

23. TESTS E INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS DE DOS POBLACIONES

Un test de hipótesis para dos muestras es similar en muchos aspectos al test para una muestra.

- Se especifica una hipótesis nula, en la mayoría de los casos se propone que las medias de las dos poblaciones son iguales y se establece la hipótesis alternativa (uni o bilateral).
- Se especifica un nivel de significación α .
- Se calcula el p-valor: la probabilidad de obtener datos cuyas medias muestrales difieren tanto o más que la diferencia observada cuando H_0 es verdadera. Si esta probabilidad es pequeña (menor que α) se rechaza H_0 y se concluye que la diferencia observada no es atribuible al azar y las medias de las dos poblaciones son diferentes.

El estadístico del test dependerá de la estructura de los conjuntos de datos. En particular es importante establecer si los datos corresponden a *muestras apareadas* o *independientes*.

23.1 MUESTRAS APAREADAS

La característica distintiva de las muestras apareadas es que para cada observación del primer grupo, hay una observación relacionada en el segundo grupo. Este diseño es más complejo que el de muestras independientes (sección 23.2) pero lo presentamos en primer término porque se pueden aplicar, con sólo tomar diferencias, todos los procedimientos que hemos visto para una muestra.

Las **muestras apareadas** se obtienen cuando se realizan comparaciones sobre una misma unidad experimental:

- se determina en la misma unidad la concentración de una sustancia con dos métodos diferentes.
- se estudia un mismo individuo antes y después de un tratamiento.

Ejemplo 1: Interesa decidir si una **dieta escasa en calorías** produce un cambio en la tensión media de dióxido de carbono arterial en pacientes con problemas respiratorios crónicos

Tensión de dióxido de carbono arterial (mm Hg)
en pacientes con problemas respiratorios

PACIENTE	ANTES	DESPUÉS	DIFERENCIA
1	49	45	4
2	68	54	14
3	65	60	5
4	57	60	-3

5	76	59	17
6	62	54	8
7	49	47	2
8	53	50	3
MEDIA	59.875	53.625	6.2500
DESVÍO	9.6279	5.8782	6.5411
MINIMO	49.000	45.000	-3.0000
CUARTIL Inf.	50.000	47.750	2.2500
MEDIANA	59.500	54.000	4.5000
CUARTIL Sup.	67.250	59.750	12.500
MAXIMO	76.000	60.000	17.000

¿Qué se observa?

- Hay una gran superposición, de más del 50%, entre los dos conjuntos de datos.
- Parece haber una leve reducción del dióxido de carbono después de la dieta.
- Los valores dióxido de carbono de un mismo paciente están relacionados.

Llamemos X_i = concentración de CO_2 del paciente i antes de realizar la dieta

Y_i = concentración de CO_2 del paciente i después de realizar la dieta

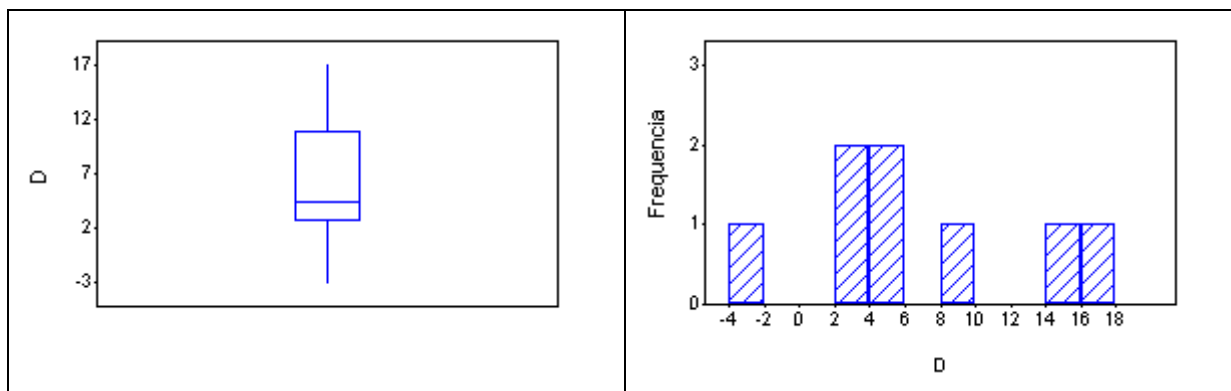
Definimos una nueva variable, la diferencia entre las concentraciones de un paciente.

$$D_i = X_i - Y_i.$$

$D_i > 0 \Rightarrow$ la concentración de CO_2 del paciente i se redujo al realizar la dieta.

$D_i = 0 \Rightarrow$ la concentración de CO_2 del paciente i se mantuvo igual.

