COMENTARIOS SOBRE LOS SUPUESTOS.

1. Los tests basados en la ditribución *t* suponen que los datos de cada muestra provienen de una distribución Normal. Generalmente no se conoce la distribución de las variables, pero si el histograma de los datos es aproximadamente simétrico y suave y el box-plot no muestra datos atípicos, la distribución Normal será una buena aproximación.
2. Apartamientos moderados de la distribución normal no modifican fuertemente las conclusiones de los tests *t*.
3. El test t para dos muestras independientes es extremadamente sensible a la heterogeneidad de varianzas cuando los tamaños de muestra son muy diferentes.
4. Si falla el supuesto de normalidad pero ambas muestras tienen tamaño suficientemente grande entonces, por el Teorema Central del Límite, se pueden utilizar los tests con nivel aproximado basados en la distribución Normal.
5. El test F para igualdad de varianzas es MUY SENSIBLE a apartamientos de la hipótesis de normalidad, mucho más que el test t. Por lo tanto, cuando tenemos dudas respecto de la normalidad de los datos y los mismos presentan evidencias de diferencias en las varianzas (por ejemplo, una desviación estándar es más que el doble de la otra) y los tamaños de muestras son muy diferentes, es preferible usar el test t aproximado para varianzas distintas.
6. El supuesto de que las observaciones son *independientes* es muy difícil de verificar. Depende del diseño de la investigación, y del modo en que se recolectaron los datos. Cuando este supuesto no se cumple no son válidos los métodos que hemos presentado ni los que vamos a presentar.

   Un modo simple de chequear un tipo de dependencia asociada a la secuencia de observaciones es el siguiente:

   - Calcular el residuo de cada dato: residuo = dato – media del grupo.
   - Graficar los residuos versus el orden en que fueron obtenidos los datos.

   Si fueran independientes, el gráfico no debería presentar estructura ni tendencia. En el ejercicio 2 de la práctica 1 (temperatura de sublimación del iridio y del rodio) vimos que la estructura de los datos cambiaba con el orden por no haberse estabilizado la reacción. Sólo eliminando las primeras observaciones podíamos suponer que los errores eran independientes.

¿Qué hacer cuando los datos se apartan fuertemente de la distribución normal y la cantidad de datos es pequeña? La alternativa es usar tests no paramétricos.

13. ALTERNATIVAS NO PARAMÉTRICAS

El modelo de Normalidad de los datos es un modelo paramétrico pues está unívocamente especificado mediante el conocimiento de los parámetros, μ y σ. También lo es un modelo que establezca que los datos provienen de una distribución exponencial o cualquier otra distribución que quede completamente especificada con el conocimiento de una cantidad finita de parámetros. En lo que sigue veremos

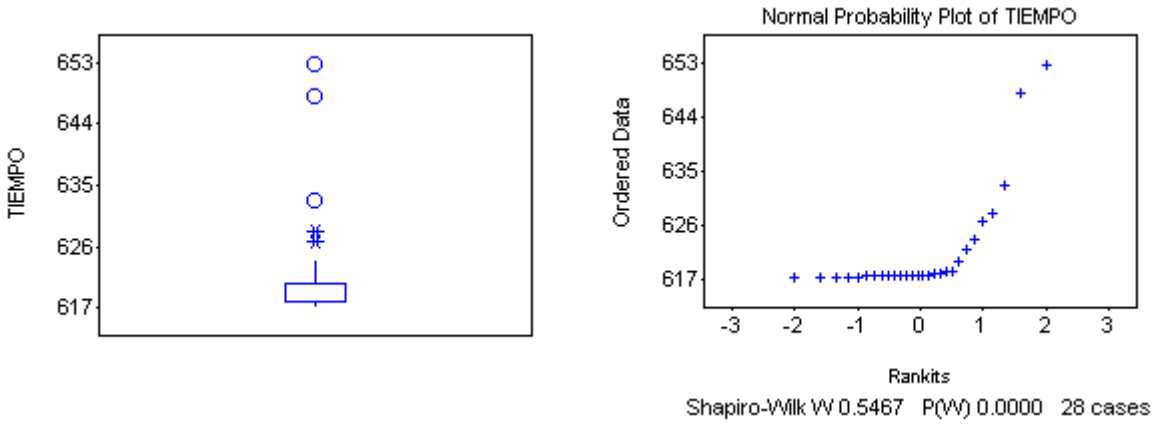
métodos que no requieren que los datos tengan una distribución paramétrica específica, son llamados métodos no paramétricos.

