Análisis Estadístico de Datos Climáticos

Pruebas de Hipótesis

(Wilks, cap. 5)

2015

PRUEBAS DE HIPÓTESIS
(o pruebas de significación)

Objetivo: A partir del análisis de una muestra de datos, decidir si se rechaza o no una hipótesis, llamada *hipótesis nula*.
La evidencia se juzga en el contexto de un modelo estadístico, de modo que es posible cuantificar, o por lo menos acotar, el riesgo de rechazar dicha hipótesis cuando ésta es cierta.
No se debe confundir la decisión de rechazar o no una hipótesis con un cierto nivel de riesgo con la determinación de su falsedad o veracidad.

TIPOS DE PRUEBAS

Las pruebas pueden apuntar a decidir sobre distintos aspectos, p. ej:

- **Establecer un valor ó un intervalo de valores para los parámetros de una variable**
  - Asociada a la construcción de Intervalos de confianza
  - Ejemplo de hipótesis: La media de una variable es 10
- **Establecer la igualdad de las distribuciones de dos ó mas variables**
  - Ejemplo de hipótesis: La media de dos poblaciones gausianas son iguales con igual variancia
- **Determinar la forma de la distribución de la variable**
  - Pruebas especificas para establecer el tipo de distribución de una variable
  - Ejemplo de hipótesis: La distribución de una variable es gausiana

PRUEBAS PARAMETRICAS Y NO PARAMETRICAS

Se denominan pruebas **paramétricas** aquellas que presuponen una distribución de probabilidad dada para los datos.

Se denominan pruebas **no paramétricas** aquellas que no presuponen una distribución de probabilidad para los datos, por ello se conocen también como pruebas de distribución libre.

Ejemplo de prueba de hipótesis (Wilks, p. 135)

En un aviso publicitario, un establecimiento turístico afirma que en ese lugar, en invierno, de cada 7 días, hay 6 días de cielo completamente despejado. Se desea tratar de verificar si tal afirmación es razonablemente cierta. Para ello, se consigue información de n = 25 días (no consecutivos ! ¿por qué?), resultando que hay 15 días de cielo despejado.
¿Qué conclusión se puede sacar de estos datos?

1) **El primer paso de la prueba de hipótesis es elegir si la prueba será paramétrica o no, y elegir un "estadístico de la prueba" que será una función de los datos.**

El problema responde a una distribución binomial porque: hay n=25 ensayos independientes con 2 resultados posibles y, suponiendo que el tiempo en que fueron relevados los datos no fue muy largo, es razonable suponer que la probabilidad de éxito (día despejado), p, es la misma cada día.

(El test es paramétrico)

El estadístico de la prueba será la variable aleatoria "número de aciertos" que en el ejemplo, toma el valor X = 15.

Ejemplo de prueba de hipótesis (cont.)

2) **Enunciar una "hipótesis nula" (H_0) , que es una referencia a contrastar con el valor del estadístico obtenido.**

   En nuestro ejemplo, elegimos como H_0 que la afirmación realizada es cierta, o sea que $p = 6/7 = 0.857$.

   Dada la naturaleza del problema, se entiende que, si la afirmación es falsa, el verdadero valor de p será <u>menor</u> que 0.857.

3) **Elegir una hipótesis alternativa H_A que, de acuerdo a lo anterior será: $p < 0.857$. En este caso, se dice que la prueba es de un extremo (one-sided), por oposición a $p \neq 0.857$, que sería de dos extremos (two-sided).**

   En general, dada una hipótesis nula, puede haber varias hipótesis alternativas para elegir.

4) **Hallar la "distribución nula" de X, es decir la distribución de X dado que H_0 es cierta, o sea suponiendo que $p = 0.857$. En esta hipótesis, X sigue una distribución binomial de parámetros $n = 25$ y $p = 0.857$.**

Ejemplo de prueba de hipótesis (cont.)

5) Comparar qué tan probable es el valor obtenido de X en la muestra (15) suponiendo que la hipótesis nula es cierta. Para ello, se calcula la probabilidad (de acuerdo a la distribución nula) del estadístico de la prueba observado *y de todos los otros resultados que sean por lo menos tan desfavorables para la hipótesis nula.*

Es decir, en nuestro caso, calculamos la probabilidad de que $X \leq 15$ en la hipótesis que $n = 25$ y $p = 0.857$, o sea:

$$\sum_{x=0}^{x=15} b(x;25;\ 0.857)$$

o en forma equivalente,

$$\Pr\{X \leq 15\} = \sum_{x=0}^{15} \binom{25}{x} 0.857^{x} (1 - 0.857)^{25-x}$$

El resultado (p-valor) es 0.0015.

