 3.2.3 Contraste de hipótesis para un parámetro
- 3.2.4 Valores de P.
- 3.2.5 Normalidad.
- 3.2.6 Homogeneidad de varianza.

## **Unidad 4. Análisis de asociación y causalidad.**

### **4.1 Correlación.**

- 4.1.1 Asociación no implica causalidad.
- 4.1.2 Enfoque paramétrico: Correlación de Pearson.
  - 4.1.2.1 Supuestos e hipótesis estadísticas.
  - 4.1.2.2 Ejecución en R: `cor.test(... method = "pearson")`
  - 4.1.2.3 Interpretación de resultados.
  - 4.1.2.4 Visualización gráfica de resultados.
- 4.1.3 Enfoque no paramétrico: correlación de Spearman y Kendall
  - 4.1.3.1 Supuestos e hipótesis estadísticas.
  - 4.1.3.2 Ejecución en R: `cor.test(... method = "spearman"/"kendall")`
  - 4.1.3.3 Interpretación de resultados: coeficiente de correlación.
  - 4.1.3.4 Visualización gráfica de resultados.

### **4.2 Regresión lineal simple y múltiple**

- 4.2.1 ¿Qué es el modelo de regresión lineal simple y cuándo se utiliza?
  - 4.2.1.1 Supuestos del modelo de regresión lineal: normalidad de residuales, homocedasticidad de residuales e independencia de datos.
  - 4.2.1.2 Hipótesis estadísticas del modelo de regresión.
  - 4.2.1.3 Ajuste de un modelo de regresión lineal simple en R: función `lm()`
  - 4.2.1.4 Validación estadística y gráfica de un modelo de regresión.
  - 4.2.1.5 Interpretación de resultados: función `summary()`.
  - 4.2.1.6 Visualización gráfica del modelo de regresión lineal simple.

#### 4.2.2 Modelo de regresión múltiple con variables continuas.

4.2.2.1 Supuestos e hipótesis estadísticas.

4.2.2.2 Validación gráfica del modelo lineal múltiple.

4.2.2.3 Interpretación de resultados.

4.2.2.4 Visualización gráfica de resultados.

4.2.2.5 Comparación de modelos.

#### 4.2.3 Modelo de regresión múltiple con variables continuas y categóricas (ANCOVA)

4.2.3.1 Supuestos e hipótesis estadísticas.

4.2.3.2 Validación gráfica del modelo lineal múltiple.

4.2.3.3 Interpretación de resultados.

4.2.3.4 Visualización gráfica de resultados.

### 4.3 Modelos lineales generalizados.

4.3.1 ¿Qué es un modelo lineal generalizado?

4.3.2 Distribuciones diferentes a la distribución normal.

4.3.3 Modelos lineales generalizados:

4.3.3.1 Modelo logístico (distribución binomial).

4.3.3.1.1 Supuestos estadísticos.

4.3.3.1.2 Ejecución en R: Función `glm(... family = "binomial")`

4.3.3.1.3 Validación gráfica del modelo.

4.3.3.1.4 Visualización gráfica de resultados.

4.3.3.2 Modelo para conteos (distribución Poisson).

4.3.3.2.1 Supuestos estadísticos.

4.3.3.2.2 Ejecución en R: Función `glm(... family = "poisson")`

4.3.3.2.3 Validación gráfica del modelo.

4.3.3.2.4 Visualización gráfica de resultados.

4.3.3.2.5 Alternativas para tratar con modelos lineales generalizados con distribución poisson sobredispersos: `quasivarianza` y distribución binomial negativa.