Ejemplo. Se midió el TIEMPO (seg.) que tarda la concentración de un compuesto reducirse a la mitad durante una reacción. Se realizaron 28 repeticiones de la reacción en condiciones independientes e idénticas.

Resultados obtenidos, ordenados de menor a mayor.

Tiempo	Tiempo	Tiempo	Tiempo
617.2	617.6	617.8	621.9
617.2	617.6	617.8	623.7
617.3	617.7	618.0	626.7
617.4	617.7	618.0	628.1
617.4	617.7	618.2	632.6
617.5	617.7	618.5	648
617.6	617.8	619.9	652.7

Box-plot y el gráfico de probabilidad Normal de los valores de la tabla anterior.



El box-plot y el gráfico de probabilidad Normal muestran que la distribución de la variable TIEMPO es fuertemente asimétrica a derecha. Mediante el test de Shapiro Wilk se rechaza la hipótesis de que la distribución de la variable es normal.

Si se desea resumir esta variable en un parámetro de posición, es preferible usar, en vez de la media, la **mediana**, ya que la interpretación de esta medida no depende de la forma de la distribución.
Estimamos la mediana poblacional θ, con la mediana muestral, med = 617.80

13.1 TEST DEL SIGNO

El test del signo nos permitirá decidir si la mediana, el parámetro de centralidad, de la población de la cual provienen los datos coincide o no con cierto valor, y no se requiere ningún supuesto aparte del de independencia sobre la distribución de los mismos.

Si la muestra proviene de una población cuya mediana es igual al valor θ_0, "valor nulo" especificado en la hipótesis nula uno esperaría encontrar, aproximadamente, igual cantidad de observaciones en la muestra por encima y por debajo de θ_o.

Definimos una nueva variable "diferencia": $D = X - \theta_0$. **Si H_0: $\theta = \theta_0$ es verdadera**, entonces $P(D > 0) = P(D < 0) = 0$ y la cantidad de diferencias positivas (o negativas) en la muestra tiene una distribución Bi (n,p) donde n = tamaño de muestra y p =1/2 es la probabilidad de que ocurra una diferencia positiva (o negativa).

Si en nuestra muestra encontramos gran cantidad de diferencias positivas (o negativas), esto será evidencia en contra de la hipótesis nula. Las diferencias iguales a 0 se ignoran y se trabaja con las diferencias distintas de cero, tanto positivas como negativas.

Modelo. $X_1, X_2,...,X_n$ observaciones independientes provenientes de una distribución con mediana θ.

Hipótesis nula: H_0: $\theta = \theta_0$

Hipótesis alternativas posibles

a) H_a: $\theta \neq \theta_0$	b) H_a: $\theta < \theta_0$	c) Ha: $\theta > \theta_0$

El estadístico del test está basado en las variable $D_i = X_i - \theta_o$ que suponemos independientes e igualmente distribuidas

Estadístico del test: cantidad de i's tal que la diferencia D_i es positiva (o negativa).

p-valor. Se calcula utilizando la distribución Bi(n,1/2) ó, si el tamaño de la muestra es suficientemente grande, la aproximación de la binomial por la Normal.

¿Cómo hacer el test del signo usando Statistix?

Para realizar el test del signo con el Statistix es necesario usar un artificio, generando una nueva variable que tome el valor θ_0. Esto se debe a que no ofrece como opción el test del signo para una muestra, sino para muestras apareadas (caso que presentaremos más adelante).

Trabajaremos sobre el ejemplo del tiempo que tarda la concentración de un compuesto en caer a la mitad durante una reacción.

Nos interesan las hipótesis H_0: $\theta = 620$ seg versus H_a: $\theta < 620$ segundos.

1) Generamos la variable VALORNU = 620
   Data -> Transformations
   `Transformation Expression: valornu = 620`

2) Statistics->One, Two, Multi-sample tests -> Sign Test
   Movemos a la casilla Sample Variables la variable que contiene los datos "TIEMPO" y la variable "VALORNU" que indica el valor nulo propuesto en H_0.

