
TEST DE COMPARACIÓN DE DOS POBLACIONES

Prueba t para muestras independientes

Test - t

H_0 : Las respuestas medias son iguales en los dos grupos

H_1 : Las respuestas medias son distintas en los dos grupos

Para poder usar el Test – t tenemos que asumir que en los dos grupos la variable que se estudia es NORMAL y que todos los datos son independientes

Prueba t para muestras independientes

Se quieren comparar las concentraciones de Chl a en la cabecera y en la cola de los embalses. Se toman medidas en distintos embalses.

Cola	18,6	21,7	24,5	12,9	12,5	34,1	40,2	21,6	34,5	11,8
Cabecera	10,1	20,4	18,5	5,9	12,2	33,9	37,8	16,7	29,2	11,6

H_0 : Las respuestas medias son las mismas en las dos zonas

H_1 : Son distintas

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias					95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Chl a	Se han asumido varianzas iguales	,040	,843	,775	18	,448	3,60700	4,65323	-6,16906	13,38306
	No se han asumido varianzas iguales			,775	17,928	,448	3,60700	4,65323	-6,17186	13,38586

El p-valor es 0.448

NO hemos encontrado evidencia para rechazar H_0 , no podemos afirmar que los niveles medios de Chl a son significativamente distintos en cola y en cabecera

Prueba t para muestras relacionadas

Se quieren comparar las concentraciones de Chl a en la cabecera y en la cola de los embalses. Se toman medidas en 10 embalses

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Cola	18,6	21,7	24,5	12,9	12,5	34,1	40,2	21,6	34,5	11,8
Cabecera	10,1	20,4	18,5	5,9	12,2	33,9	37,8	16,7	29,2	11,6

Cuando los datos están emparejados entre las dos muestras

Se utiliza el **test-t para muestras dependientes**

H_0 : Las respuestas medias son iguales en los dos grupos

H_1 : Las respuestas medias son distintas en los dos grupos

Para poder usar el Test – t tenemos que asumir que en los dos grupos la variable que se estudia es **NORMAL**

Prueba t para muestras relacionadas

Se quieren comparar las concentraciones de Chl a en la cabecera y en la cola de los embalses. Se toman medidas en 10 embalses.

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Cola	18,6	21,7	24,5	12,9	12,5	34,1	40,2	21,6	34,5	11,8
Cabecera	10,1	20,4	18,5	5,9	12,2	33,9	37,8	16,7	29,2	11,6

H_0 : Las respuestas medias son las mismas en las dos zonas

H_1 : Son distintas

Prueba de muestras relacionadas

		Diferencias relacionadas				t	gl	Sig. (bilateral)	
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior				Superior
Par 1	Cola - Cabecera	3,60700	3,10468	,98179	1,38604	5,82796	3,674	9	,005

El p-valor es 0.005

Hemos encontrado evidencia para rechazar H_0 , podemos afirmar que los resultados son significativamente distintos en la cabecera y en la cola

Comparación de dos poblaciones con datos NO normales

Tablas de contingencia y Test Chi-cuadrado

Los individuos de las dos poblaciones (columnas) se clasifican por la categoría de su respuesta (filas)

Opinión de 167 individuos sobre la reforma del Plan Hidrológico Nacional

	Menor de 25	Mayor de 25
A favor	8	28
En contra	12	44
NS/NC	22	53

Tabla de contingencia 3x2

Si la respuesta no es categórica se divide en clases y cada individuo se asigna a una de ellas

Test Chi-cuadrado

H₀: Las poblaciones (*columnas*) son **HOMOGÉNEAS** en sus respuestas (*filas*)

H₁: no son **HOMOGÉNEAS**

Para poder usar el Chi-cuadrado **NO** tenemos que asumir nada sobre la distribución de la variable, únicamente que todos los datos son independientes

Test Chi-cuadrado

H₀: Las poblaciones (*columnas*) son **HOMOGÉNEAS** en sus respuestas (*filas*)

