

## **ESQUEMA DE R Y R-COMMANDER PARA ESTADÍSTICA DESCRIPTIVA.**

### **DATOS**

1. Abrir datos:  
*Datos -> Cargar Archivo de Datos*
2. Importar datos desde texto:  
*Datos -> Importar Datos -> Desde Archivo de Texto, Portapapeles, etc.*
3. Importar datos desde Excel:  
*Datos -> Importar Datos -> Desde Conjunto de Datos Excel, Access o Dbase.*
4. Exportar datos desde R-Commander a texto:  
*Datos -> Conjunto de Datos Activo -> Exportar el Conjunto de Datos Activo.*
5. Guardar datos:  
*Datos -> Conjunto de Datos Activo -> Guardar el Conjunto de Datos Activo.*
6. Resumir el conjunto de datos:  
*Estadísticos -> Resumen -> Conjunto de Datos Activo.*
7. Cambiar variable numérica a cualitativa:  
*Datos -> Modificar Variables del Conjunto de Datos Activo -> Convertir Variable Numérica en Factor.*
8. Agrupar variable numérica en intervalos:  
*Datos -> Modificar Variables del Conjunto de Datos Activo -> Segmentar Variable Numérica.*  
*Nombres de los niveles: Rangos*  
*Método de segmentación: Segmentos Equidistantes*

### **ESTADÍSTICA DESCRIPTIVA UNIVARIANTE**

1. Realizar un análisis descriptivo de variable numérica:  
*Estadísticos -> Resúmenes -> Resúmenes numéricos.*
2. Realizar un análisis descriptivo de variable numérica en función de variable cualitativa o de variable numérica agrupada por intervalos:  
*Estadísticos -> Resúmenes -> Resúmenes numéricos (Resumir por grupos).*
3. Construir una tabla de frecuencias o análisis descriptivo de variable cualitativa:  
*Estadísticos -> Resúmenes -> Distribución de frecuencias.*
4. Realizar representaciones gráficas:  
*Gráficas -> Diagrama de barras*  
*Gráficas -> Sectores*  
*Gráficas -> Histograma*  
*Gráficas -> Diagrama de caja*
5. Diagrama de caja para variable numérica en función de variable cualitativa o variable numérica agrupada por intervalos:  
*Gráficas -> Diagrama de caja (Resumir por grupos).*
6. Calcular coeficientes de asimetría y curtosis:  
*En ventana de instrucciones escribir órdenes:*  
*kurtosis(NombreBaseDatos\$NombreVariable)*  
*skewness(NombreBaseDatos\$NombreVariable)*  
*Y pulsar ejecutar. Si no funciona cargar el paquete e1071, en Rgui:*  
*Paquetes -> Cargar Paquete.*

## ESTADÍSTICA DESCRIPTIVA BIVARIANTE

1. Construir tabla de doble entrada para variables cualitativas o cuantitativas previamente agrupadas por intervalos:  
*Estadísticos -> Tablas de contingencia -> Tabla de doble entrada.*  
*Para marginales marcar opción: Porcentajes Totales*  
*Para coeficiente Chi-cuadrado marcar opción: Test de Independencia Chi-Cuadrado*  
*Para frecuencias esperadas marcar opción: Imprimir Frecuencias Esperadas*
2. Obtener coeficiente de correlación por rangos de Spearman y coeficiente de correlación de Pearson:  
*Estadísticos -> Resúmenes -> Matriz de Correlaciones.*
3. Realizar el diagrama de dispersión:  
*Gráfica -> Diagrama de dispersión.*
4. Examinar la matriz completa de diagramas de dispersión:  
*Gráfica -> Matriz de diagramas de dispersión.*
5. Realizar un ajuste lineal o regresión mínimo cuadrática:  
*Estadísticos -> Ajuste de Modelos -> Regresión Lineal.*  
*A destacar:*  
*Expresión de la recta, valor de los coeficientes de la recta y bondad del ajuste.*
6. Comparar modelos alternativos.  
*Estadísticos -> Ajuste de modelos -> Modelo lineal.*