Obtenemos la salida siguiente:

```
SIGN TEST FOR TIEMPO - VALORNU

NUMBER OF NEGATIVE DIFFERENCES             21
NUMBER OF POSITIVE DIFFERENCES              7
NUMBER OF ZERO DIFFERENCES (IGNORED)        0

PROBABILITY OF A RESULT AS
OR MORE EXTREME THAN OBSERVED          0.0063 ⇐ p-valor (una cola)

A VALUE IS COUNTED AS A ZERO IF ITS
ABSOLUTE VALUE IS LESS THAN 0.00001

CASES INCLUDED 28    MISSING CASES 0
```

Observaciones

- Las diferencias iguales a cero se ignoran.
- El p-valor que calcula Statistix es a una cola.
- Para una hipótesis bilateral el p-valor sería el doble.
- En este ejemplo como hay más diferencias negativas que las esperadas por puro azar (7 positivas vs 21 negativas) rechazamos la hipótesis nula y concluimos que la mediana del tiempo de vida del compuesto es significativamente menor que 620 minutos (p = 0.0063).
- Este test NO HACE SUPUESTOS acerca de la forma de la distribución de la variable.

13.2 TEST DE RANGOS SIGNADOS DE WILCOXON

El test del signo solamente tiene en cuenta si las observaciones son mayores o menores que la mediana propuesta en H_0. Propondremos ahora un test no paramétrico que compara las mismas hipótesis que el test del signo, pero que tiene en un poco más en cuenta la magnitud de las observaciones.
Definimos el **rango** de una observación como la posición que ocupa en la muestra ordenada de menor a mayor. Cuando hay observaciones repetidas se les asigna el rango promedio.
Consideremos los siguientes datos:

Datos ordenados ⇒	35	85	96	96	180	200
Rango ⇒	1	2	3	4	5	6
			3.5			

¿Cómo se construye el test de Wilcoxon?

Modelo: $X_1, X_2, \ldots, X_n$ son observaciones independientes provenientes de una distribución continua **simétrica** con mediana θ.

Hipótesis nula: $H_0: \theta = \theta_0$

Hipótesis alternativas posibles

a) $H_a: \theta \neq \theta_0$	b) $H_a: \theta < \theta_0$	c) Ha: $\theta > \theta_0$

Estadístico del test:

- Se obtienen las diferencias $D_i = X_i - \theta_o$
- Se ordenan las diferencias sin tener en cuenta su signo (es decir, se ordenan los valores
- absolutos de las diferencias) y se le asigna a cada una un rango.
- Se calcula la suma de los rangos de las diferencias positivas (o negativas). Este valor se denomina W y es el ESTADÍSTICO del test.

p-valor Statistix calcula el p-valor y el estadístico del test automáticamente.

Para calcular el p-valor del test debemos conocer la *distribución del estadístico*. Esta distribución no es simple y se encuentra tabulada para $n \leq 25$. Para muestras grandes la distribución se aproxima a una Normal con media $\mu = \frac{n(n+1)}{4}$ y varianza $\sigma^2 = \frac{n(n+1)(2n+1)}{24}$.

Si la hipótesis nula es verdadera, la suma de rangos positivos será aproximadamente igual a la mitad de la suma de rangos totales. Pero si H_0 es falsa, la suma de rangos positivos será notablemente mayor o menor que la mitad de la suma total.

Consideremos el siguiente ejemplo, con $\theta_0 = 100$

Datos ordenados (X) ⇒	35	85	96	140	180	200	240	289	360	400
Diferencias ⇒	–75	–15	–4	40	80	100	140	189	260	300
Diferencias absolutas ⇒	75	15	4	40	80	100	140	189	260	300
Rangos ⇒	4	2	1	3	5	6	7	8	9	10

Suma rangos negativos = 4 + 2 + 1 = 7
Suma rangos positivos = 3 + 5 + 6 + 7 + 8 + 9 + 10 = 48
Parecen diferentes!! deberíamos calcular la probabilidad de encontrar estos valores cuando H_o es verdadera. Lo haremos inmediatamente con Statistix.
Supongamos que interesa hacer un test para las hipótesis:

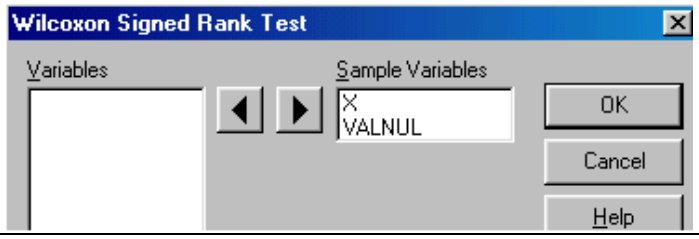
$H_o: \theta = 100$ vs $H_o: \theta \neq 100$

¿Cómo hacer el test de Wilcoxon usando Statistix?

