Este material está basado en los apuntes de la Dra. Diana Kelmansky, disponibles en
http://www.dm.uba.ar/materias/estadistica_Q/2010/2/

TEST DEL SIGNO

Test del signo para una muestra

En el caso de una muestra, el test del signo permite decidir si la mediana de la distribución de la variable en la población de la cual provienen los datos coincide o no con cierto valor, Hipótesis nula: H_0: $\theta = \theta_0$

Este test requiere el supuesto de independencia de las observaciones.

Modelo X1, ... , Xn una muestra aleatoria de una distribución cualquiera con mediana θ.

Definimos una nueva variable "diferencia": $D_i = X_i - \theta_0$, i=1,2, ...,n

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

Si H_0: $\theta = \theta_0$ **es verdadera**, la variable aleatoria D_i tiene la misma probabilidad de tomar valores positivos que negativos es decir:

$$P(D_i > 0) = P(D_i < 0) = 0{,}5 \text{ (suponiendo que } P(X_i = \theta_0)=0).$$

Definimos ahora otra nueva variable aleatoria.

S_n = cantidad de la cantidad de diferencias positivas (o negativas) en la muestra

> $S_n \sim Bi\,(n,p)$, n = tamaño de muestra y p es la probabilidad de que ocurra una diferencia positiva (o negativa). Cuando H0 es verdadera p = 1/2.

<u>Estadístico del test</u>: $S_n \sim Bi\,(n,1/2)$ bajo H_0

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

<u>Hipótesis alternativas posibles</u>

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

Hipótesis alternativas posibles

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

Regiones de rechazo de nivel α

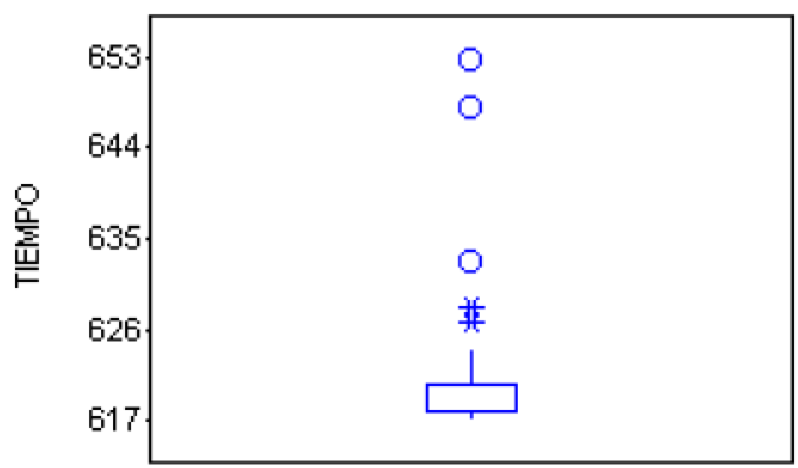
a) H_a: $\theta \neq \theta_0$	b) H_a: $\theta < \theta_0$	c) Ha: $\theta > \theta_0$
$S_n \geq n-k$ ó $S_n \leq k$	$S_n \leq k$	$S_n \geq n-k$
donde $P(Bi(n,1/2) \leq k) = \alpha/2$	donde $P(Bi(n,1/2) \leq k) = \alpha$	donde $P(Bi(n,1/2) \leq k) = \alpha$

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.

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

Resultados ordenados por columna de menor a mayor.



X= tiempo que tarda la concentración de un compuesto en caer a la mitad durante una reacción.

H_0: θ = 620 seg $\qquad$ H_a: θ < 620 segundos.

El valor observado de $S = card\{i / X_i - 620 > 0\}$ = 7 y recordemos que rechazábamos H_o si $S \leq$ k con k tal que $P(Bi(28,1/2) \leq k) = \alpha$.

Si tomamos k = 9, se obtiene un nivel α = 0.0436. Claramente a este nivel rechazamos H_o.

```
STATISTIX 7.0                       EJEMPLO_TEST_SIGNO

SIGN TEST FOR TIEMPO - MEDIANAHO
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
A VALUE IS COUNTED AS A ZERO IF ITS
ABSOLUTE VALUE IS LESS THAN 0.00001
CASES INCLUDED 28    MISSING CASES 0
```

← p-valor (una cola)

Observaciones

- La distribución del estadístico del test bajo H_0 es Bi(n,1/2) y el p-valor puede calcularse exactamente utilizando la distribución binomial. Si el tamaño de la muestra es suficientemente grande también puede utilizarse la aproximación de la binomial por la Normal.
- El test únicamente tiene en cuenta el signo de las diferencias y no su magnitud.
- Las diferencias iguales a cero se ignoran.
- El test NO REQUIERE SUPUESTOS acerca de la FORMA de la distribución de la variable subyacente a los datos, únicamente sobre la INDEPENDENCIA de las observaciones.
- Para realizar el test del signo para una muestra con el Statistix es necesario usar un artificio, generando una nueva variable que tome el valor θ_0.
- El p-valor que calcula Statistix es a una cola. Para obtener el p-valor de una hipótesis bilateral, el p-valor de la salida se duplica.