En Matlab: binocdf(15,25,6/7);

¿Qué conclusión podemos sacar?

Ejemplo de prueba de hipótesis (cont.)

Hemos llegado a que la probabilidad de tener 15 (o menos) días despejados, en la hipótesis nula (p=0.857), es 0.0015.

En general, se toma (arbitrariamente) 0.05 como valor de referencia para comparar con el p-valor (también se toma 0.1 o 0.01). Es decir, se rechaza H_0 si el p-valor es menor que 0.05.

La elección de ese valor (0.05 u otro) se hace de antemano.

Como $0.0015 < 0.05$, decimos que se rechaza la hipótesis nula con un nivel de significación o significancia de 0.05. (En este caso, también la rechazaríamos con un nivel de significancia de 0.1 o 0.01.). También se dice que se rechaza H_0 con un nivel de confianza de 0.95, (o 0.90 o 0.99).

Naturalmente, ese nivel de significancia debe ser "pequeño", de modo que rechazamos H_0 si la probabilidad de lo ocurrido (el valor de X o menor) es baja.

En resumen, los datos obtenidos dan fuerte evidencia de que $p < 0.857$.

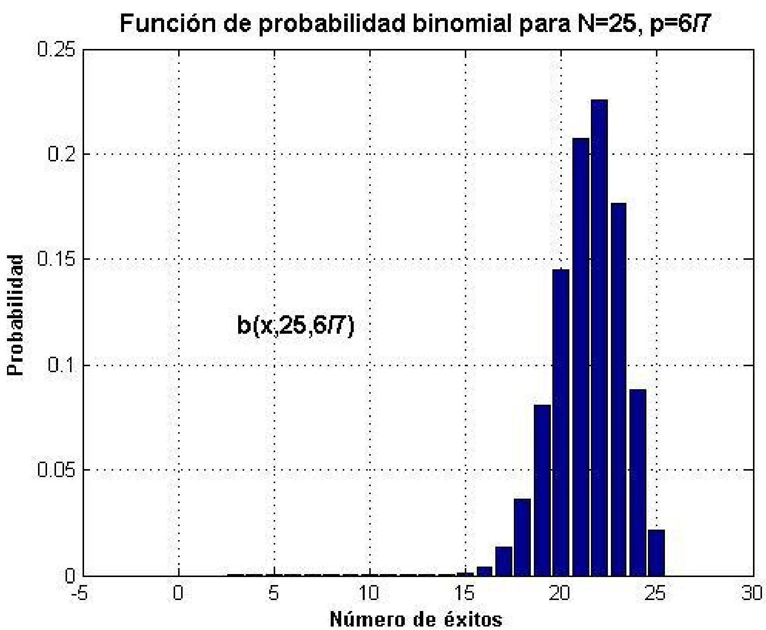
Observar que el p-valor hallado (0.0015) es la probabilidad de rechazar la hipótesis nula dado que es cierta, y nos da una cuantificación del riesgo de rechazar H_0 cuando es cierta.

Observar que, para decidir si rechazar o no H_0, no es suficiente hacer la simple comparación 15/25 < 6/7, ya que de esa forma no estamos teniendo en cuenta el azar que determina que X tenga variabilidad.

Por ej., si hubiéramos obtenido X = 18, se cumple también que 18/25 < 6/7, pero el p-valor es $P(X \leq 18)$ = binocdf(18, 25, 6/7) = 0.056 > 0.05, y no habríamos rechazado la hipótesis nula con un nivel de 0.05.

(Idem con X = 19, 20 y 21; para este el p-valor da 0.488)

La región de rechazo al nivel 0.05 es $X \leq 17$.
$P(X \leq 17) = 0.019$



Densidad de probabilidad de H_0
(n=25, p = 6/7 = 0.857)

HIPOTESIS NULA y ALTERNATIVA

- **Hipótesis nula:** es una hipótesis en principio "aceptada", cuyo grado de validez se contrasta con el resultado de la prueba. Se especifica de una forma exacta: p. ej:

$H_0 : \theta = \theta_0$

- **Hipótesis alternativa:** se especifica de manera más general, p. ej :

$H_A: \theta \neq \theta_0$
$H_A: \theta > \theta_0$
$H_A: \theta < \theta_0$.