H₁: no son **HOMOGÉNEAS**

Opinión de 167 individuos sobre la reforma del Plan Hidrológico Nacional

	Menor de 25	Mayor de 25
A favor	8	28
En contra	12	44
NS/NC	22	53

Tabla de contingencia 3x2

H₀: La opinión es independiente de la edad **H₁**: Depende de la edad

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	1,273	2	,529

El p-valor es 0.529

NO hemos encontrado evidencia para rechazar **H₀**, por tanto no podemos afirmar que la opinión dependa de la edad

Test **NO** PARAMÉTRICO de *Mann-Whitney*

H_0 : Las poblaciones son homogéneas

H_1 : Las poblaciones **NO** son homogéneas

Para poder usar el *Test de Mann Whitney* **NO** tenemos que asumir nada sobre la distribución de la variable, únicamente que todos los datos son independientes

Test **NO** PARAMÉTRICO de *Wilcoxon*

H_0 : Las poblaciones son homogéneas

H_1 : Las poblaciones **NO** son homogéneas

El *Test de Wilcoxon* se usa cuando los datos están emparejados

Se quieren comparar las concentraciones de Chl a en la cabecera y en la cola de los embalses. Se toman medidas en 10 embalses.

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Cola	18,6	21,7	24,5	12,9	12,5	34,1	40,2	21,6	34,5	11,8
Cabecera	10,1	20,4	18,5	5,9	12,2	33,9	37,8	16,7	29,2	11,6

H_0 : Las respuestas son homogéneas en las dos zonas

H_1 : No son homogéneas

Test **NO** PARAMÉTRICO de *Wilcoxon*

Estadísticos de contraste^b

	Cabecera - Cola
Z	-2,005 ^a
Sig. asintót. (bilateral)	,005

a. Basado en los rangos positivos.

b. Prueba de los rangos con signo de Wilcoxon

El p-valor es 0.005

Test de *Mann-Whitney*

Estadísticos de contraste

	Chl a
U de Mann-Whitney	34,000
W de Wilcoxon	89,000
Z	-1,209
Sig. asintót. (bilateral)	,226
Sig. exacta [2 ^a (Sig. unilateral)]	,247 ^a

a. No corregidos para los empates.

b. Variable de agrupación: Punto de muestreo

Hemos encontrado evidencia para rechazar H_0 , podemos afirmar que los resultados son significativamente distintos en la cabecera y en la cola

Comentarios:

- La diferencia entre usar un test – t (paramétrico) y un test no paramétrico está en las **hipótesis que asumimos**
- El test – t es de aplicación más restrictiva, pero **más potente**
- Existen alternativas a los test de MANN-WITHEY y Chi-cuadrado, por ejemplo los tests KRUSHAL-WALLIS tienen la misma H_0 y H_1 , por tanto sólo tenemos que **mirar al p-valor**
- Los test de MANN-WITHEY y de WILCOXON son test de rangos y sirven también para **comparar más de dos poblaciones**
- Existen otros test, que son no paramétricos, para “juzgar” si se cumple la hipótesis de normalidad

Test de bondad de ajuste a la normal

Kolmogorov-Smirnov (Lilliefors) , Shapiro-Wilk, ...

H_0 : La distribución de los datos es **NORMAL**

H_1 : La distribución de los datos no es normal

Pruebas de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
variable	,028	245	,200*	,996	245	,859

*. Este es un límite inferior de la significación verdadera.

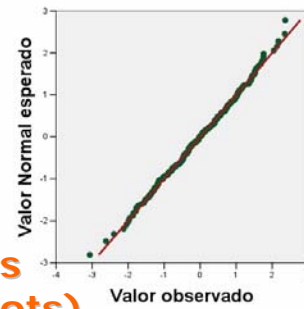
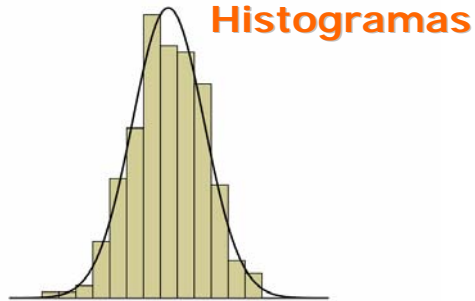
a. Corrección de la significación de Lilliefors