$D_i < 0 \Rightarrow$ la concentración de CO_2 del paciente i aumentó al realizar la dieta.



Cuando graficamos el box-plot y el histograma de la variable diferencia observamos que en la mayoría de los pacientes la concentración de CO_2 se redujo al realizar la dieta

23.1.1 Consideraciones generales bajo Normalidad de las diferencias

Sean (X_i, Y_i) ($1 \leq i \leq n$) las observaciones realizadas sobre n unidades experimentales, para cada par (X_i, Y_i) ; $D_i = X_i - Y_i$

Modelo D_1, \dots, D_n v.a.i.i.d $N(\mu_D, \sigma_D^2)$ con $\mu_D = \mu_X - \mu_Y$

No realizamos ningún supuesto sobre la distribución de cada variable, solamente sobre la diferencia. Luego

$$\frac{\bar{D} - \mu_D}{S_D / \sqrt{n}} \sim t_{n-1} \quad \Leftrightarrow \quad \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_D / \sqrt{n}} \sim t_{n-1}$$

$$\text{donde } \bar{D} = \frac{\sum_{i=1}^n D_i}{n} = \bar{X} - \bar{Y} \quad y \quad S_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}$$

Veremos a continuación que si interesa realizar un test o construir un intervalo de confianza para diferencia de medias con muestras apareadas, hay que realizarlo con los mismos procedimientos desarrollados para una muestra tomando las diferencias.

23.1.2 Tests e intervalos de confianza para diferencias de medias con datos apareados.

Un **intervalo de confianza** de nivel $1 - \alpha$ para $\mu_X - \mu_Y$, $IC(\mu_X - \mu_Y)$ está dado por

$$\bar{X} - \bar{Y} \pm t_{n-1, \alpha/2} \frac{S_D}{\sqrt{n}}$$

Un **test** para la **hipótesis nula**: $H_0: \mu_X - \mu_Y = \delta$ (o equivalentemente $H_0: \mu_D = \delta$) estará basado en el siguiente

$$\text{Estadístico del test} \quad T = \frac{(\bar{X} - \bar{Y}) - \delta}{S_D / \sqrt{n}} \sim t_{n-1} \quad \text{bajo } H_0$$

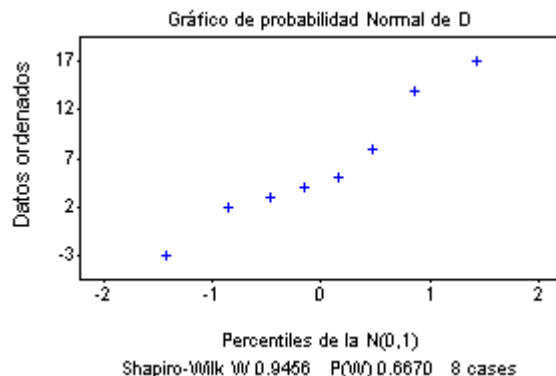
Observación: El test basado en el estadístico anterior se denomina **test t apareado** para diferencia de medias.

Región de rechazo de nivel α para cada tipo de hipótesis alternativa y su p-valor

Tipo de Hipótesis alternativa	Región de Rechazo de nivel α	p-valor
a) $H_a: \mu_X - \mu_Y \neq \delta$	$ T \geq t_{n-1, \alpha/2}$	$p = 2 P(T \geq T_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$T \geq t_{n-1, \alpha}$	$p = P(T \geq T_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$T \leq -t_{n-1, \alpha}$	$p = P(T \leq T_{\text{obs}})$

--	--	--

Ejemplo 1: Se ha observado, para 8 pacientes con problemas respiratorios crónicos, una diferencia de 6.25 mm Hg en las medias muestrales de la tensión de dióxido de carbono. Se quiere decidir si esta diferencia es estadísticamente distinta de cero a nivel $\alpha = 0.05$.



No se rechaza el supuesto de Normalidad de las diferencias (p-valor = 0.6670) que requiere el test t.

Hipótesis nula: $H_0: \mu_X - \mu_Y = 0$ no hay diferencia entre la **media** de la concentración de CO_2 **en la población** de todos los posibles pacientes con problemas respiratorios crónicos antes de realizar la dieta y después de realizarla.

El valor observado del estadístico del test es:

$$t_{\text{obs}} = \frac{(\bar{x} - \bar{y}) - 0}{\frac{s_D}{\sqrt{8}}} = \frac{6.25 - 0}{2.31} = 2.70$$

Para un test a **dos colas**, de la tabla de la distribución t con 7 (n-1) grados de libertad, obtenemos que el valor - p se encuentra entre 0.02 y 0.05. Por lo tanto **se rechaza la hipótesis de igualdad de medias** a nivel 5% ($p < 0.05$).