TIPOS DE ERROR

Rechazar una hipótesis no significa que ésta sea falsa, como tampoco el no rechazarla significa que sea verdadera. La decisión tomada no está libre de error.

Error I: Rechazar una hipótesis que es verdadera. (Rechazamos una hipótesis cuando no debiera ser rechazada).

Error II: No rechazar una hipótesis que es falsa (No rechazamos una hipótesis que debiera ser rechazada).

NIVEL DE SIGNIFICACION y POTENCIA DE LA PRUEBA

α es la **Probabilidad de cometer un Error tipo I. Se llama** ***nivel de significación, nivel de rechazo o nivel de la prueba.***
El nivel de la prueba se elige antes de realizar la prueba. Luego, el p-valor obtenido se compara con α.

β es la **probabilidad de cometer un Error tipo II**
(1 - β recibe el nombre de "potencia de la prueba": es la probabilidad de rechazar una hipótesis que es falsa)

Es deseable que estas dos probabilidades de error (α y β) sean pequeñas, pero eso no es siempre posible.

CUADRO DE DECISIONES Y TIPOS DE ERRORES

		Estado de la Naturaleza	
		H_0 verdadera	H_0 falsa
Decisión	No rechazo H_0	Acierto $1 - \alpha$ Nivel de confianza	Error Tipo II β
	Rechazo H_0	Error Tipo I α Nivel de significación	Acierto $1 - \beta$ Potencia de prueba

Ocurre que la disminución de un tipo de error va acompañada del aumento del otro.

Se pueden disminuir ambos si se aumenta el tamaño de la muestra, lo que no siempre es posible.

ERRORES TIPO I Y II
Y NIVEL DE SIGNIFICACION

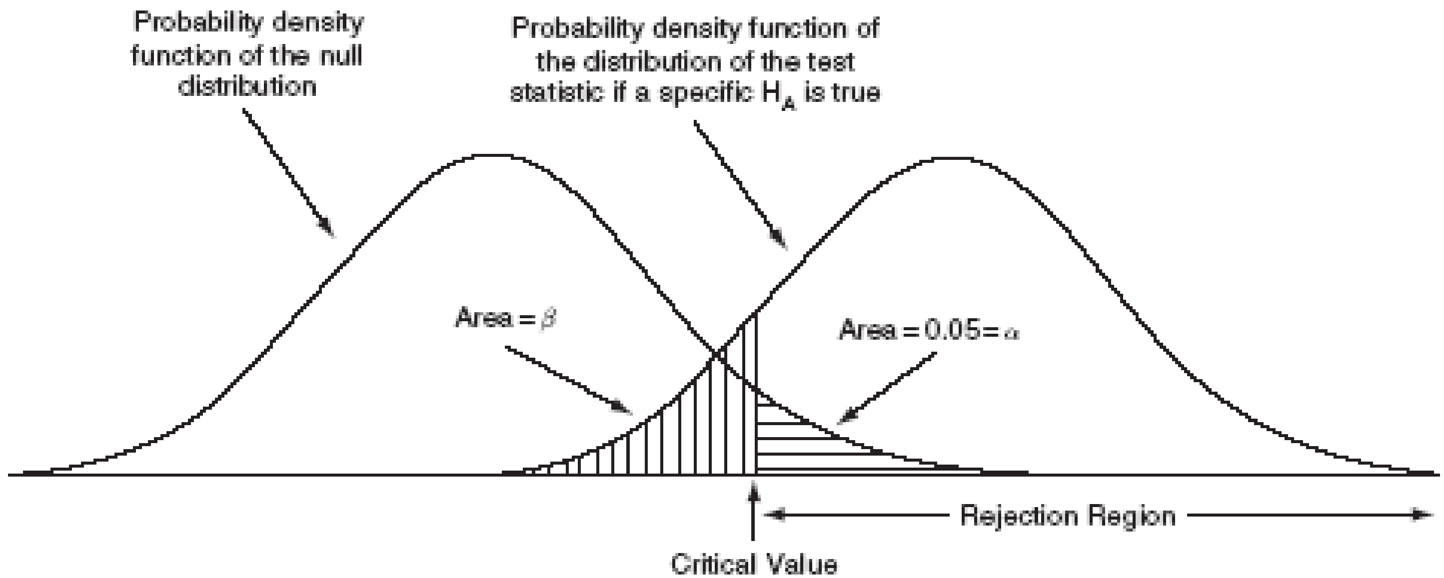


FIGURE 5.1 Illustration of the relationship of the rejection level, α, corresponding to the probability of a Type I error (horizontal hatching); and the probability of a Type II error, β (vertical hatching); for a test conducted at the 5% level. The horizontal axis represents possible values of the test statistic. Decreasing the probability of a Type I error necessarily increases the probability of a Type II error, and vice versa.