## Unidad 5. Comparando poblaciones

### 5.1 Pruebas paramétricas y no paramétricas para comparar dos poblacionales.

5.1.1 Enfoque paramétrico:

5.1.1.1 Prueba t-Student para muestras independientes.

5.1.1.1.1 Supuestos e hipótesis estadísticas.

5.1.1.1.2 Ejecución en R: `t.test()`

5.1.1.1.3 Interpretación de resultados y parámetros estadísticos.

5.1.1.1.4 Visualización gráfica de resultados.

5.1.1.2 Prueba t-Student para muestras relacionadas.

5.1.1.2.1 Supuestos e hipótesis estadísticas.

5.1.1.2.2 Ejecución en R: `t.test(... paired = T)`

5.1.1.2.3 Interpretación de resultados y parámetros estadísticos.

5.1.1.2.4 Visualización gráfica de resultados

**5.1.2 Enfoque no paramétrico:**

**5.1.2.1 Prueba U de Mann-Whitney (muestras independientes):**

5.1.2.1.1 Supuestos e hipótesis estadísticas.

5.1.2.1.2 Ejecución en R: `wilcox.test()`.

5.1.2.1.3 Interpretación de resultados.

5.1.2.1.4 Visualización gráfica de resultados.

**5.1.2.2 Prueba de Wilcoxon (muestras dependientes o relacionadas):**

5.1.2.2.1 Supuestos e hipótesis estadísticas.

5.1.2.2.2 Ejecución en R: `wilcox.test(... paired = T)`.

5.1.2.2.3 Interpretación de resultados.

5.1.2.2.4 Visualización gráfica de resultados.

**5.2 Pruebas paramétricas y no paramétricas para comparar tres o más poblacionales I.**

**5.2.1 Enfoque paramétrico:**

5.2.1.1 Análisis de varianza (ANOVA o ANDEVA) de una vía.

5.2.1.2 Supuestos estadísticos: normalidad, homogeneidad de varianza e independencia de datos.

5.2.1.3 Ejecución en R: Función `aov()` y Función `lm()`.

5.2.1.4 Validación gráfica del modelo.

5.2.1.5 Comparaciones: pruebas de contrastes y pruebas post hoc.

5.2.1.6 Representación gráfica de los resultados.

**5.2.2 Enfoque no paramétrico:**

5.2.2.1 Prueba de Kruskal-Wallis.

5.2.2.2 Supuestos estadísticos: homogeneidad de varianza e independencia de datos.

5.2.2.3 Ejecución en R: Función `kruskal.test()`.

5.2.2.4 Comparaciones múltiples: prueba de Wilcoxon.

5.2.2.5 Representación gráfica de los resultados.

**5.3 Pruebas paramétricas y no paramétricas para comparar tres o más poblacionales II.**

**5.3.1 Enfoque paramétrico:**

5.3.1.1 Análisis de varianza de dos vías.

5.3.1.1.1 Supuestos e hipótesis estadísticas.

5.3.1.1.2 Validación gráfica del modelo.

5.3.1.1.3 Comparaciones múltiples.

5.3.1.1.4 Visualización gráfica de resultados.

**5.3.2 Enfoques alternativos al paramétrico:**

5.3.2.1 Análisis de permutación.

5.3.2.2 Modelos lineales generalizados.

## **5.4 Pruebas paramétricas y no paramétricas para comparar tres o más poblacionales III.**

### **5.4.1 Enfoque paramétrico:**

- 5.4.1.1 Análisis de varianza para medidas repetidas.
- 5.4.1.2 Supuestos e hipótesis estadísticas.
- 5.4.1.3 Ejecución en R:
- 5.4.1.4 Validación gráfica y estadística de los supuestos.
- 5.4.1.5 Comparaciones múltiples.
- 5.4.1.6 Visualización gráfica de resultados.

### **5.4.2 Enfoque no paramétrico:**

- 5.4.2.1 Prueba de Friedman.
- 5.4.2.2 Supuestos e hipótesis estadísticas.
- 5.4.2.3 Comparaciones múltiples.
- 5.4.2.4 Visualización gráfica de resultados.

## **Unidad 6. Análisis multivariado**

### **6.1 Análisis Factorial Exploratorio**

- 6.1.1 Uso del paquete psych para AFE.
  - 6.1.1.1 Número de factores latentes.
  - 6.1.1.2 AFE con variables categóricas.