El cuadro siguiente muestra los resultados del test t anterior, dados por el Statistix:
Statistics -> One, Two, Multi-sample Tests -> Paired T Test

PAIRED T TEST FOR ANTES - DESPUES	
NULL HYPOTHESIS: DIFFERENCE = 0	
ALTERNATIVE HYP: DIFFERENCE <> 0	
MEAN	6.2500
STD ERROR	2.3126
LO 95% CI	0.7815
UP 95% CI	11.718
T	2.70
DF	7

P	0.0305
---	--------

Se obtiene, inmediatamente, un intervalo de 95% de confianza para la diferencia de medias de dióxido de carbono, ANTES - DESPUÉS, en pacientes sometidos a una dieta escasa en calorías: (0.78 , 11.72). No contiene al cero. Por lo tanto se rechaza la hipótesis de igualdad de medias a nivel 5% ($p < 0.05$).

Utilizando el p-valor = 0.0305 se llega a la misma conclusión, los datos proveen suficiente evidencia a nivel 0.05 para decidir que la tensión media de dióxido de carbono arterial en pacientes con problemas respiratorios crónicos se modifica mediante una dieta escasa en calorías.

23.1.3 Test del signo y el Test de rangos signados de Wilcoxon para dos muestras apareadas

El test del signo y el test de rangos signados de Wilcoxon permiten decidir si la mediana poblacional de la variable diferencia es cero o es un valor especificado. Los procedimientos son idénticos a los presentados para una muestra, aplicados a las diferencias $D_i = X_i - Y_i$ donde (X_i, Y_i) ($1 \leq i \leq n$) son las observaciones realizadas sobre n unidades experimentales.

Para el test del signo únicamente debe satisfacerse el supuesto de independencia de cada par de observaciones. Para el test de Wilcoxon deben agregarse los supuestos de continuidad simetría de la diferencia.

Ejemplo 1 cont. Las tablas siguientes muestran los resultados de aplicar el test del signo y el test de rangos signados de Wilcoxon a los datos de pacientes con problemas respiratorios crónicos.

SIGN TEST FOR ANTES - DESPUES

NUMBER OF NEGATIVE DIFFERENCES	1
NUMBER OF POSITIVE DIFFERENCES	7
NUMBER OF ZERO DIFFERENCES (IGNORED)	0

PROBABILITY OF A RESULT AS OR MORE EXTREME THAN OBSERVED	0.0352	← p-valor (una cola)
---	--------	----------------------

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

CASES INCLUDED 8 MISSING CASES 0

WILCOXON SIGNED RANK TEST FOR ANTES - DESPUES

SUM OF NEGATIVE RANKS	-2.5000
SUM OF POSITIVE RANKS	33.500
EXACT PROBABILITY OF A RESULT AS OR MORE EXTREME THAN THE OBSERVED RANKS (1 TAILED P-VALUE)	0.0117
NORMAL APPROXIMATION WITH CONTINUITY CORRECTION	2.100
TWO TAILED P-VALUE FOR NORMAL APPROXIMATION	0.0357
TOTAL NUMBER OF VALUES THAT WERE TIED	2
NUMBER OF ZERO DIFFERENCES DROPPED	0
MAX. DIFF. ALLOWED BETWEEN TIES	0.00001
CASES INCLUDED 8	MISSING CASES 0

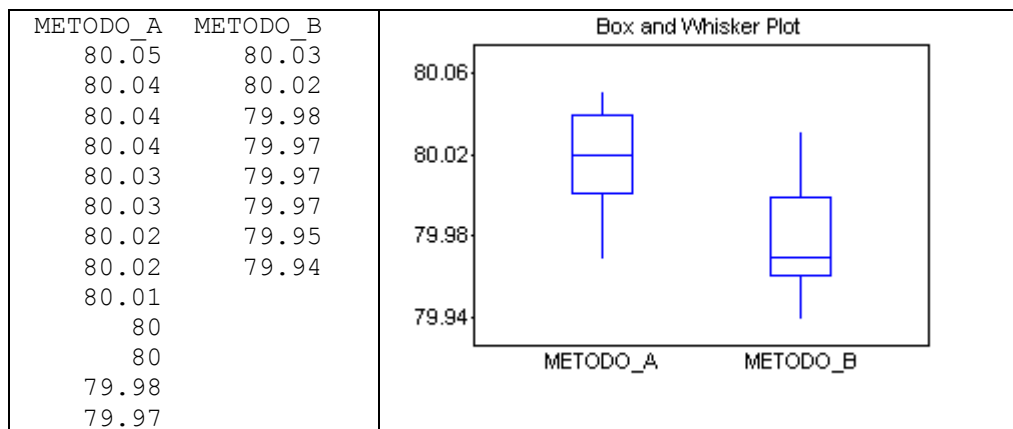
Observación.