El nivel α de la prueba puede prescribirse, pero la probabilidad de error Tipo II, (β), en general no, porque la hipótesis alternativa H_A usualmente es la unión de muchas hipótesis alternativas (ej. $p < p_0$). En la figura se muestra una de las infinitas posibles hipótesis alternativas. A veces es útil estudiar cómo varía la potencia ($1 - \beta$) para un conjunto de posibles H_A .

NIVEL DE SIGNIFICACION Y NIVEL DE CONFIANZA

En la práctica, es frecuente un nivel de significación de 0,05 ó 0,95 de ***nivel de confianza.***

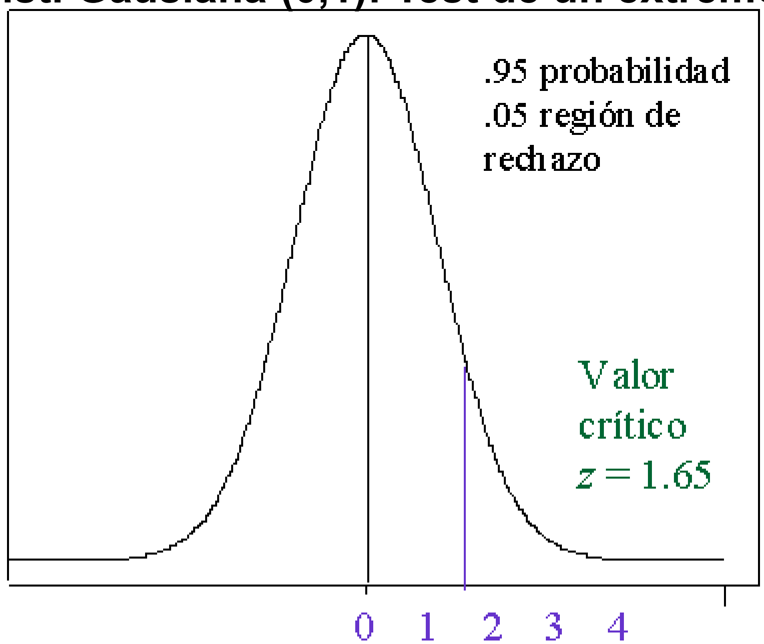
Si por ejemplo se elige el nivel de significación 0,05 (ó 5%), entonces hay unas cinco (05) oportunidades entre 100 de rechazar la hipótesis cuando no debiera haberse rechazado; es decir, tenemos un 95% de ***confianza*** de que hemos adoptado la decisión correcta.

El nivel 0.1 (0.90) se suele llamar "marginalmente significativo"
El nivel 0.05 (0.95) " " " "significativo"
El nivel 0.01 (0.99) " " " "muy significativo"

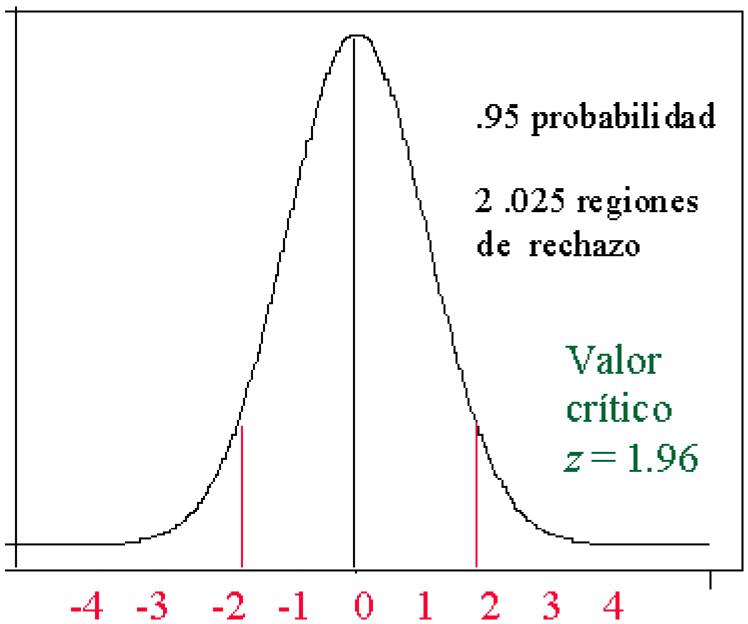
Pruebas de Uno y Dos Extremos.