NO hemos encontrado evidencia en contra de la normalidad

El p-valor es mayor que 0.2

El p-valor es 0.859

Otras formas de ver la normalidad



Gráficos probabilísticos normales (Q-Q y P-P plots)

Estadística Descriptiva:

Resúmenes numéricos y gráficos de los datos

Resúmenes gráficos de una dimensión

Diagrama de barras
Pictograma o tarta } Cualitativos o discretos

Diagrama de tallo y hojas
Histograma
Barras de error
Diagrama de cajas - Boxplot } Cuantitativos

Resúmenes gráficos de varias dimensiones

Boxplot múltiple
Diagrama de dispersión
Matriz de diagramas de dispersión
Gráficos de estrellas o caras } Cuantitativos

Resúmenes numéricos de datos cuantitativos

Medidas que indican la posición de los datos

Media

Se suman todos los datos y se divide entre el número total

Mediana

Se ordenan los datos y se toma el que divide a la muestra en dos

Moda

Es el valor que aparece con más frecuencia en la muestra

Media recortada

Se calcula igual que la media pero quitando los datos extremos más pequeños y los más grandes, en un porcentaje fijo

Medidas de la variabilidad o dispersión de los datos

Recorrido

Diferencia entre el dato más grande y el más pequeño

Varianza

Promedio de todas las distancias de cada dato a la media (centro de la muestra)

Desviación típica

Raíz cuadrada de la varianza, con las mismas unidades que los datos

Rango Intercuartílico

Diferencia entre el tercer y el primer cuartil

MEDA

Mediana de las desviaciones absolutas a la mediana

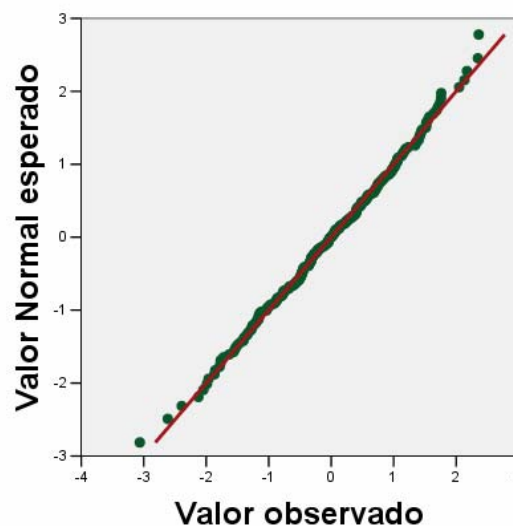
Gráficos probabilísticos normales

Se representan los valores esperados si la distribución es normal, frente a los valores observados

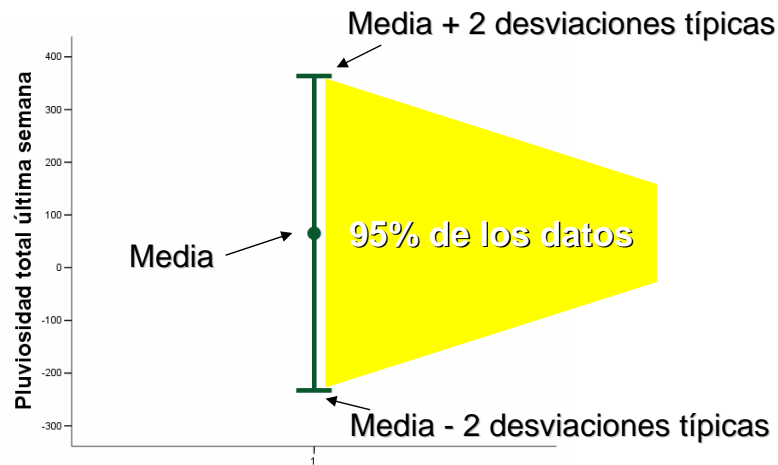
Si los datos son normales, los puntos tienen que situarse próximos a la diagonal