Si los supuestos para realizar el test de t se satisfacen éste test es más potente, para un nivel dado, que las correspondientes alternativas no paramétricas. Siendo el test de rangos signados de Wilcoxon más potente que el test del signo.



**23.2 TESTS E INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE
MEDIAS DE DOS POBLACIONES NORMALES:
MUESTRAS INDEPENDIENTES - VARIANZAS IGUALES**

Ejemplo 2: Se utilizaron dos métodos para determinar el calor latente de fusión del hielo. Interesa saber si los métodos difieren. La tabla muestra el calor total absorbido por el hielo al pasar de una temperatura de -72°C a agua 0°C en calorías por gramo de masa.



Para los datos correspondientes al método A podemos observar simetría dentro de la caja y la cola de los valores altos un poco corta. En los datos del método B hay una leve asimetría hacia los valores mayores más marcada dentro de la caja. Los dos conjuntos de datos presentan variabilidad similar. Vemos también que hay diferencias en la posición relativa de los dos conjuntos de datos, el 75% de los datos mayores del método A es mayor que el 75% de los datos menores del método B. Vamos a evaluar más formalmente esta diferencia mediante intervalos de confianza y tests de hipótesis.

Planteo general

Consideraremos el caso en que tenemos dos conjuntos de observaciones Normales independientes entre si y también entre los dos grupos, con igual varianza. Es decir que las observaciones satisfacen el siguiente modelo.

Modelo:

$$\left. \begin{array}{l} X_1, \dots, X_n \text{ i.i.d } N(\mu_X, \sigma^2) \\ Y_1, \dots, Y_m \text{ i.i.d } N(\mu_Y, \sigma^2) \end{array} \right\} \text{ independientes}$$

Entonces

$$\left. \begin{array}{l} \bar{X} \sim N(\mu_X, \sigma^2/n) \\ \bar{Y} \sim N(\mu_Y, \sigma^2/m) \end{array} \right\} \Rightarrow \bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \sigma^2(\frac{1}{n} + \frac{1}{m}))$$

Nos interesa construir tests e intervalos de confianza para $\mu_X - \mu_Y$.

a) σ^2 conocido

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0,1)$$

Por lo tanto, un **intervalo de confianza** de nivel $1 - \alpha$ para $\mu_X - \mu_Y$ está dado por

$$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sigma \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Un **test** para la **hipótesis nula**: $H_0: \mu_X - \mu_Y = \delta$ estará basado en el siguiente

Estadístico del test
$$Z = \frac{(\bar{X} - \bar{Y}) - \delta}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0,1) \text{ bajo } H_0$$

Región de rechazo de nivel α para cada tipo de hipótesis alternativa

Tipo de Hipótesis alternativa	Región de Rechazo de nivel α	p-valor
a) $H_a: \mu_X - \mu_Y \neq \delta$	$ Z \geq z_{\alpha/2}$	$p = 2 P(Z \geq Z_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$Z \geq z_{\alpha}$	$p = P(Z \geq Z_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$Z \leq -z_{\alpha}$	$p = P(Z \leq Z_{\text{obs}})$

b) σ^2 desconocido, lo debemos estimar.

Hemos supuesto que las varianzas de las dos poblaciones de interés son iguales. Por lo tanto S_X^2 y S_Y^2 son dos estimadores de σ^2 . Parece razonable obtener una única estimación tomando un promedio pesado de ambos estimadores. Le daremos más peso al estimador que se obtuvo con la mayor cantidad de observaciones.

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n+m-2}$$

Puede demostrarse que si se satisface el

Modelo:
$$\left. \begin{array}{l} X_1, \dots, X_n \text{ i.i.d } N(\mu_X, \sigma^2) \\ Y_1, \dots, Y_m \text{ i.i.d } N(\mu_Y, \sigma^2) \end{array} \right\} \text{ independientes} \quad (8)$$

entonces

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

Un **intervalo de confianza** de nivel $1 - \alpha$ para $\mu_X - \mu_Y$ está dado por

$$\bar{X} - \bar{Y} \pm t_{n+m-2, \alpha/2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Un **test** para la **hipótesis nula**: $H_0: \mu_X - \mu_Y = \delta$ estará basado en el siguiente

Estadístico del test $T = \frac{(\bar{X} - \bar{Y}) - \delta}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \text{ bajo } H_0$

Observación: El test basado en el estadístico anterior se denomina **test t** para diferencia de medias