Cuando consideramos ambos lados de la media lo llamamos prueba de dos extremos o de dos colas.
Con frecuencia no obstante, estaremos interesados tan sólo en valores extremos a un lado de la media (o sea, en uno de los extremos de la distribución), tal como sucede cuando se contrasta la hipótesis de que un proceso es mejor que otro. Tales contrastes se llaman unilaterales, o de un extremo. En tales situaciones, la región crítica o de rechazo es una región situada a un lado de la distribución, con área igual al nivel de significación.
Los tests de dos extremos son apropiados cuando valores del estadístico tanto en un extremo como en el otro son desfavorables para la hipótesis nula. Es el caso en que H_A es: "H_0 no es cierta"

Dist. Gausiana (0,1): Test de un extremo



Dist.gausiana (0,1): Test de 2 extremos



INTERVALOS DE CONFIANZA

El problema de determinar intervalos de confianza para un parámetro se puede interpretar como el problema inverso de realizar una prueba de hipótesis sobre un parámetro.

Las pruebas de hipótesis evalúan probabilidades asociadas al valor observado de un estadístico (o variable aleatoria), suponiendo que es cierta una determinada hipótesis nula. Reciprocamente, los intervalos de confianza se construyen encontrando valores del estadístico que no estén en la región crítica, o de rechazo.

En el contexto de estimar un parámetro, un ***intervalo de confianza es un rango de valores (calculado a partir de una muestra) en el cual se encuentra el verdadero valor del parámetro, con una probabilidad determinada*****.**

La probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo construido se denomina **nivel de confianza**, y se denota 1- α. La probabilidad de equivocarnos se llama **nivel de significación** y se simboliza α.. Generalmente se construyen intervalos con confianza
1- α.= 95% (o significación α = 5%).

El intervalo de confianza es aleatorio pues depende de la muestra. El verdadero valor del parámetro no es aleatorio; es fijo, aunque suele ser desconocido.

Fijado el nivel de confianza (1 – α), cuanto mayor es el intervalo de confianza, mayor es la incertidumbre sobre el verdadero valor del parámetro.

Ejemplo: intervalo de confianza para un promedio (distribución gausiana)

Supongamos que tenemos una muestra de valores de n variables aleatorias independientes: X_1, X_2,…X_n , todas con distribución N(μ, σ^2), donde σ se supone conocida y μ es desconocida.

El promedio $\bar{X} = \dfrac{X_1 + \ldots + X_n}{n}$ **es una variable aleatoria gausiana que tiene media μ y**

desviación estándar $\dfrac{\sigma}{\sqrt{n}}$

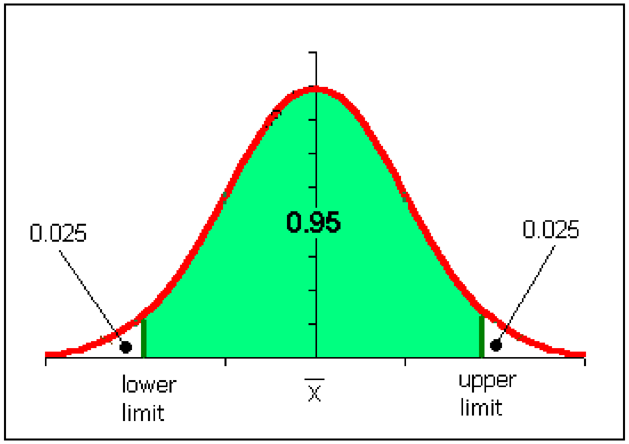
Se busca construir un intervalo de confianza al 95% de μ, es decir un intervalo que contenga a μ, con probabilidad 0.95.

La variable aleatoria $\dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ **tiene distribución N(0,1).**

Entonces queremos que $-1.96 \leq \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96$

Despejando μ se tiene: $\bar{X} - 1.96\,\dfrac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\,\dfrac{\sigma}{\sqrt{n}}$

que es el intervalo de confianza buscado.