Igual que con el test del signo, por ser otra opción para muestras apareadas, es necesario generar una variable artificial que tome el valor θ_o.

1) Generamos la variable artificial VALNUL = 100
   Data ->Transformations
   `Transformation Expression:  VALNUL = 100`

2) Statistics -> One, Two, Multi-sample tests -> Wilcoxon Signed Rank Test
   Indicamos la variable que contiene los datos (X) y la variable artificial que indica el valor propuesto en H_o.




```
WILCOXON SIGNED RANK TEST FOR X - VALNUL

SUM OF NEGATIVE RANKS                                  -7.0000
SUM OF POSITIVE RANKS                                   48.000

EXACT PROBABILITY OF A RESULT AS OR MORE
EXTREME THAN THE OBSERVED RANKS (1 TAILED P-VALUE)   0.0186

NORMAL APPROXIMATION WITH CONTINUITY CORRECTION        2.039
TWO TAILED P-VALUE FOR NORMAL APPROXIMATION           0.0415

TOTAL NUMBER OF VALUES THAT WERE TIED      0
NUMBER OF ZERO DIFFERENCES DROPPED         0
MAX. DIFF. ALLOWED BETWEEN TIES      0.00001

CASES INCLUDED 10    MISSING CASES 0
```

La suma de rangos positivos y negativos coincide con los calculados más arriba.

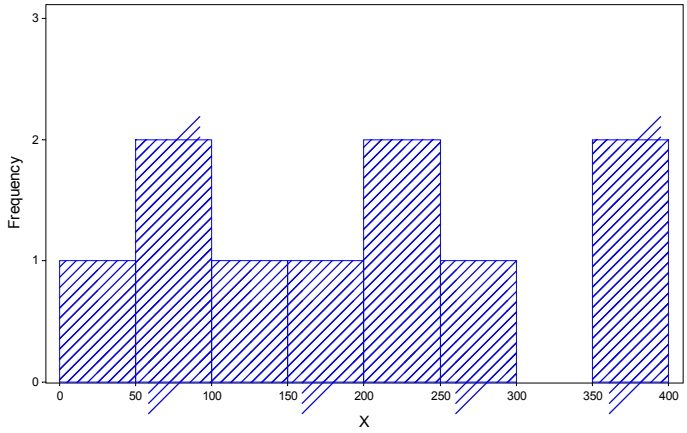
Como el test es a dos colas p = 2 * 0.0186 = 0.0372 (p-valor basado en la distribución exacta).

Si aplicamos el test del signo a estos datos obtenemos el siguiente resultado:

```
SIGN TEST FOR X - VALNUL

NUMBER OF NEGATIVE DIFFERENCES              3
NUMBER OF POSITIVE DIFFERENCES              7
NUMBER OF ZERO DIFFERENCES (IGNORED)        0

PROBABILITY OF A RESULT AS
OR MORE EXTREME THAN OBSERVED          0.1719 ⇐  p = 2*0.172 = 0.344


A VALUE IS COUNTED AS A ZERO IF ITS
ABSOLUTE VALUE IS LESS THAN 0.00001
```

Como vemos el test del signo falla en rechazar H_0, mientras que el test de Wilcoxon rechaza H_0. Esto se debe a que el estadístico W no sólo tiene en cuenta la cantidad de datos mayores que el "valor nulo" para la mediana propuesto en H_0 sino además la distancia expresada a través de los rangos.

¿Qué supuestos deben cumplir los datos para poder aplicar el test de rangos de Wilcoxon?
La distribución debe ser simétrica. En este caso, no parece haber apartamientos groseros de la simetría (ver histograma), por lo que concluimos que es válido usar este test.