Región de rechazo de nivel α para cada tipo de hipótesis alternativa y su p-valor

Tipo de Hipótesis alternativa	Región de Rechazo de nivel α	p-valor
a) $H_a: \mu_X - \mu_Y \neq \delta$	$ T \geq t_{n+m-2, \alpha/2}$	$p = 2 P(T \geq T_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$T \geq t_{n+m-2, \alpha}$	$p = P(T \geq T_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$T \leq -t_{n+m-2, \alpha}$	$p = P(T \leq T_{\text{obs}})$

Observaciones

- La hipótesis nula más utilizada es $H_0: \mu_X - \mu_Y = 0$.
- Cuando no se rechaza $H_0: \mu_X - \mu_Y = 0$ decimos que no existe evidencia a nivel α para decidir que las medias muestrales difieren significativamente.
- Cuando sí se rechaza $H_0: \mu_X - \mu_Y = 0$ a favor de
 - a) $H_a: \mu_X - \mu_Y \neq 0$ -> decimos que las medias muestrales difieren significativamente.
 - b) $H_a: \mu_X - \mu_Y > 0$ -> decimos que la media muestral del grupo (X) es significativamente mayor que la del grupo (Y).

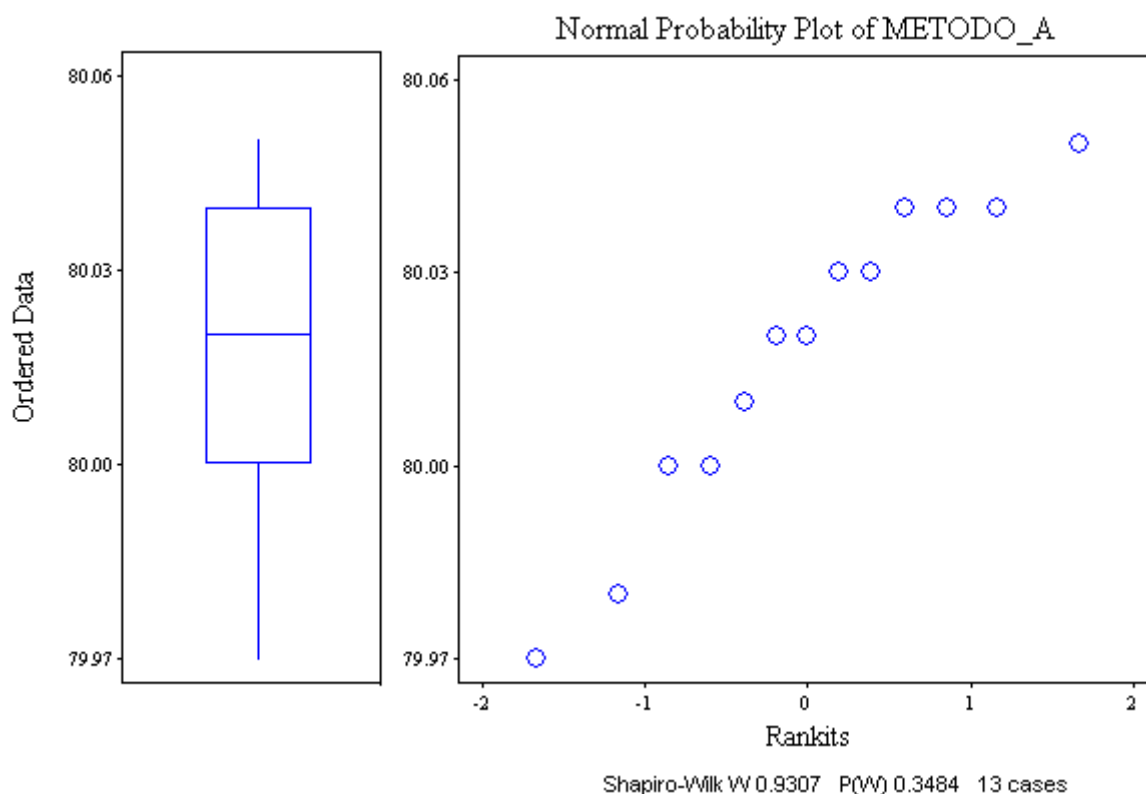
Volvamos al ejemplo 2 del calor de fusión del hielo. Sean

X_i = calor total absorbido por el hielo en la repetición i -ésima
cuando se utiliza el método A