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. El siguiente test no paramétrico compara las mismas hipótesis que el test del signo, tiene en un poco más en cuenta la magnitud de las observaciones pero requiere el supuesto de SIMETRÍA.

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: $\theta > \theta_0$

Construcción del Estadístico del test:

- $D_i = X_i - \theta_o$

- Se ordenan D_i 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.

  Sea $|D|^{(1)} \leq |D|^{(2)} .... \leq |D|^{(n)}$, la muestra de valores absolutos ordenados y

  R_j = rango ($|D_j|$) es decir $|D_j| = |D|^{(Rj)}$

- La suma de los rangos de las diferencias positivas (o negativas) es el ESTADÍSTICO del test. Este valor se lo denota por W

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.

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 $$\mu = \frac{n(n+1)}{4}$$

varianza $$\sigma^2 = \frac{n(n+1)(2n+1)}{24}.$$

Ejemplo

H_0 $\theta = 100$ vs	H_0 $\theta \neq 100$

Es decir que $\theta_0 = 100$.

La tabla siguiente ejemplifica los cálculos previos para obtener el valor del estadístico del test

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 de diferencias negativas = 4 + 2 + 1 = 7
Suma rangos de diferencias positivas = 3 + 5 + 6 + 7 + 8 + 9 + 10 = 48

Usando Statistix

```
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
```

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

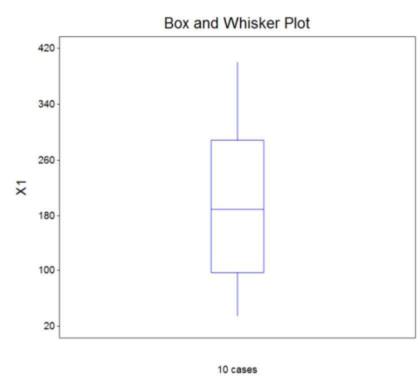
Si aplicamos el test del signo a estos datos obtenemos el siguiente resultado:

```
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
```

El test del signo no rechaza H_0, mientras que el test de Wilcoxon sí 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 propuesta en H_0 sino además el orden de las diferencias 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 alejamientos groseros de la simetría



¿Cómo se asignan los rangos si hay observaciones repetidas?

El test asigna el rango promedio a las observaciones repetidas.

Datos ordenados (X)	35	85	96	140	140	180	200	240	289	360	400
Diferencias	–75	–15	–4	40	40	80	100	140	189	260	300
Diferencias absolutas	75	15	4	40	40	80	100	140	189	260	300
Posición	5	2	1	3	4	6	7	8	9	10	11
Rangos asignados	5	2	1	3.5	3.5	6	7	8	9	10	11

```
WILCOXON SIGNED RANK TEST FOR X - VALNUL

SUM OF NEGATIVE RANKS -8.0000
SUM OF POSITIVE RANKS 58.000

EXACT PROBABILITY OF A RESULT AS OR MORE
EXTREME THAN THE OBSERVED RANKS (1 TAILED P-VALUE)   0.0122

NORMAL APPROXIMATION WITH CONTINUITY CORRECTION        2.178
TWO TAILED P-VALUE FOR NORMAL APPROXIMATION          0.0294

CASES INCLUDED 11    MISSING CASES 0
```

TEST DE MANN-WHITNNEY – TEST DE WILCOXON

Consideramos a continuación 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). Ambas muestras proviene de poblaciones cuyas distribuciones solamente difieren en la posición esto es:

$X_1, X_2, \dots, X_n$ i.i.d ; distribución F(x)

$Y_1, Y_2, \dots, Y_m$ i.i.d ; distribución G(x) = F(x+c)

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

Llamamos θ_X a la mediana de las observaciones con distribución F y θ_Y a la mediana de las observaciones con distribución G,

Hipótesis nula (1): H_0: $\theta_X = \theta_Y$

Hipótesis alternativas posibles H_a:

a) $\theta_X \neq \theta_Y$	b) $\theta_X < \theta_Y$	c) $\theta_X > \theta_Y$

Modelo (2) 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 ; distribución G

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

Hipótesis alternativa (2) 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 cantidad de observaciones

Cuando hay empates entre las observaciones se reemplaza el rango de cada empate por el promedio de los rangos de los empates

La distribución del estadístico está tabulada para tamaños de muestra muy pequeños. Para tamaños de muestras moderados mayores a 10 se usa una aproximación Normal, modificando el estadístico por una versión estandarizada y se utilizan valores críticos de la tabla de N(0,1).