13.3 TEST DE MANN-WHITNNEY – TEST DE WILCOXON para dos muestras independientes

En el test de t para dos muestras se comparan las medias muestrales de dos conjuntos de datos y se establece si la diferencia es estadísticamente significativa o es atribuible al azar. Ahora en vez de realizar los cálculos sobre los datos se toman los rangos de los mismos en una muestra combinada ordenada de menor a mayor. El estadístico del test es equivalente a comparar las medias de los rangos de cada una de las muestras.

Si no existieran diferencias entre los dos grupos y ambos tuvieran **la misma cantidad de observaciones**, esperaríamos que la suma de rangos de la Muestra 1 fuera "similar" a la suma de rangos de la Muestra 2. Lo que nos indicaría que los datos de las dos muestras aparecen alternadamente en la muestra total ordenada.

Si proponemos como estadístico del test *la suma de rangos de cada muestra*, una suma demasiado grande o demasiado pequeña será indicativa de diferencias entre las dos poblaciones de las cuales fueron obtenidas las muestras. Por lo tanto, la hipótesis nula de que las dos poblaciones no difieren debería ser rechazada cuando la suma de rangos de una muestra tiende a ser notablemente mayor (o menor) que los de la otra muestra.

¿Por qué preferir un test de rangos?
1. Si los datos no son numéricos y corresponden a categorías ordinales, los rangos contienen la misma información que los datos.

2. Si la variable es numérica y su distribución no es normal, no valen los tests que hemos estudiado para muestras pequeñas. La teoría de los métodos basados en rangos es relativamente simple y no es necesario especificar una distribución para las variables.

El test que presentaremos se conoce como Test de Mann-Whitney pero también como Test de Wilcoxon, ya que los primeros autores generalizaron el procedimiento que el segundo propuso para el problema de muestras independientes de igual tamaño.

Presentaremos dos modelos posibles para el problema. Cada modelo permite testear diferentes hipótesis respecto de las poblaciones de las cuales provienen los datos, pero el test es el mismo.

Modelo (1). Proponemos que ambas muestras provengan de poblaciones con la misma **distribución F cualquiera**.

$X_1, X_2, ..., X_n$ i.i.d ; distribución F con mediana θ_X.

$Y_1, Y_2, ..., Y_m$ i.i.d ; distribución F con mediana θ_Y.

Si hay una diferencia entre ellas se debe SOLO a la posición de la distribución.

Hipótesis nula (1): $H_0: \theta = \theta_0$

Hipótesis alternativas posibles

a) $H_a: \theta \neq \theta_0$	b) $H_a: \theta < \theta_0$	c) Ha: $\theta > \theta_0$

Modelo (2) El modelo propone que los datos de cada muestra provienen de diferentes distribuciones.

$X_1, X_2, ..., X_n$ i.i.d ; distribución F

$Y_1, Y_2, ..., Y_m$ i.i.d ; disribución G

Hipótesis (2): H_0: $F(x) = G(x)$ para todo x versus

H_a: $F(x) \neq G(x)$ para algún x

La hipótesis nula afirma que las dos distribuciones poblacionales son iguales (es equivalente a la H_0 del Modelo (1)), pero la alternativa dice que las dos distribuciones difieren de algún modo, pero no dice de qué modo.

Estadístico del test es el mismo para ambos modelos

T = Suma de rangos de la muestra con menor número de observaciones

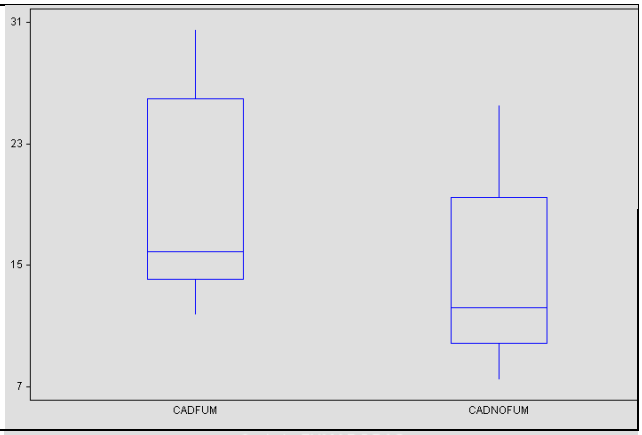
Si hay muchos empates se estandariza la suma de rangos restando su media y dividiendo por la desviación estándar.