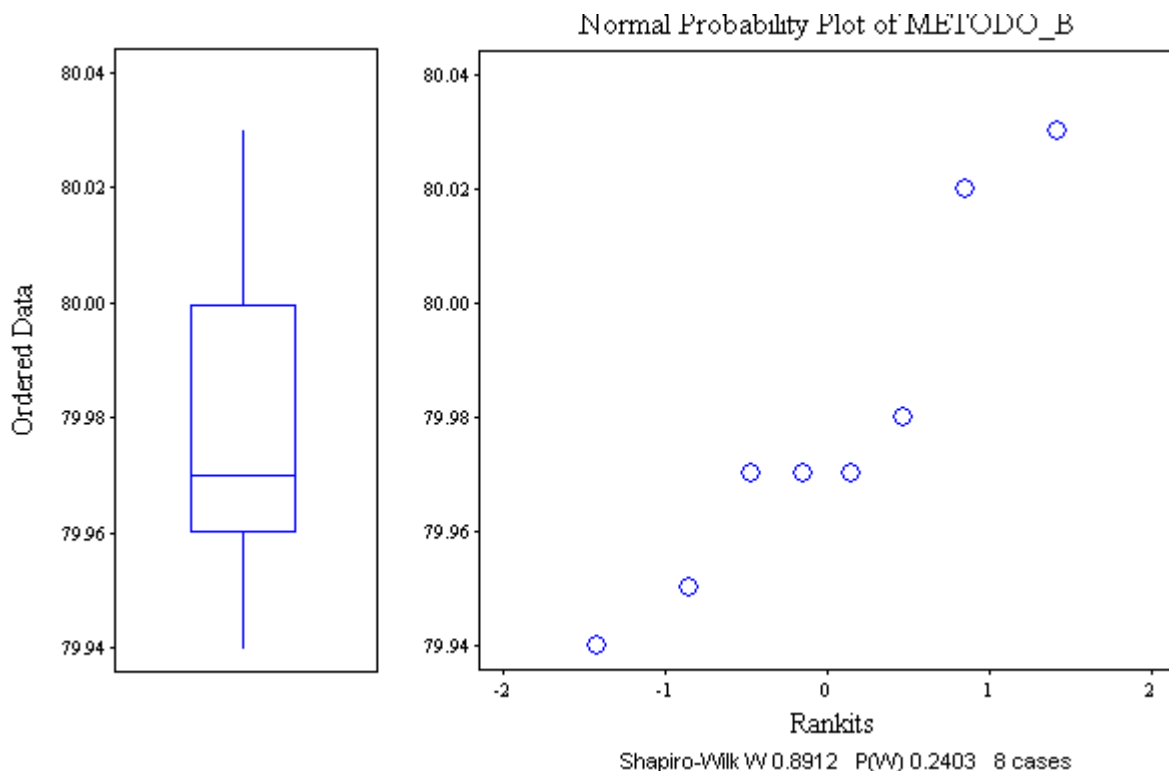
Y_i = calor total absorbido por el hielo en la repetición i -ésima
cuando se utiliza el método B

Supongamos a demás que las repeticiones se realizan en condiciones indep. e idénticas.

Supuesto de normalidad de los datos. Hemos visto que los datos del método B tienen una leve asimetría hacia los valores mayores y los de A la asimetría es más leve aún y hacia los valores menores. ¿Pueden estos alejamientos de la distribución Normal deberse a variaciones aleatorias?



Los p-valores de los estadísticos de Shapiro-Wilk resultan mayores que 0.20 para los dos métodos, no rechaza el supuesto de Normalidad de para cada conjunto de datos.



Supuesto de homogeneidad de varianzas. Los boxplots nos permiten suponer que los dos conjuntos de datos provienen de poblaciones con igual varianza. Evaluaremos posteriormente este supuesto mediante un test de hipótesis.

Podemos realizar un test t para muestras independientes. Las hipótesis de interés en este problema son:

$$H_0: \mu_X - \mu_Y = 0 \quad H_a: \mu_X - \mu_Y \neq 0$$

Con el Statistix (Statistics -> Summary Statistics -> Descriptive Statistics) obtenemos la media y el desvío de los datos correspondientes cada uno de los métodos.

VARIABLE	N	MEAN	SD
METODO_A	13	80.018	0.0245
METODO_B	8	79.979	0.0314

De la tabla anterior resulta:

$$n=13 \quad m=8 \quad \bar{x} = 80.018, \quad \bar{y} = 79.979, \quad s_x = 0.0245 \quad s_y = 0.0314$$

Por lo tanto

$$s_p^2 = \frac{12s_X^2 + 7s_Y^2}{19} = 0.00074235 \quad y \quad s_p = 0.02724617$$

y el valor observado del estadístico del test es

$$T_{\text{obs}} = \frac{(\bar{x} - \bar{y}) - 0}{s_p \sqrt{\frac{1}{13} + \frac{1}{8}}} = \frac{(80.018 - 70.979) - 0}{0.027 \sqrt{\frac{1}{13} + \frac{1}{8}}} = 3.18$$

Regla de decisión a nivel $\alpha = 0.05$: Rechace H_0 si $|T| \geq t_{19,0.025} = 2.09$

Conclusión: Las medias de las observaciones del método A y del método B difieren significativamente, es decir que la diferencia observada no es atribuible al azar.