Los dos más habituales son el **Q-Q plot** y el **P-P plot**

Se pueden usar con otras distribuciones



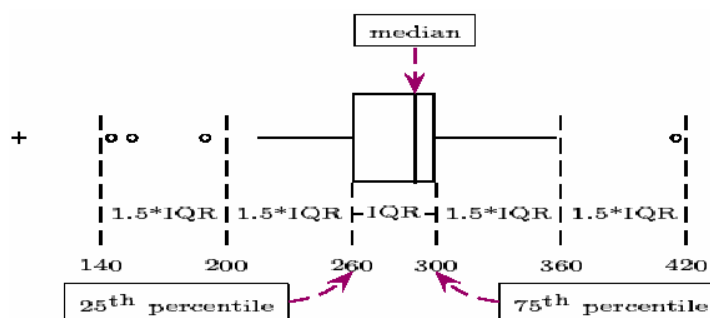
Barras de error



- La desviación típica es muy sensible a los datos atípicos
- Siempre es un gráfico simétrico
- Es una buena herramienta cuando los datos son normales

Diagrama de cajas o Boxplot

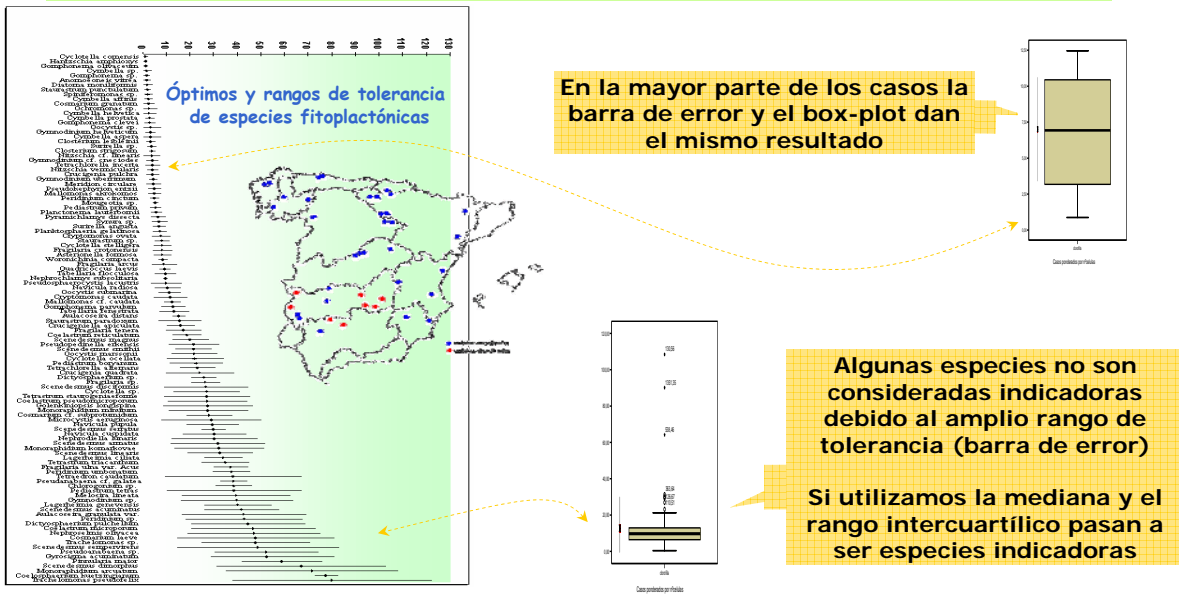
1. Ordenar la muestra
2. Calcular la mediana, el primer y el tercer cuartil
3. Calcular el rango y el rango intercuartílico (IQR)



¿ Box-plot o barra de error ?