Tipo de Modelo de Ajuste	Expresión: $Y = f(X)$	Fórmula en R
Recta de regresión $Y/X$	$Y = a + bX$	$Y \sim X$
Recta de regresión sin constante	$Y = bX$	$Y \sim X - 1$
Ajuste Parabólico o Cuadrático	$Y = a + bX + cX^2$	$Y \sim X + I(X^2)$
Ajuste Cúbico	$Y = a + bX + cX^2 + dX^3$	$Y \sim X + I(X^2) + I(X^3)$
Ajuste Logarítmico	$Y = a + b \log(X)$	$Y \sim \log(X)$
Ajuste Inverso X	$Y = a + \frac{b}{X}$	$Y \sim I(1/X)$
Ajuste Exponencial	$Y = \exp(a + bX)$	$\log(Y) \sim X$
Ajuste Potencial o Multiplicativo	$Y = aX^b$	$\log(Y) \sim \log(X)$

Tabla 4.1. Fórmulas para distintos tipos de ajustes en R.

7. Realizar predicciones.  
*Escribir orden en instrucciones:*  
*predict (NombreModelo,data.frame(NombreDeVariableX=c(valor1,valor2,valor3)))*  
*Pulsar ejecutar.*

## GUARDAR

19. Guardar datos:  
*Datos -> Conjunto de datos activo -> Guardar el conjunto de datos activos.*
20. Guardar resultados:  
*Fichero -> Guardar los resultados.*
21. Guardar gráficos:  
*Gráficas -> Guardar gráfico en fichero -> Como mapa de bits para guardarlo como imagen.*
22. Guardar instrucciones:  
*Fichero -> Guardar instrucciones*
23. Guardar modelos de regresión:  
*Fichero -> Guardar área de trabajo.*

# ESQUEMA DE R Y R-COMMANDER PARA VARIABLE ALEATORIA E INFERENCIA PARAMÉTRICA.

## VARIABLE ALEATORIA DISCRETA.

*Distribuciones → Distribuciones discretas*

Distribuciones discretas utilizadas:

- *Distribución binomial*
- *Distribución de Poisson*
- *Distribución geométrica*
- *Distribución hipergeométrica*
- *Distribución binomial negativa*

1. Obtener percentiles (ejemplo binomial):

*Distribuciones → Distribuciones discretas → Distribución binomial → Cuantiles binomiales:*

*Cola izquierda: Para obtener  $k$  tal que  $P(X \leq k) = p$*

*Cola derecha: Para obtener  $k$  tal que  $P(X > k) = p$*

2. Obtener valores de función de distribución (ejemplo binomial):

*Distribuciones → Distribuciones discretas → Distribución binomial → Probabilidades binomiales acumuladas:*

*Cola izquierda: Para obtener  $P(X \leq k)$*

*Cola derecha: Para obtener  $P(X > k)$*

3. Obtener la función de probabilidad (ejemplo binomial):

*Distribuciones → Distribuciones discretas → Distribución binomial → Probabilidades binomiales.*

4. Obtener gráficas de la función de probabilidad y la función de distribución (ejemplo binomial):

*Distribuciones → Distribuciones discretas → Distribución binomial → Gráfica de la distribución binomial.*

5. Generar números aleatorios (ejemplo binomial):

*Distribuciones → Distribuciones discretas → Distribución binomial → Muestra de una distribución binomial.*

*Nota: En la distribución hipergeométrica, al introducir parámetros: “m” indica el número de elementos en la población que poseen la característica objeto de estudio, “n” será el resto de elementos y “k” el número de elementos seleccionados.*

## VARIABLE ALEATORIA CONTINUA.

*Distribuciones → Distribuciones continuas*

Distribuciones continuas utilizadas:

- *Distribución normal*
- *Distribución t*
- *Distribución Chi-cuadrado*
- *Distribución F*
- *Distribución exponencial*

1. Obtener percentiles (ejemplo normal):  
*Distribuciones → Distribuciones continuas → Distribución normal → Cuantiles normales:*  
*Cola izquierda: Para obtener  $k$  tal que  $P(X < k) = p$*   
*Cola derecha: Para obtener  $k$  tal que  $P(X > k) = p$*
2. Obtener valores de función de distribución (ejemplo normal):  
*Distribuciones → Distribuciones continuas → Distribución normal → Probabilidades normales:*  
*Cola izquierda: Para obtener  $P(X < k)$*   
*Cola derecha: Para obtener  $P(X > k)$*
3. Obtener gráficas de la función de densidad y la función de distribución (ejemplo normal):  
*Distribuciones → Distribuciones continuas → Distribución normal → Gráfica de la distribución normal.*
4. Generar números aleatorios (ejemplo normal):  
*Distribuciones → Distribuciones continuas → Distribución normal → Muestra de una distribución normal.*