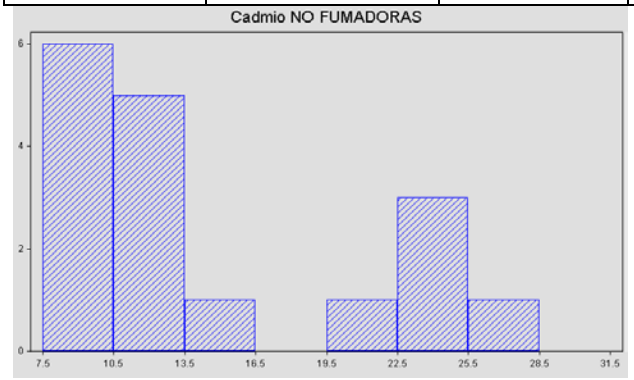
La distribución de este estadístico para tamaños de muestra pequeños está tabulada. Para tamaños de muestras grandes se usa una aproximación normal a la distribución del estadístico.

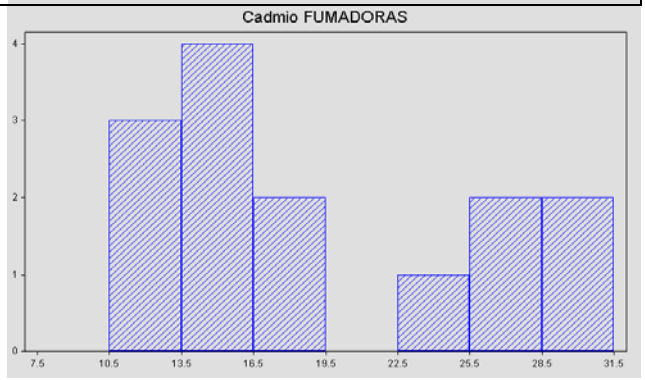
¿Cómo hacer el test de Mann-Whitney usando Statistix?

Ejemplo. En un estudio sobre consecuencias adversas del tabaquismo durante la gestación, se midieron los niveles de cadmio (ng/g) en tejido de placenta de una muestra de 14 mujeres fumadoras y 18 no fumadoras. Los datos están organizados mediante dos variables: una que indica el nivel de cadmio y otra que indica el grupo. Los resultados se resumen en la tabla y los gráficos siguientes.

	No fumadoras	**Fumadoras**
Media	14.71	19.09
Desvío estándar	6.39	6.78
n	17	14
Shapiro-Wilk (p-valor)	0.829 (0.005)	0.849 (0.021)



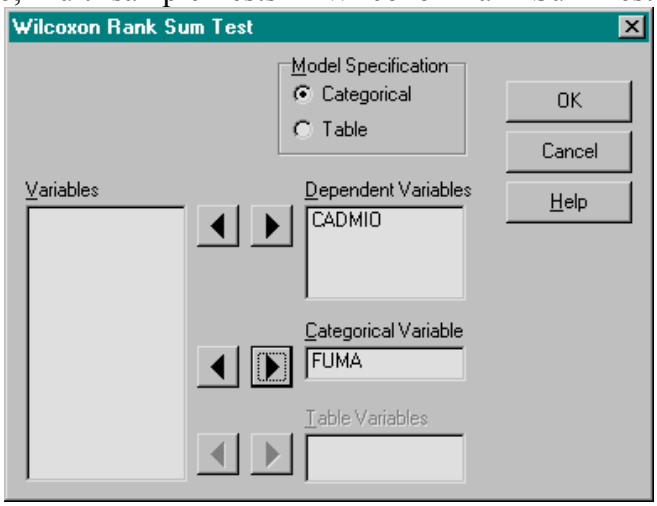




Concluimos que:

- Los datos provienen de distribuciones no normales.
- Los histogramas y los box-plots muestran distribuciones asimétricas, con asimetría derecha. Las dos distribuciones son razonablemente similares.