UTILIZACIÓN DEL FITOPLACTON COMO INDICADOR BIOLÓGICO PARA LA EVALUACIÓN DE LA EUTROFIZACIÓN EN LOS EMBALSES ESPAÑOLES

C. NUÑO, C. DE HOYOS, A. JUSTEL



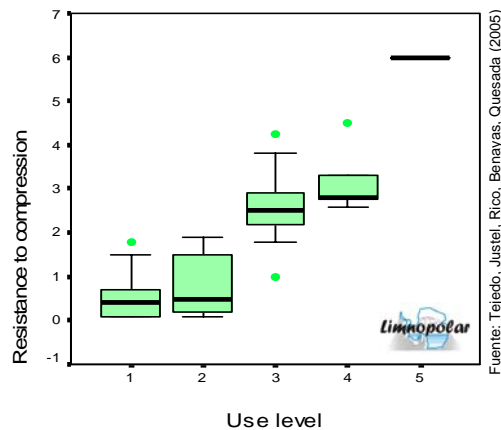
Gráficos para conjuntos de datos

Comparación de conjuntos de datos

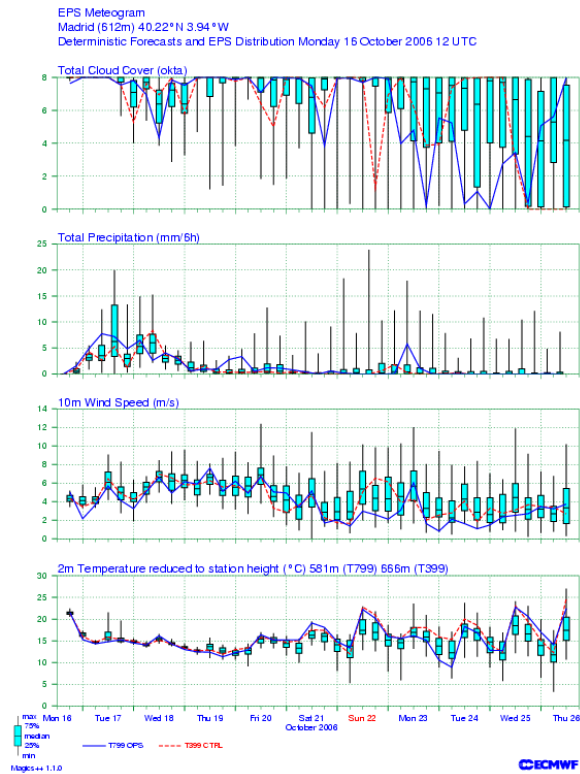
Los datos corresponden a la misma medición en dos clases de individuos

¿Son los tipos (x_1^1, \dots, x_n^1) e (x_1^2, \dots, x_m^2) distintos?

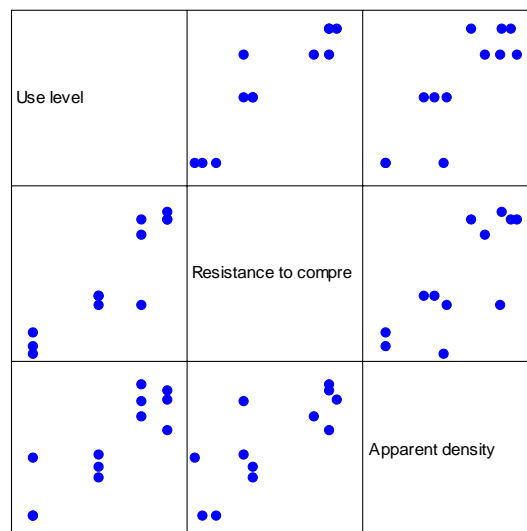
Boxplot múltiple



Boxplot múltiples para mostrar la información de las predicciones meteorológicas



Matriz de diagramas de dispersión



Fuente: Limnopolar, 2005

✚ **Utilizando cualquier software estadístico es sencillo**

✚ **Con EXCEL, también es posible...**

El procedimiento básico "insertar funciones" sirve para calcular la media, la varianza, etc.

Los gráficos habituales se pueden utilizar para hacer diagramas de barras, pictogramas, cajas y dispersión

Para otros procedimientos más completos y la representación de histogramas hay que activar una herramienta

Herramientas ↔ Complementos ↔ Herramientas para análisis ↔ Aceptar

Con la opción activada se accede al menú estadístico

Herramientas ↔ Análisis de datos