## INFERENCIA PARAMÉTRICA PARA UNA POBLACIÓN

1. Intervalo de confianza y contraste de hipótesis para la media:  
*Estadísticos → Medias → Test t para una muestra:*  
**Salida:**  
*t: valor del estadístico de contraste*  
*p-value: p-valor (si p-valor <  $\alpha$  rechazar  $H_0$ , si p-valor >  $\alpha$  aceptar  $H_0$ )*  
*confidence interval: intervalo de confianza (nos da los límites entre los que se encuentra el parámetro estimado con una probabilidad  $1 - \alpha$ )*  
*mean of x: media muestral*
2. Intervalo de confianza y contraste de hipótesis para la proporción (para variables cualitativas dicotómicas):  
*Estadísticos → Proporciones → Test de proporciones para una muestra.*  
**Nota:** *El intervalo que se obtiene es para la primera categoría en orden alfabético.*  
**Tipo de prueba:** *Aproximación normal.*  
**Salida:**  
*X-squared: valor del estadístico de contraste*  
*p-value: p-valor (si p-valor <  $\alpha$  rechazar  $H_0$ , si p-valor >  $\alpha$  aceptar  $H_0$ )*  
*confidence interval: intervalo de confianza (nos da los límites entre los que se encuentra el parámetro estimado con una probabilidad  $1 - \alpha$ )*  
*p: proporción muestral*
3. Intervalo de confianza para la varianza:  
*Escribimos en R-Gui para crear la función:*  

```
>var.interval=function(data,conf.level=0.95){
+df=length(data-1)
+chilower=qchisq((1-conf.level)/2,df)
+chiupper=qchisq((1-conf.level)/2,df,lower.tail=FALSE)
+v=var(data)
+c(df*v/chiupper, df*v/chilower)
+}
```

**Para construir intervalo:**  

```
>var.interval(NombreDatos$NombreVariable)
```

Para guardar función:

En Rgui: Archivo → Guardar área de trabajo

Para volver a utilizar función:

En Rgui: Archivo → Cargar área de trabajo

Para construir intervalo variando intervalo de confianza (ejemplo para 99%):

>var.interval(NombreDatos\$NombreVariable, conf.level=0.99)

**Salida:**

Obtenemos los límites entre los que se encuentra el parámetro estimado con una probabilidad  $1 - \alpha$

#### 4. Contraste de hipótesis para la varianza (hallamos p-valor):

Escribimos en Rgui:

>n=474 (Introducimos el tamaño de muestra)

>sam.var=8.32 (Introducimos la varianza muestral)

>value=10 (Introducimos el valor de la varianza en la hipótesis nula)

En caso de que la hipótesis alternativa vaya con "<":

>pchisq((n-1)\*sam.var/value, n-1)

En caso de que la hipótesis alternativa vaya con ">":

>pchisq((n-1)\*sam.var/value, n-1, lower.tail=FALSE)

En caso de que la hipótesis alternativa vaya con "≠":

Habrá que sumar el p-valor para las dos regiones correspondientes.

**Salida:**

Nos devuelve el p-valor (si p-valor <  $\alpha$  rechazar  $H_0$ , si p-valor >  $\alpha$  aceptar  $H_0$ )

**Nota:** Para obtener los límites correctos de los intervalos de confianza, la hipótesis alternativa del contraste de hipótesis debe ir con "≠"

## INFERENCIA ESTADÍSTICA PARA DOS POBLACIONES INDEPENDIENTES.

### 1. Intervalo de confianza y contraste de hipótesis para la diferencia de medias (previamente deben tenerse apilados los datos de las variables implicadas):

Estadísticos → Medias → Test t para muestras independientes.