Statistics -> One, Two, Multi-sample Tests -> Wilcoxon Rank Sum Test



WILCOXON RANK SUM TEST FOR CADMIO BY FUMA

FUMA	RANK SUM	SAMPLE SIZE	U STAT	MEAN RANK
no fuma	214.00	17	61.000	12.6
fuma	282.00	14	177.00	20.1
TOTAL	496.00	31		

NORMAL APPROXIMATION WITH CORRECTIONS FOR CONTINUITY AND TIES 2.284
TWO-TAILED P-VALUE FOR NORMAL APPROXIMATION **0.0224**

TOTAL NUMBER OF VALUES THAT WERE TIED 11
MAXIMUM DIFFERENCE ALLOWED BETWEEN TIES 0.00001

p-valor (2 colas)

CASES INCLUDED 31 MISSING CASES 0

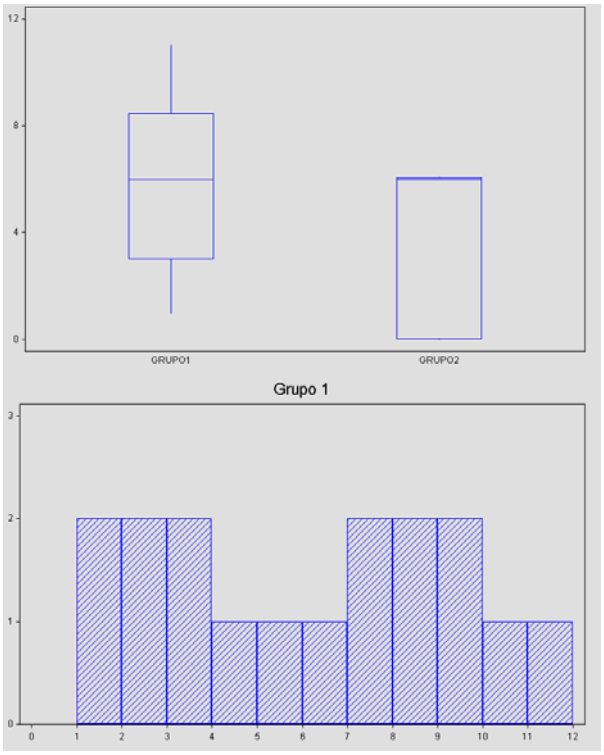
¿Qué hipótesis estamos testeando?
Podemos suponer que la distribución de cadmio es similar en las dos poblaciones, por lo tanto, planteamos las hipótesis sobre las medianas poblacionales.
Sea θ_X = mediana poblacional de cadmio en no fumadoras
θ_Y = mediana poblacional de cadmio en fumadoras

$$H_0: \theta_X = \theta_Y \text{ versus } H_a: \theta_X \neq \theta_Y$$

Conclusión:
Si pretendemos un test de nivel 5%, como el p-valor = 0.0224 < 0.05, entonces, se rechaza H_0 y concluimos que la mediana del nivel de cadmio en placenta es mayor en la población de mujeres fumadoras que en las no fumadoras.

¿Es válido usar el test de Mann-Whitney si la distribución de los datos en las dos muestras es claramente diferente?
Debemos tener en cuenta en este caso que el test de Mann-Whitney / Wilcoxon no es un test para el parámetro de posición, por lo tanto, si rechazamos la hipótesis nula, podemos concluir que las distribuciones difieren pero no sabemos en de que modo difieren.
Consideremos el siguiente ejemplo.

Datos ficticios



Las distribuciones son diferentes (no hay duda), pero la mediana es la misma en los dos grupos (ver box-plots). Veamos el resultado del test de Mann-Whitney.

```
WILCOXON RANK SUM TEST FOR GRUPO1 VS GRUPO2

                        SAMPLE
VARIABLE     RANK SUM    SIZE      U STAT    MEAN RANK
---------  ---------  ------  ---------  ---------
GRUPO1        361.50      17      208.50       21.3
GRUPO2        233.50      17      80.500       13.7
TOTAL         595.00      34

NORMAL APPROXIMATION WITH CORRECTIONS FOR CONTINUITY AND TIES   2.216
TWO-TAILED P-VALUE FOR NORMAL APPROXIMATION                   0.0267

TOTAL NUMBER OF VALUES THAT WERE TIED        18
MAXIMUM DIFFERENCE ALLOWED BETWEEN TIES 0.00001

CASES INCLUDED 34    MISSING CASES 0
```

Se rechaza la hipótesis nula y concluimos que las dos distribuciones difieren, pero ¿cómo difieren? Sólo podremos describir la forma en que las dos distribuciones difieren describiendo los histogramas. Pero no podemos resumir estos datos a través de las medianas y acompañando a la conclusión del test de Mann-Whitney – Wilcoxon.
Consideraremos a continuación un test que no hace supuestos sobre las distribuciones de las dos muestras.