Ejemplo :

En una clase de una escuela de enseñanza superior hay 48 alumnos varones, de los cuáles 12 viven en el campo y 36 en la ciudad. Se desarrolló un test para determinar la condición física de los alumnos. Se aplicó este test a los 48 alumnos y se asignó a cada uno una puntuación. Una puntuación baja indica mala condición física.

Los resultados son los siguientes:

Campo (X_i)		**Ciudad (Y_i)**					
14.8	10.6	12.7	16.9	7.6	2.4	6.2	9.9
7.3	12.5	14.2	7.9	11.3	6.4	6.1	10.6
5.6	12.9	12.6	16.0	8.3	9.1	15.3	14.8
6.3	16.1	2.1	10.6	6.7	6.7	10.6	5.0
9.0	11.4	17.7	5.6	3.6	18.6	1.8	2.6
4.2	2.7	11.8	5.6	1.0	3.2	5.9	4.0

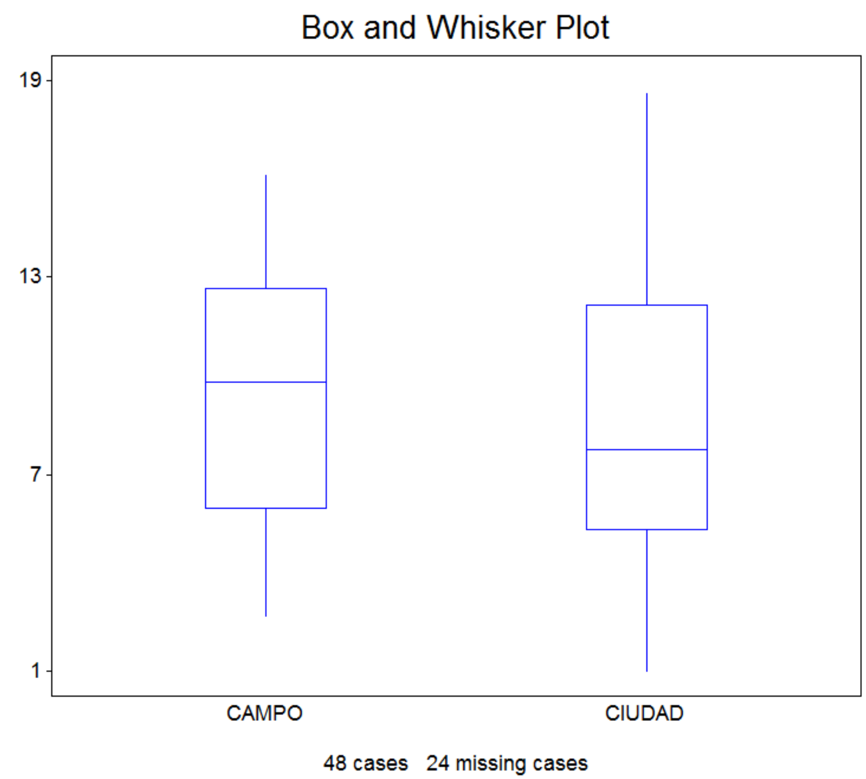
Se testea

H_o: los alumnos del campo y de ciudad tienen la misma condición física

H_1: los alumnos del campo tienen $\neq$ condición física que los de ciudad

DESCRIPTIVE STATISTICS

	CAMPO	CIUDAD
N	12	36
MEAN	9.4500	8.6500
SD	4.2839	4.9143
1ST QUARTI	5.7750	5.1500
MEDIAN	9.8000	7.7500
3RD QUARTI	12.8000	12.4000
MAD	3.3000	3.6500



X	Y	Rango	X	Y	Rango	X	Y	Rango
	1.0	1		6.2	17		11.3	33
	1.8	2	6.3		18	11.4		34
	2.1	3		6.4	19		11.8	35
	2.4	4		6.7	20.5	12.5		36
	2.6	5		6.7	20.5		12.6	37
2.7		6	7.3		22		12.7	38
	3.2	7		7.6	23	12.9		39
	3.6	8		7.9	24		14.2	40
	4.0	9		8.3	25		14.8	41.5
4.2		10	9.0		26	14.8		41.5
	5.0	11		9.1	27		15.3	43
	5.6	13		9.9	28		16.0	44
	5.6	13		10.6	30.5	16.1		45
5.6		13		10.6	30.5		16.9	46
	5.9	15	10.6		30.5		17.7	47
	6.1	16		10.6	30.5		18.6	48

$$T = \sum_{i=1}^{n} R(X_i) = 321$$

STATISTIX 7.0 EJEMPLO_MANN-WHITNEY

WILCOXON RANK SUM TEST FOR CAMPO VS CIUDAD

VARIABLE	RANK SUM	SAMPLE SIZE	U STAT	MEAN RANK
CAMPO	**321.00**	**12**	**243.00**	**26.8**
CIUDAD	**855.00**	**36**	**189.00**	**23.8**
TOTAL	**1176.0**	**48**		