**Salida:**

**t:** valor del estadístico de contraste

**p-value:** p-valor (si p-valor <  $\alpha$  rechazar  $H_0$ , si p-valor >  $\alpha$  aceptar  $H_0$ )

**confidence interval:** intervalo de confianza (nos da los límites entre los que se encuentra la diferencia de medias estimada con una probabilidad  $1 - \alpha$ .. Si el intervalo incluye el 0, las medias podrán considerarse iguales con el nivel de confianza establecido)

**sample estimates:** medias muestrales de las dos variables implicadas.

### 2. Intervalo de confianza y contraste de hipótesis para el cociente de varianzas (previamente deben tenerse apilados los datos de las variables implicadas):

Estadísticos → Varianzas → Test F para dos varianzas.

**Salida:**

**F:** valor del estadístico de contraste

**p-value:** p-valor (si p-valor <  $\alpha$  rechazar  $H_0$ , si p-valor >  $\alpha$  aceptar  $H_0$ )

**confidence interval:** intervalo de confianza (nos da los límites entre los que se encuentra el cociente de varianzas estimado con una probabilidad  $1 - \alpha$ .. Si el intervalo incluye el 1, las varianzas podrán considerarse iguales con el nivel de confianza establecido)

**ratio of variances:** cociente entre varianzas muestrales.

- Intervalo de confianza y contraste de hipótesis para la diferencia de proporciones (para variables cualitativas dicotómicas):

*Estadísticos → Proporciones → Test de proporciones para dos muestras.*

**Nota:** Las categorías se ordenan por orden alfabético.

**Tipo de prueba:** Aproximación normal.

**Salida:**

**X-squared:** valor del estadístico de contraste

**p-value:** p-valor (si p-valor  $< \alpha$  rechazar  $H_0$ , si p-valor  $> \alpha$  aceptar  $H_0$ )

**confidence interval:** intervalo de confianza (nos da los límites entre los que se encuentra la diferencia de proporciones estimada con una probabilidad  $1 - \alpha$ . Si el intervalo incluye el 0, las proporciones podrán considerarse iguales con el nivel de confianza establecido)

**Sample estimates:** proporciones muestrales

## **INFERENCIA ESTADÍSTICA PARA DOS POBLACIONES RELACIONADAS O PAREADAS.**

- Intervalo de confianza y contraste de hipótesis para la diferencia de medias de muestras pareadas:

*Estadísticos → Medias → Test t para datos relacionados.*

**Salida:**

**t:** valor del estadístico de contraste

**p-value:** p-valor (si p-valor  $< \alpha$  rechazar  $H_0$ , si p-valor  $> \alpha$  aceptar  $H_0$ )

**confidence interval:** intervalo de confianza (nos da los límites entre los que se encuentra la diferencia de medias estimada con una probabilidad  $1 - \alpha$ . Si el intervalo incluye el 0, las medias podrán considerarse iguales con el nivel de confianza establecido)

**mean of the differences:** diferencia de las medias muestrales

**Nota:** Para obtener los límites correctos de los intervalos de confianza, la hipótesis alternativa del contraste de hipótesis debe ir con “ $\neq$ ”

## **TRANSFORMACIONES EN LOS DATOS.**

- Hacer que la primera categoría en una variable cualitativa no sea la primera por orden alfabético:

*Datos → Modificar variables del conjunto de datos activo → Reordenar niveles de factor*

*Esta opción ordena las categorías en orden alfabético inverso.*

*Datos → Modificar variables del conjunto de datos activo → Recodificar variables.*

**Ejemplo:** “mujer”=1, “hombre”=2

- Filtrar el conjunto de datos:

*Datos → Conjunto de datos activo → Filtrar el conjunto de datos activo (Se crea conjunto de datos nuevo)*

*Seleccionar una variable (por defecto se incluyen todas)*

**Ejemplo de expresión de filtrado:** minoría= “sí”

- Apilar datos:

*Datos → Conjunto de datos activo → Apilar variables del conjunto de datos activo*

# ESTADÍSTICA DESCRIPTIVA CON STATGRAPHICS CENTURION

## DATOS

1. Abrir datos:  
[Archivo](#) → [Abrir](#) → [Abrir Datos](#) → [Abrir datos desde StatGraphics](#)
2. Guardar datos (StatFolio y StatReporter):  
[Archivo](#) → [Guardar](#) → [Guardar Datos \(StatFolio o StatReporter\)](#)
3. Introducir nuevos datos:  
[Hacer doble click sobre una columna de la hoja de datos para editar antes de introducir datos \(introducir el tipo de datos\).](#)