13.4 TEST DE LA MEDIANA

Este test se puede generalizar a más de dos grupos y es una alternativa al test de Mann Whitney cuando interesa un test para el parámetro de posición.
Puede ser usado con datos numéricos o categóricos ordinales.

Modelo:

$X_1, X_2, ..., X_n$ i.i.d ; distribución F con mediana θ_X.

$Y_1, Y_2, ..., Y_m$ i.i.d ; disribución G con mediana θ_Y.

Hipótesis: H_0: $\theta_X = \theta_Y$ versus H_a: $\theta_X \neq \theta_Y$

(Este test tal como lo calculan la mayoría de los paquetes no acepta hipótesis alternativas unilaterales).

Estadístico: No daremos la expresión formal del estadístico, sino que presentaremos la idea de como se construye.

- Se ordenan los n+m datos y se calcula la mediana θ general.
- Se cuenta el número de observaciones menores o iguales que la mediana θ en cada muestra (m_x y m_y) y el número de observaciones mayores que la mediana θ (M_X y M_Y).
  Estos datos se vuelcan a una tabla de doble entrada como la siguiente:

	Muestra X's	Muestra Y´s
$\leq \theta$	m_X	m_Y
$> \theta$	M_X	M_Y
	n	m

- Si H_o es verdadera las proporciones de datos menores que la mediana y mayores que la mediana deberían ser similares en las dos muestras, es decir, esperamos que $\frac{m_X}{n} \cong \frac{m_Y}{m}$ y $\frac{M_X}{n} \cong \frac{M_Y}{m}$.

El estadístico del test mide la distancia entre lo observado y lo esperado cuando H_0 es verdadera. Si las muestras son relativamente grandes, el estadístico tiene distribución aproximada χ^2 (chi-cuadrado) con 1 grado de libertad.

Aplicaremos el test de la mediana para los datos de los dos últimos ejemplos.

```
MEDIAN TEST FOR CADMIO BY FUMA

                        FUMA
                 -----------------
                 no fuma     fuma      TOTAL
                 -------   -------   -----
ABOVE MEDIAN           6         9      15
BELOW MEDIAN          11         4      15
TOTAL                 17        13      30
TIES WITH MEDIAN       0         1       1

MEDIAN VALUE      14.400
```

```
CHI-SQUARE  3.39   DF 1   P-VALUE 0.0654

MAX. DIFF. ALLOWED BETWEEN A TIE  0.00001

CASES INCLUDED 31    MISSING CASES 0
```

Conclusión: A nivel 5% no hay suficiente evidencia para rechazar H_o.
Recordemos que el test de Mann-Whitney rechazó la misma hipótesis nula con un p-valor = 0.022. El test basado en rangos es más potente que el test de la mediana (usado en condiciones comparables).

Datos *ficticios*

```
MEDIAN TEST FOR GRUPO1 - GRUPO2

                   GRUPO1      GRUPO2      TOTAL
                 ---------   ---------   ---------
ABOVE MEDIAN           8           8          16
BELOW MEDIAN           8           8          16
TOTAL                 16          16          32
TIES WITH MEDIAN       1           1           2

MEDIAN VALUE      6.0000

CHI-SQUARE  0.00   DF 1   P-VALUE 1.0000

MAX. DIFF. ALLOWED BETWEEN A TIE  0.00001

CASES INCLUDED 34    MISSING CASES 0
```

Conclusión: No se rechaza la hipótesis que los dos grupos tienen la misma mediana. La conclusión es francamente diferente de la que obtuvimos con el test de Mann-Whitney porque este último NO es un test para el parámetro de posición.

El test de la mediana puede transformarse en un *test de percentiles* para la hipótesis nula que las dos poblaciones tienen el mismo percentil *p*. Simplemente ordenamos todas las observaciones, calculamos el percentil *p* general y luego contamos que proporción de los datos de cada muestra caen por debajo del percentil general. Los datos se vuelcan en una tabla de doble entrada, y el estadístico tiene distribución aproximada chi-cuadrado.