Es un hecho importante que la longitud del intervalo de confianza disminuye cuando aumenta el tamaño de la muestra

Comparación de medias de dos muestras

Prueba t de Student

Se utiliza para hacer una prueba de hipótesis de la diferencia de las medias de dos poblaciones cuando las desviaciones estándar de ambas poblaciones no son conocidas.

Se considera una muestra de cada población, de tamaños $\mathbf{n_X}$ y $\mathbf{n_Y}$ respectivamente:
$\{x_1, x_2, \ldots\ldots, x_{nX}\}$ y $\{y_1, y_2, \ldots\ldots\ldots, y_{nY}\}$

Se debe cumplir que:

- **todas las observaciones de ambas poblaciones son independientes entre sí**
- **ambas poblaciones siguen una distribución gausiana** (a veces se invoca el teorema del límite central para aceptar esta hipótesis, o se transforma la variable).
- **ambas distribuciones tienen <u>la misma varianza</u> (aunque desconocida)**

La hipótesis nula H_0 es que ambas medias son iguales.

$\widehat{\mu}_X$ y $\widehat{\mu}_Y$ son las medias muestrales

y $$S_p^2 = \frac{\sum_{i=1}^{n_X}(\mathbf{x}_i - \widehat{\mu}_X)^2 + \sum_{i=1}^{n_Y}(\mathbf{y}_i - \widehat{\mu}_Y)^2}{n_X + n_Y - 2}$$ es una estimación combinada de la desviación estándar común.

Se demuestra que bajo esas hipótesis (en particular se debe cumplir H_0), el estadístico

$$t = \frac{\widehat{\mu}_X - \widehat{\mu}_Y}{S_p\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \quad (1)$$

tiene una distribución t de Student con $n_X + n_Y - 2$ grados de libertad

Una variable aleatoria T tiene una distribución t de Student con <u>m</u> grados de libertad si

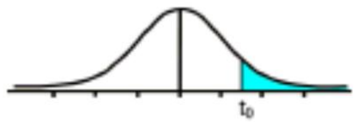
$$\mathbf{T} = \frac{\mathbf{A}}{\sqrt{\mathbf{B}/m}}$$

con A ~ N(0,1) y B ~ χ_m^2, siendo

A y B independientes.

(Si la varianza es conocida, se sustituye su valor σ por S_P en (1); el estadístico resultante tiene distribución N (0,1))

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3007	1.6794	2.0141	2.4121	2.6896
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800

Tabla t de Student

Matlab: tinv (p,v) da la inversa de la distribución acumulada con v grados de libertad; no es lo mismo que la gráfica!! (hay que cambiar p por 1-p)

Hay otras pruebas de Student que se pueden usar cuando los datos están apareados o cuando dentro de de las poblaciones existe autocorrelación.

Más adelante veremos cómo usar la distribución t de Student para determinar si la correlación entre 2 series es significativa a un cierto nivel.

El test de Student no es resistente.

En el caso en que las varianzas de ambas poblaciones sean diferentes, no se obtiene directamente una distribución de Student.

Se encuentra que el estimador:

$$t = \frac{\widehat{\mu}_X - \widehat{\mu}_Y}{\sqrt{S_X^2/n_X + S_Y^2/n_Y}}$$

tiene *aproximadamente* una distribución de Student en que el número de grados de libertad se estima a partir de los datos con:

$$df = \frac{(S_X^2/n_X + S_Y^2/n_Y)^2}{\frac{(S_X^2/n_X)^2}{n_X - 1} + \frac{(S_Y^2/n_Y)^2}{n_Y - 1}}$$

PRUEBAS DE HIPOTESIS NO PARAMETRICAS

Se denominan pruebas **no paramétricas** aquellas que **no presuponen** una distribución de probabilidad para los datos, por ello se conocen también como de distribución libre (*distribution free*).

En la mayor parte de ellas los resultados estadísticos se derivan únicamente a partir de procedimientos de ordenación y recuento, por lo que su base lógica es de fácil comprensión. Cuando trabajamos con muestras pequeñas ($n \leq 20$) en las que se desconoce si es válido suponer la normalidad de los datos, conviene utilizar pruebas no paramétricas, al menos para corroborar los resultados obtenidos a partir de la utilización de la teoría basada en la normal.