## ESTADÍSTICA DESCRIPTIVA UNIVARIANTE Y BIVARIANTE

### Datos Numéricos

[Describir](#) → [Datos Numéricos](#)

**Si tenemos una sola variable** → [Análisis de una variable](#):

Tablas:        [Resumen estadístico](#) (Para hallar medidas estadísticas)  
                  [Tabla de frecuencias](#) (Para hallar una tabla de frecuencias con la variable agrupada en intervalos)  
                  [Percentiles](#) (Para hallar percentiles)

Gráficos:     [Histograma](#)  
                  [Gráfico de caja y bigotes](#)

**Si tenemos más de una variable** → [Análisis multivariado \(incluir solamente casos completos\)](#):

Tablas:        [Resumen estadístico](#) (Para hallar medidas estadísticas)  
                  [Correlaciones](#) (Para hallar el coeficiente de correlación de Pearson)  
                  [Correlaciones de rango](#) (Para hallar el coeficiente de correlación por rangos de Spearman)  
                  [Covarianzas](#) (Para hallar la covarianza)

Gráficos:     [Matriz de dispersión](#) (Para hallar los diagramas de dispersión)

**Nota:** Con el botón derecho del ratón, en opciones de ventana, se modifican siempre las opciones que vienen por defecto.

### Datos Cualitativos

[Describir](#) → [Datos Categóricos](#)

**Si tenemos una sola variable** → [Tabulación](#):

Tablas:        [Tabla de frecuencias](#) (Para hallar una tabla de frecuencias)

Gráficos:     [Diagrama de barras](#)  
                  [Diagrama de sectores](#)

## **Si tenemos dos variables → Tabulación cruzada:**

- Tablas: [Tabla de frecuencias](#) (Para hallar la tabla de doble entrada)  
[Pruebas de independencia](#) (Para hallar el coeficiente chi-cuadrado)  
[Resumen estadístico](#) (Ahí encontraremos el coeficiente de contingencia (entre otros))
- Gráficos: [Diagrama de barras](#)

Nota: Con el botón derecho del ratón, en opciones de ventana, se modifican siempre las opciones que vienen por defecto, por ejemplo, en la tabla de frecuencias podemos pedirle que imprima las frecuencias esperadas.

## **REGRESIÓN LINEAL**

### **Regresión simple**

[Relacionar](#) → [Un factor](#) → [Regresión Simple \(seleccionamos modelo de regresión\)](#)

- Tablas: [Resumen del análisis](#) (Para obtener la ecuación del modelo y coeficientes de correlación y bondad de ajuste)  
[Pronósticos](#) (Para hacer predicciones)  
[Comparación de modelos alternativos](#) (Para hacer una comparación de los distintos modelos lineales ordenados de mejor a peor en función de la bondad de ajuste)
- Gráficos: [Gráfico del modelo ajustado](#)

Nota 1: Con el botón derecho del ratón, en opciones de ventana, se modifican siempre las opciones que vienen por defecto, por ejemplo, en la tabla de predicciones, cambiaremos el valor de la X.

Nota 2: Con el botón derecho del ratón, en opciones de análisis, cambiamos entre distintos modelos.

### **Regresión polinomial (parabólica y cúbica)**

[Relacionar](#) → [Un factor](#) → [Regresión Polinomial \(orden 2 para modelo parabólico y 3 para cúbico\)](#)

- Tablas: [Resumen del análisis](#) (Para obtener la ecuación del modelo y bondad de ajuste)  
[Pronósticos](#) (Para hacer predicciones)
- Gráficos: [Gráfico del modelo ajustado](#)

Nota 1: Con el botón derecho del ratón, en opciones de ventana, se modifican siempre las opciones que vienen por defecto, por ejemplo, en la tabla de predicciones, cambiaremos el valor de la X.

Nota 2: Con el botón derecho del ratón, en opciones de análisis, cambiamos el orden del polinomio.

## Regresión múltiple

Relacionar → Varios factores → Regresión Múltiple (Mínimos cuadrados ordinarios)

Tablas: [Resumen del análisis](#) (Para obtener la ecuación del modelo y y bondad de ajuste)

Nota: Con el botón derecho del ratón, en opciones de análisis, podemos seleccionar “regresión paso a paso hacia atrás”, para ver qué variables conviene quitar del modelo.