

Retomemos el ejemplo (págs. 132 - 134) en el que interesa decidir si un espectrofotómetro está calibrado y se obtienen 5 determinaciones de un gas estándar cuya concentración de CO es de 70 ppm (datos: 78, 83, 68, 72, 88). Vimos que el test t de nivel $\alpha=0.05$ realizado resultó en no rechazo de la hipótesis nula: $H_0: \mu = 70$ para la hipótesis alternativa bilateral $H_a: \mu \neq 70$ y con el mismo test sí se rechazó H_0 cuando consideramos la hipótesis alternativa unilateral $H_a: \mu > 70$.

Las tablas siguientes muestran los resultados obtenidos al realizar los mismos tests con el Statistix: **Statistics -> One, Two, Multi-Sample Tests -> One-Sample T test**

ONE-SAMPLE T TEST FOR CONC		ONE-SAMPLE T TEST FOR CONC
NULL HYPOTHESIS: MU = 70	← Hipótesis nula →	NULL HYPOTHESIS: MU = 70
ALTERNATIVE HYP: MU <> 70	← Hipótesis alternativa →	ALTERNATIVE HYP: MU > 70
MEAN 77.800	← Media muestral →	MEAN 77.800
STD ERROR 3.6111	← s / \sqrt{n} →	STD ERROR 3.6111
MEAN - H0 7.8000	← Media muestral - valor nulo →	MEAN - H0 7.8000
LO 95% CI -2.2260	← Extremo inferior del I. C. nivel 95% →	LO 95% CI -2.2260
UP 95% CI 17.826	← Extremo superior del I. C. nivel 95% →	UP 95% CI 17.826
T 2.16	← Valor observado del estadístico del test t →	T 2.16
DF 4	← Grados de libertad de la distribución t →	DF 4
P 0.0969	← p-valor →	P 0.0484
CASES INCLUDED 5		CASES INCLUDED 5

El p-valor correspondiente al test bilateral 0.0969 (> 0.05) se corresponde con el no rechazo de la hipótesis nula a un nivel del 5%. Este resultado es equivalente al que se obtiene del intervalo con nivel de confianza del 95% para el error sistemático (sesgo) de las mediciones del espectrofotómetro: (-2.2260, 17.826). El cero pertenece a dicho intervalo, por lo tanto concluimos que los datos no proveen suficiente evidencia a nivel 0.5 para decidir que las mediciones del espectrofotómetro tienen sesgo. Las únicas diferencias entre el test bilateral y el unilateral se encuentran en la hipótesis alternativa y el p-valor. El intervalo de confianza es únicamente equivalente al test bilateral.

Observación: Varios de los procedimientos que hemos visto para el caso en que la muestra es pequeña requieren del supuesto de Normalidad de los datos. Para esos métodos si los datos presentan una *fuerte asimetría*, tienen *valores atípicos* o tienen

un histograma *marcadamente diferente* de la curva Normal en algún sentido, el test NO ES VÁLIDO. Lo mismo ocurre con el intervalo de confianza.

10. EVALUACIÓN DEL SUPUESTO DE NORMALIDAD

Dado un conjunto de datos consideraremos dos estrategias para evaluar si es razonable suponer que en la población de la cual proviene la muestra, la variable de interés tiene una distribución aproximadamente normal.

Métodos Gráficos

- Box-plot
- Histograma
- Gráfico de tallo-hojas
- Gráfico de Probabilidad normal (Q-Q plot)

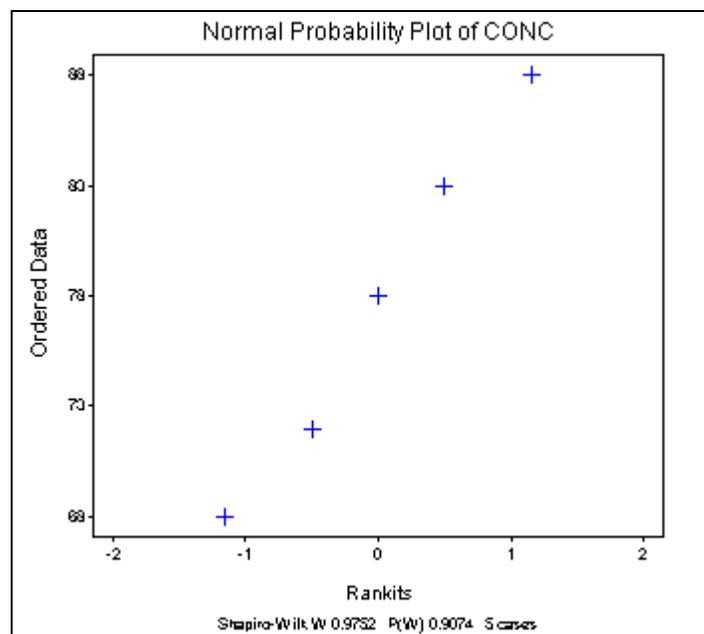
Métodos Analíticos (Tests de hipótesis)