En estos casos se emplea como parámetro de centralización la **mediana**, que es aquel punto para el que el valor de X está el 50% de las veces por debajo y el 50% por encima.

PRUEBAS NO PARAMETRICAS

Veremos algunas pruebas no paramétricas, que en buena medida son paralelas a las versiones paramétricas (t Student, F, etc.):

Caso de dos grupos independientes
Prueba de Mann-Whitney-----(paralela a la t de grupos independientes)

Caso de "a" grupos independientes
Prueba de Kruskal-Wallis-----(paralela a la F unifactorial entre-sujetos)

Prueba de Mann-Whitney
(comparación de dos grupos independientes)

Este procedimiento es una buena alternativa cuando no se puede utilizar la prueba t de Student, por no cumplir con los requisitos que esta prueba exige.

La hipótesis nula es que ambas muestras provienen de una misma población o distribución.

n_1 = tamaño de la muestra del grupo 1.
n_2 = tamaño de la muestra del grupo 2.

Se ordenan todos los datos del menor al mayor valor, y se le asignan rangos.

La idea es que si ambas muestras provienen de la misma distribución, los rangos de uno y otro grupo van a tener una magnitud global similar (relativa al tamaño de las muestras).

Se definen:

R_1 = sumatoria de los rangos del grupo 1.
R_2 = sumatoria de los rangos del grupo 2.

• **Si n_1 y n_2 son menores que 20, (muestras pequeñas) se definen los valores estadísticos de Mann-Whitney:**

$$U_1 = R_1 - \frac{n_1}{2}(n_1 + 1)$$

$$U_2 = R_2 - \frac{n_2}{2}(n_2 + 1)$$

Se elige el más pequeño de ambos para comparar con los valores críticos de U de la tabla de probabilidades asociadas al test de Mann-Whitney

• Si n_1 y n_2 son mayores que 20 (muestras grandes), los estadísticos U se distribuye en forma aproximadamente gausiana y se calcula el valor $Z_{emp} \sim N\ (0,1)$

$$Z_{emp} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}$$

Luego se decide si se acepta o rechaza la hipótesis, de acuerdo al nivel de significación elegido.

En Matlab, se usa el comando ranksum:

```
[p, h] = ranksum(x,y)
```

x e y son los vectores de datos; si h=1, se rechaza la hipótesis nula al 5%; si h=0, se acepta la hipótesis nula al 5%; p da el p-valor

Prueba de Kruskal-Wallis

(comparación de "a" grupos independientes)

La prueba de Kruskal-Wallis, es una alternativa a la prueba F del análisis de varianza para diseños de clasificación simple. En este caso se comparan varios grupos. Es una generalización del test de Mann-Whitney.

Pasos:

1. pasar todas las puntuaciones a rangos (conjuntamente en los "a" grupos)

2. computar la suma de los rangos en cada grupo (son las R_j)

Estadístico de contraste:

$$H = \frac{12}{N(N+1)} \left(\Sigma \frac{R_j^2}{n_j} \right) - 3\,(N+1)$$

Si la Hipótesis nula es cierta (es decir, que no haya diferencias entre los grupos), y hay más de 5 datos en cada grupo, H se distribuye según la distribución Chi-cuadrado con a-1 grados de libertad

Se puede aplicar esta prueba cuando no se cumplan los supuestos de homogeneidad de varianzas ni el de normalidad del ANOVA unifactorial entre sujetos.

Matlab: kruskalwallis.m

Test de Bondad de ajuste Chi Cuadrado ($\chi 2$)

El Test Chi - Cuadrado puede utilizarse para determinar la calidad del ajuste mediante distribuciones teóricas (como la distribución normal o la binomial) de distribuciones empíricas (o sea las obtenidas de los datos de la muestra).

La **prueba de Chi-cuadrado** mide la discrepancia entre una distribución observada y otra teórica (bondad de ajuste), comparando esencialmente un histograma con la PDF teórica, indicando en la hipótesis nula, la probabilidad de obtener un resultado "igual o peor" que el observado.

La hipótesis nula es que los datos provienen de una distribución teórica dada.

Los datos se ordenan en forma creciente y se dividen en clases. Las clases no tienen por qué tener la misma amplitud ni la misma cantidad de datos, pero sí se recomienda que haya por lo menos 5 datos en cada clase.