Podemos realizar el test t para diferencia de medias utilizando el Statistix para la hipótesis alternativa $H_a: \mu_X - \mu_Y \neq 0$

Statistics -> One, Two, Multi-sample Tests -> Two Sample T-Test

TWO-SAMPLE T TESTS FOR METODO_A VS METODO_B				
VARIABLE	MEAN	SAMPLE SIZE	S.D.	S.E.
METODO_A	80.018	13	0.0245	6.81E-03
METODO_B	79.979	8	0.0314	0.0111 (A)
DIFFERENCE	0.0389			
NULL HYPOTHESIS: DIFFERENCE = 0 (B)				
ALTERNATIVE HYP: DIFFERENCE <> 0				
ASSUMPTION	T	DF	P	95% CI FOR DIFFERENCE
EQUAL VARIANCES (C)	3.18	19	0.0049	(0.0133, 0.0646) (D)
UNEQUAL VARIANCES	2.99	12.3	0.0110	(0.0107, 0.0672)
TESTS FOR EQUALITY OF VARIANCES	F	NUM DF	DEN DF	P
	1.63	7	12	0.2173
CASES INCLUDED 21 MISSING CASES 5				

(A) $S.E. = S.D. / \sqrt{(\text{Tamaño de la muestra})} = 0.0314 / \sqrt{8} = 0.0111$

(B) Hipótesis.

(C) Valor observado del estadístico (T), grados de libertad (DF) y p-valor (P) del test que supone varianzas iguales para las hipótesis indicadas en (B).

(D) Intervalo de confianza para $\mu_X - \mu_Y$, basado en el estadístico que supone igualdad de varianzas.

Completaremos el análisis de esta salida después de abordar el tema de la homogeneidad de varianzas.

Test F de homogeneidad de varianzas

Este test solo es válido bajo el supuesto de Normalidad de los datos

Modelo:

$$\left. \begin{array}{l} X_1, \dots, X_n \text{ i.i.d } N(\mu_X, \sigma_X^2) \\ Y_1, \dots, Y_m \text{ i.i.d } N(\mu_Y, \sigma_Y^2) \end{array} \right\} \text{ independientes}$$

entonces

$$F = \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} \sim f_{n-1, m-1}$$

distribución F de Fisher
con n-1 grados de libertad del numerador
y m-1 grados de libertad del denominador

Hipótesis: $H_0: \sigma_X^2 = \sigma_Y^2$ versus $H_a: \sigma_X^2 \neq \sigma_Y^2$

Estadístico del test: $F = \frac{S_X^2}{S_Y^2} \sim f_{n-1, m-1}$

Región de rechazo: $F = \frac{S_X^2}{S_Y^2} \geq f_{n-1, m-1, \alpha/2}$ ó $F = \frac{S_X^2}{S_Y^2} \leq f_{n-1, m-1, 1-\alpha/2}$

donde $f_{n-1, m-1, \alpha/2}$ es el valor crítico $\alpha/2$ (deja un área $\alpha/2$ a derecha bajo la curva de densidad $f_{n-1, m-1}$).

Cuando H_0 es verdadera el estadístico tiende a tomar valores muy cercanos a 1 mientras que si H_0 es falsa, el estadístico tiende a tomar valores alejados de 1.

Cuando H_0 es verdadera el estadístico del test tiene una distribución denominada F de Fisher que depende de dos parámetros: los grados de libertad del numerador ($n-1$) y los grados de libertad del denominador ($m-1$). Esta distribución está tabulada. Statistix permite calcular probabilidades bajo la misma (Statistics -> Probability Functions) y provee el p-valor para el test de igualdad de varianzas en forma automática.

P-VALOR: el p-valor de este test generalmente se considera a dos colas porque estamos tratando de detectar apartamientos de H_0 en cualquier dirección.