### **6.2 Análisis de componentes principales**

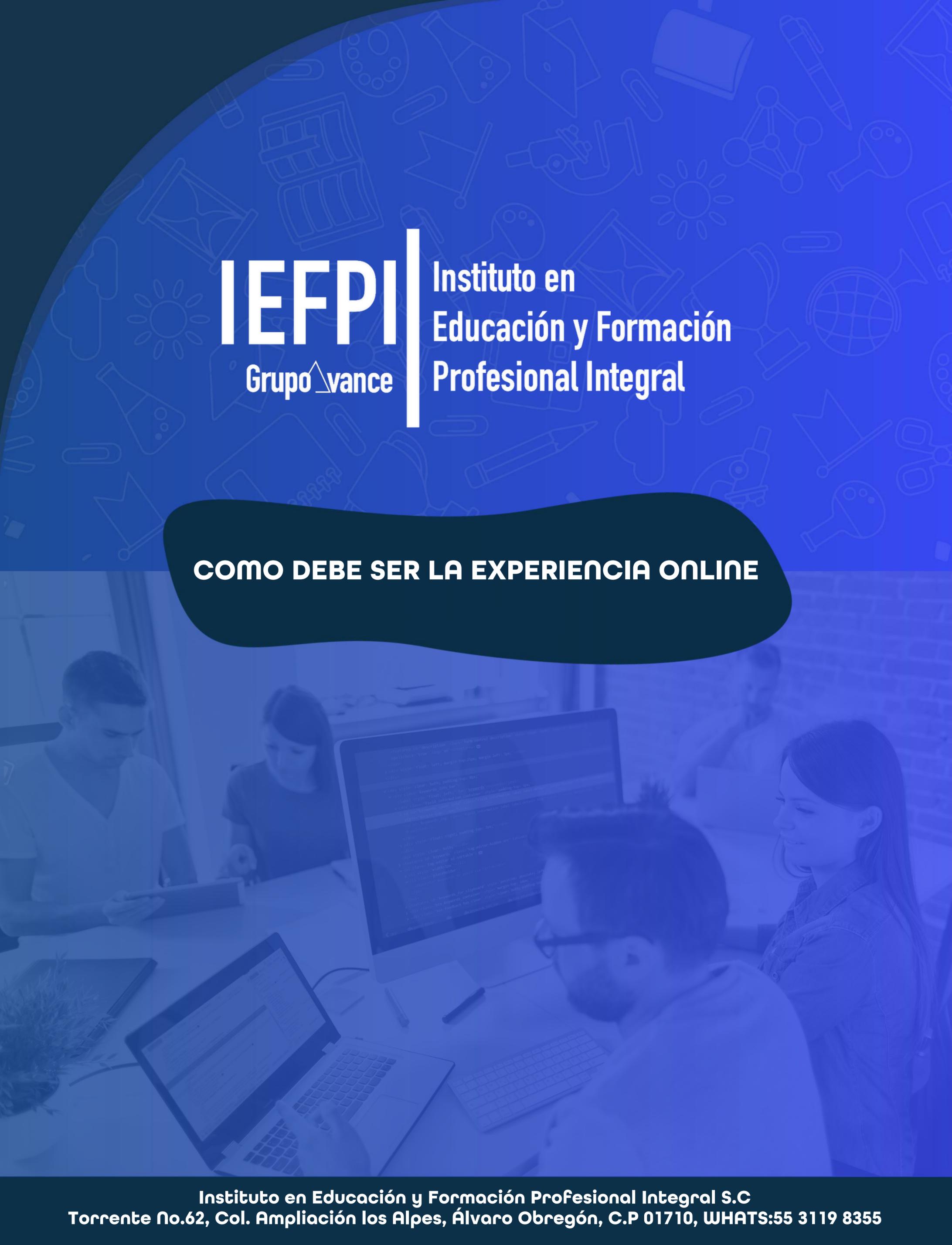
- 6.2.1 ¿Qué es el análisis de componentes principales?
- 6.2.2 Implementación con FactoMineR::PCA()
- 6.2.3 Métodos de rotación.
- 6.2.4 Visualizaciones: screeplot y factor maps.

### **6.3 Técnicas agrupamiento**

- 6.3.1 ¿Qué son las técnicas de agrupamiento?
- 6.3.2 Agrupamiento con kmeans()
- 6.3.3 Agrupamiento por hclust()
- 6.3.4 Número de grupos
- 6.3.5 Identificación de membresía
- 6.3.6 Agrupamientos basados en modelos:
  - 6.3.6.1 Modelos de mezclas normales.
  - 6.3.6.2 Uso del paquete/librería mclust.
  - 6.3.6.3 Visualización gráfica de resultados.

The background is a solid blue color with a repeating pattern of white line-art icons. These icons include various educational symbols such as books, pencils, globes, lightbulbs, DNA helices, chemical flasks, and geometric shapes like cubes and spheres. The icons are scattered across the entire page, creating a textured, academic feel.

**IEFPI** Instituto en  
Educación y Formación  
Grupo Vance Profesional Integral



# IEFPI

Grupo  Vance

Instituto en  
Educación y Formación  
Profesional Integral

**COMO DEBE SER LA EXPERIENCIA ONLINE**