La fórmula que da el estadístico es la siguiente: $$\chi^2 = \sum_{clases} \frac{(\text{Nº observado}_i - \text{Nº esperado}_i)^2}{\text{Nº esperado}_i}$$

Cuanto mayor sea el valor de χ^2, menos verosímil es que la hipótesis sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de chi-cuadrado, más ajustadas están ambas distribuciones.

Los grados de libertad vienen dados por : g.l. = nº de clases – nº de parámetros de la distribución -1
Se compara el valor obtenido de $\chi 2$ con el de la tabla correspondiente para el nivel de significación elegido.

Test de Bondad de ajuste Chi Cuadrado (Ejemplo)

Sean 1000 valores de temperatura media horaria que dividimos en 6 clases:

38 horas han tenido una temperatura media de 0 °C
144 horas han tenido una temperatura media de 1 °C
342 horas han tenido una temperatura media de 2 °C
287 horas han tenido una temperatura media de 3 °C
164 horas han tenido una temperatura media de 4 °C
25 horas han tenido una temperatura media de 5 °C

Queremos investigar si los datos se ajustan razonablemente con una distribución gausiana

$\mu = 2.47$

$\sigma = 1.11$

Centro del intervalo (ºC)	Límites del intervalo	Variable estandarizada	Area bajo la curva en el intervalo	#Esperado= 1000*Area	#Observado	$\chi 2$
0	-0.5 0.5	-2.675 -1.775	0.03415	34.15	38	0.434
1	0.5 1.5	-1.775 -0.874	0.15318	153.18	144	0.550
2	1.5 2.5	-0.874 0.027	0.31972	319.72	342	1.553
3	2.5 3.5	0.027 0.928	0.31248	312.48	287	2.078
4	3.5 4.5	0.928 1.829	0.14304	143.04	164	3.071
5	4.5 5.5	1.829 2.730	0.03048	30.48	25	0.985
Total						8.671

El nº de g. l. es 6 -2 -1 = 3

Según la tabla, para 3 grados de libertad, la hip. nula se rechaza al 5% (pero no al 1%).

Tabla de Chi Cuadrado

TABLE B.3 Right-tail quantiles of the Chi-square distribution. For large ν, the Chi-square distribution is approximately Gaussian, with mean ν and variance 2ν.

	Cumulative Probability					
ν	0.50	0.90	0.95	0.99	0.999	0.9999
1	0.455	2.706	3.841	6.635	10.828	15.137
2	1.386	4.605	5.991	9.210	13.816	18.421
3	2.366	6.251	7.815	11.345	16.266	21.108
4	3.357	7.779	9.488	13.277	18.467	23.512
5	4.351	9.236	11.070	15.086	20.515	25.745
6	5.348	10.645	12.592	16.812	22.458	27.855
7	6.346	12.017	14.067	18.475	24.322	29.878
8	7.344	13.362	15.507	20.090	26.124	31.827
9	8.343	14.684	16.919	21.666	27.877	33.719
10	9.342	15.987	18.307	23.209	29.588	35.563
11	10.341	17.275	19.675	24.725	31.264	37.366
12	11.340	18.549	21.026	26.217	32.910	39.134
13	12.340	19.812	22.362	27.688	34.528	40.871
14	13.339	21.064	23.685	29.141	36.123	42.578
15	14.339	22.307	24.996	30.578	37.697	44.262
16	15.338	23.542	26.296	32.000	39.252	45.925

La hipótesis nula se rechaza al nivel 0.05 (pero no al nivel 0.01).
Matlab: chi2gof.m (en versiones "nuevas")

Tests de Bondad de ajuste de Kolmogorov-Smirnov (KS) y Lilliefors (L)

Mientras el test χ^2 compara histograma con PDF, los tests de KS y L comparan las CDFs empírica y teórica.

La hipótesis nula H_0 es que los datos observados fueron tomados de una distribución dada.

Ambos tests usan el estadístico $D_n = \max_x |F_n(x) - F(x)|$.

siendo $F_n(x)$ la CDF empírica estimada como $F_n(x_{(i)}) = i/n$ para el i-ésimo estadístico de orden, y $F(x)$ la CDF teórica

Si D_n es suficientemente grande, dependiendo del nivel de la prueba, H_0 se rechaza

El test KS puede usarse si los parámetros de la distribución nula NO fueron obtenidos a partir de los datos de la muestra.

De lo contrario, debe usarse el test L.

Matlab: kstest.m y lillietest.m (sólo para algunas distribuciones)