- Test de Shapiro-Wilk
- Otros tests que no detallaremos tales como: Lilliefords, Kolmogorov-Smirnov, etc.

En general, se decidirá si no existen alejamientos GROSEROS de la distribución normal. Si alguno de los métodos muestra claramente que la distribución de los datos no puede suponerse normal, entonces, habrá que abandonar los métodos estadísticos que tienen como supuestos esta distribución.

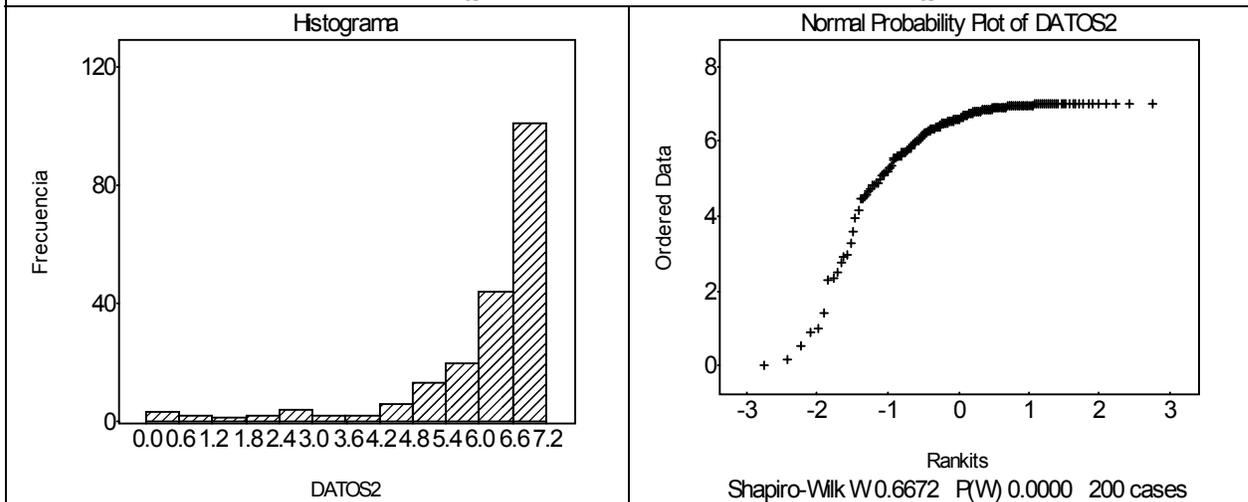
Gráfico de Probabilidad normal (Q-Q plot)

La figura siguiente muestra el gráfico de probabilidad normal para los datos de concentración de CO (78, 83, 68, 72, 88).