23.3 TESTS E INTERVALOS DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS DE DOS POBLACIONES NORMALES: MUESTRAS INDEPENDIENTES - VARIANZAS DESIGUALES

Modelo:

$$\left. \begin{array}{l} X_1, \dots, X_n \text{ i.i.d } N(\mu_X, \sigma_X^2) \\ Y_1, \dots, Y_m \text{ i.i.d } N(\mu_Y, \sigma_Y^2) \end{array} \right\} \text{ independientes}$$

Entonces

$$\left. \begin{array}{l} \bar{X} \sim N(\mu_X, \sigma_X^2/n) \\ \bar{Y} \sim N(\mu_Y, \sigma_Y^2/m) \end{array} \right\} \Rightarrow \bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, (\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}))$$

Nos interesa construir tests e intervalos de confianza para $\mu_X - \mu_Y$.

a) σ_X^2 y σ_Y^2 conocidos

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$$

Por lo tanto, un **intervalo de confianza** de nivel $1 - \alpha$ para $\mu_X - \mu_Y$ está dado por

$$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

Un **test** para la **hipótesis nula**: $H_0: \mu_X - \mu_Y = \delta$ estará basado en el siguiente

Estadístico del test bajo H_0

$$Z = \frac{(\bar{X} - \bar{Y}) - \delta}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$$

Región de rechazo de nivel α para cada tipo de hipótesis alternativa y su p-valor

Tipo de Hipótesis alternativa	Región de Rechazo de nivel α	p-valor
a) $H_a: \mu_X - \mu_Y \neq \delta$	$ Z \geq z_{\alpha/2}$	$p = 2 P(Z \geq Z_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$Z \geq z_{\alpha}$	$p = P(Z \geq Z_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$Z \leq -z_{\alpha}$	$p = P(Z \leq Z_{\text{obs}})$

b) σ_X^2 y σ_Y^2 son desconocidos los estimamos. Un **test** para la **hipótesis nula**:

$H_0: \mu_X - \mu_Y = \delta$ estará basado en el siguiente Estadístico del test de Welch

$$T^* = \frac{(\bar{X} - \bar{Y}) - \delta}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

Este estadístico fue propuesto por Welch–Satterthwaite quienes demostraron que tiene una **distribución t de Student** con **v grados de libertad aprox.** cuando $H_0: \mu_X - \mu_Y = \delta$ es verdadera. El parámetro v es siempre menor que $n+m-2$ y se calcula como la parte entera de la siguiente expresión:

$$v = \text{parte entera} \left(\frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}} \right) \quad (9)$$

Región de rechazo de nivel α para cada tipo de hipótesis alternativa y su p-valor

Tipo de Hipótesis alternativa	Región de Rechazo de nivel α	p-valor
a) $H_a: \mu_X - \mu_Y \neq \delta$	$ T^* \geq t_{v, \alpha/2}$	$p = 2 P(T^* \geq T^*_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$T^* \geq t_{v, \alpha}$	$p = P(T^* \geq T^*_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$T^* \leq -t_{v, \alpha}$	$p = P(T^* \leq T^*_{\text{obs}})$

El valor $t_{v, \alpha/2}$ es el percentil de la distribución t con v **grados de libertad** que deja a su derecha un área $\alpha/2$.

La única diferencia que tiene un test basado en el estadístico T^* con el test t para dos muestras independientes consiste en que su distribución es t aproximada y los grados de libertad v se obtienen mediante el cálculo dado en (9).

Bajo este modelo un Intervalo de Confianza de nivel aproximado $1-\alpha$ para $\mu_X - \mu_Y$ es

$$(\bar{X} - \bar{Y}) \pm t_{v, \alpha/2} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$$

Consideremos nuevamente la última porción de la salida de Statistix para los datos de calor latente de fusión

TWO-SAMPLE T TESTS FOR METODO_A VS METODO_B				
NULL HYPOTHESIS: DIFFERENCE = 0 (B)				
ALTERNATIVE HYP: DIFFERENCE <> 0				
ASSUMPTION	T	DF	P	95% CI FOR DIFFERENCE
-----	-----	-----	-----	-----
EQUAL VARIANCES (C)	3.18	19	0.0049	(0.0133, 0.0646) (D)
UNEQUAL VARIANCES (E)	2.99	12.3	0.0110	(0.0107, 0.0672) (F)
TESTS FOR EQUALITY	F	NUM DF	DEN DF	P
OF VARIANCES (G)	1.63	7	12	0.2173 (H)
CASES INCLUDED 21 MISSING CASES 5				

- (E) Valor observado del estadístico del test de Welch que no supone igualdad de varianzas, grados de libertad y p-valor, para las hipótesis indicadas en (B).
- (F) Intervalo de confianza para $\mu_X - \mu_Y$, basado en el estadístico que no supone igualdad de varianzas.
- (G) Valor del estadístico del test para $H_0: \sigma_X^2 = \sigma_Y^2$ versus $H_a: \sigma_X^2 \neq \sigma_Y^2$
- (H) No se rechaza H_0 ($p=0.2173$) por lo tanto, es correcto haber considerado el test para varianzas iguales cuyos resultados se muestran en (C) y (D).