NORMAL APPROXIMATION WITH CORRECTIONS FOR CONTINUITY AND TIES 0.631
TWO-TAILED P-VALUE FOR NORMAL APPROXIMATION 0.5279

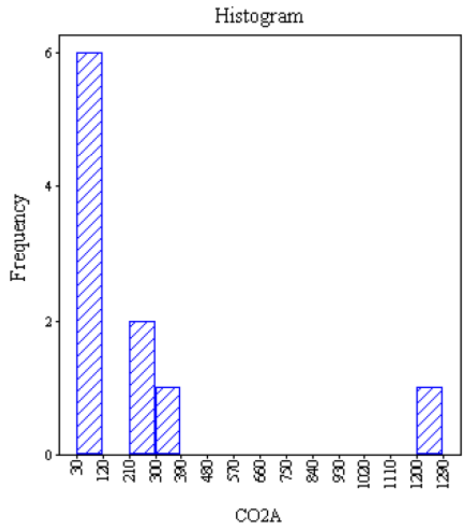
TOTAL NUMBER OF VALUES THAT WERE TIED 11
MAXIMUM DIFFERENCE ALLOWED BETWEEN TIES 0.00001

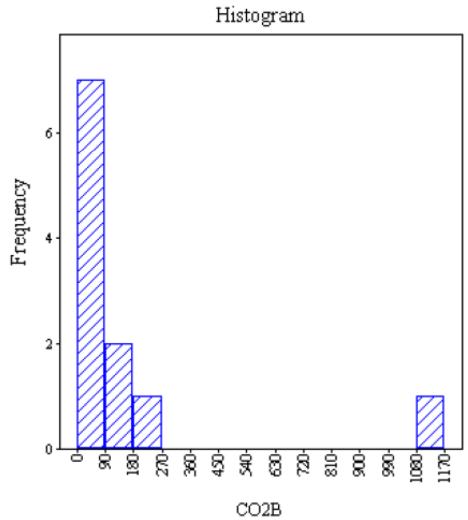
CASES INCLUDED 48 MISSING CASES 24

Ejemplo. Interesa comparar la cantidad de dióxido de carbono desprendido (mol CO_2/g suelo/hr) en dos tipos de de suelos, A con escasa y B con abundante vegetación, como medida de la actividad microbiana en los mismos ya que ésta afecta el crecimiento de las plantas.

Se obtuvieron 10 determinaciones en suelos tipo A y 11 en suelos tipo B. La tabla siguiente muestra un resumen de los resultados:

VARIABLE	N	MEAN	SD	MINIMUM	MAXIMUM
CO2A	10	263.70	362.83	60.000	1250.0
CO2B	11	155.45	339.37	1.0000	1160.0





Es adecuado considerar válido el Modelo (1): ambas muestras provienen de poblaciones con la misma **distribución F cualquiera**.

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

que las medianas de la cantidad de dióxido de carbono desprendido (mol CO_2/g suelo/hr) en dos tipos de de suelos coinciden, contra las medianas de la cantidad de dióxido de carbono desprendido difieren

```
WILCOXON RANK SUM TEST FOR CO2A VS CO2B

                             SAMPLE
VARIABLE     RANK SUM     SIZE      U STAT    MEAN RANK
---------   ---------   ------   ---------   ---------
CO2A          142.00      10      87.000       14.2
CO2B          89.000      11      23.000        8.1
TOTAL         231.00      21

EXACT PERMUTATION TEST TWO-TAILED P-VALUE     0.0879

NORMAL APPROXIMATION WITH CORRECTIONS FOR CONTINUITY AND TIES   2.218
TWO-TAILED P-VALUE FOR NORMAL APPROXIMATION                    0.0265

TOTAL NUMBER OF VALUES THAT WERE TIED          0
MAXIMUM DIFFERENCE ALLOWED BETWEEN TIES 0.00001

CASES INCLUDED 21    MISSING CASES 1
```

Se rechaza la hipótesis nula con un nivel $\alpha = 0.10$. Decimos que las medianas muestrales de la cantidad de dióxido de carbono desprendido difieren significativamente al 10% entre los dos tipos de suelo.