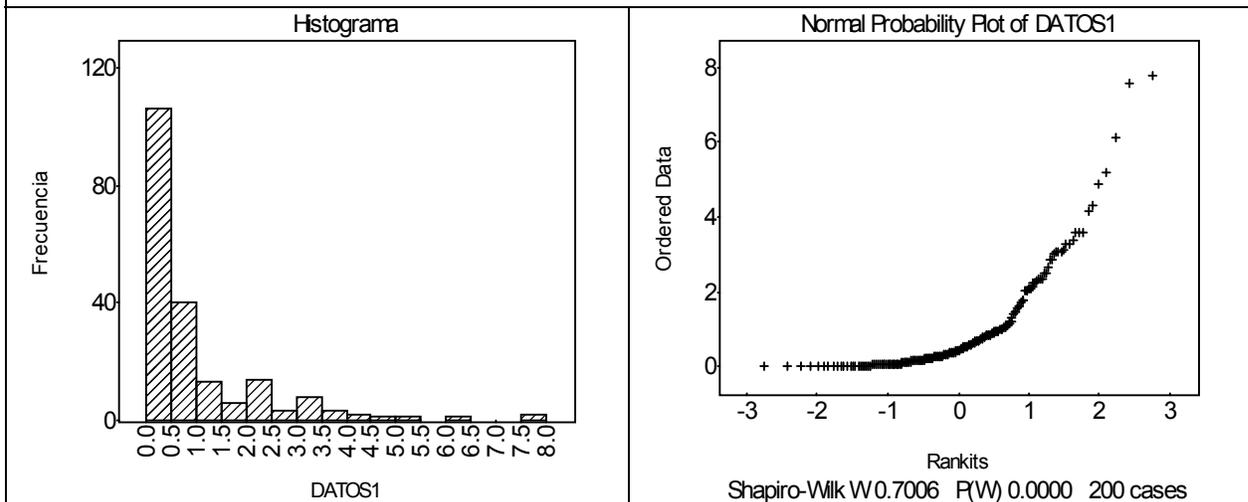
El gráfico de probabilidad normal es un diagrama de dispersión de los percentiles empíricos (datos ordenados) versus los percentiles teóricos de la distribución Normal. Si la muestra proviene de una distribución normal los puntos se encontrarán, salvo por fluctuaciones aleatorias, sobre una recta. Apartamientos de la distribución normal producen diferentes curvaturas que sugieren qué tipo de distribución puede tener la variable. Para los datos de concentración de CO no se observan importantes alejamientos de la linealidad. Las figuras siguientes muestran los histogramas de conjuntos de datos con diferentes alejamientos de la Normalidad y sus correspondientes gráficos de probabilidad normal.

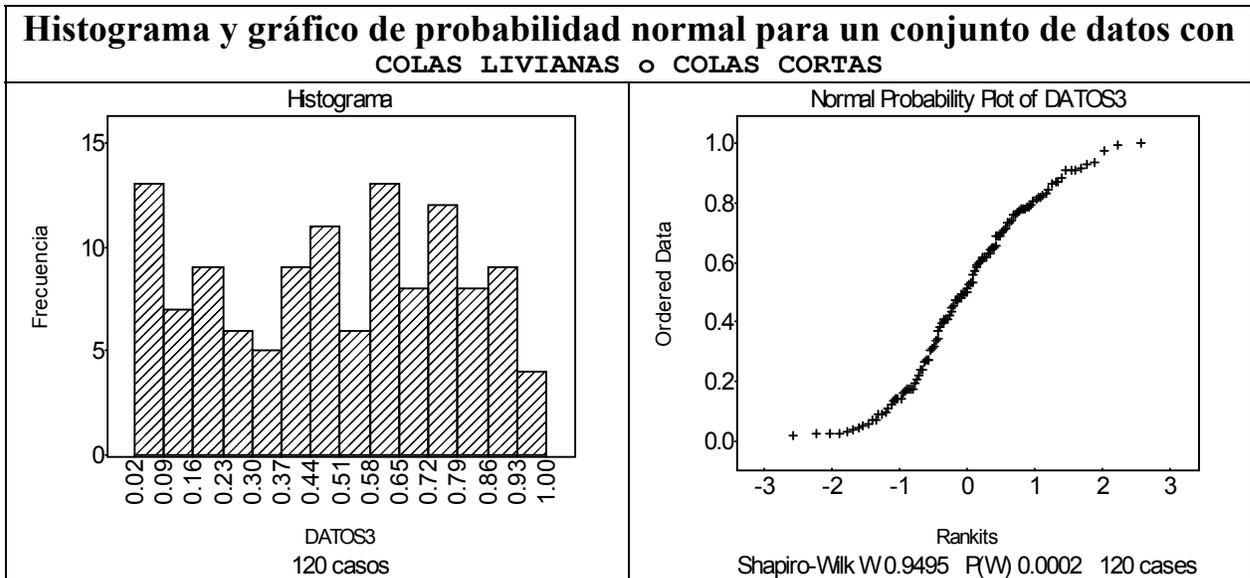
Histograma y gráfico de probabilidad normal para un conjunto de datos con ASIMETRÍA A IZQUIERDA o COLA PESADA A IZQUIERDA



Los gráficos de probabilidad normal se obtienen mediante: Statistics -> Randomness-Normality test ->Normal Probability plot

Histograma y gráfico de probabilidad normal para un conjunto de datos con ASIMETRÍA A DERECHA o COLA PESADA A DERECHA





Test de Shapiro-Wilk

Las hipótesis del test son:

H_0 : la variable tiene distribución normal (con cualquier media y varianza)

H_a : la variable no tiene distribución normal

El estadístico del test puede interpretarse como una medida de la asociación lineal entre los percentiles observados en la muestra y los percentiles teóricos de la normal que muestra el gráfico de probabilidad normal. El Statistix muestra el valor del estadístico del test y el correspondiente p-valor en el q-q plot. También se lo obtiene del siguiente modo:

Statistics -> Randomness-Normality test

-> Shapiro-Wilk Normality test

Para los datos correspondientes a las 5 determinaciones de la concentración de CO obtenemos

SHAPIRO-WILK NORMALITY TEST

VARIABLE	N	W	P
CONC	5	0.9752	0.9074

Conclusión: El p-valor = 0.9074 (> 0.05) es altísimo. No hay evidencia suficiente para rechazar la hipótesis de que la distribución de la concentración de CO es normal. Cuanto mayor sea el p-valor tanto mayor será la evidencia a favor de la hipótesis nula.

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

Hemos utilizado los tests estadísticos para comparar la media desconocida de una población con algún valor conocido, fijo, μ_0 . En la práctica, es más común el problema de comparar las medias desconocidas de dos poblaciones utilizando dos grupos de observaciones. A menudo los dos grupos corresponden a mediciones realizadas con dos métodos diferentes o son los resultados de diferentes tratamientos.

Introduciremos a continuación el procedimiento que permite decidir si la diferencia observada en las medias muestrales es suficientemente grande para que no pueda ser atribuida al azar. Un test de hipótesis para dos muestras es similar en muchos aspectos al test para una muestra.

- Especificamos una hipótesis nula, en la mayoría de los casos esta propone que las medias de las dos poblaciones son iguales y establecemos si estamos interesados en una hipótesis uni o bilateral.
- Especificamos un nivel de significación α .
- Calculamos el p-valor: la probabilidad de obtener un par de muestras con medias que difieren tanto o más que la diferencia observada cuando H_0 es verdadera. Si esta probabilidad es pequeña (menor que α) rechazamos H_0 y concluimos que las medias de las dos poblaciones son diferentes.

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

11.1 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.

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

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

Un diseño apareado permite frecuentemente controlar fuentes de variación que podrían de otro modo influenciar los resultados de la comparación y reducir la precisión de la comparación.

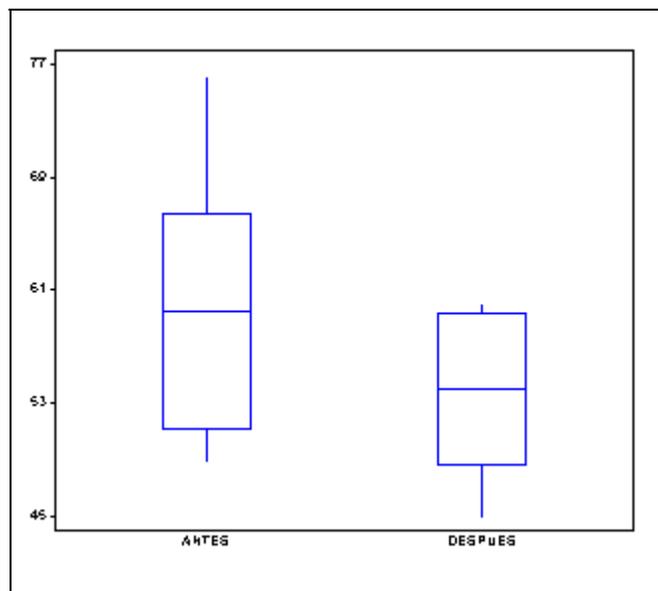
Ejemplo: Interesa decidir si una dieta escasa en calorías produce un cambio en la tensión media de dióxido de carbono arterial en pacientes con problemas respiratorios crónicos

**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?

- Los resultados de los distintos sujetos son muy variables.
- 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

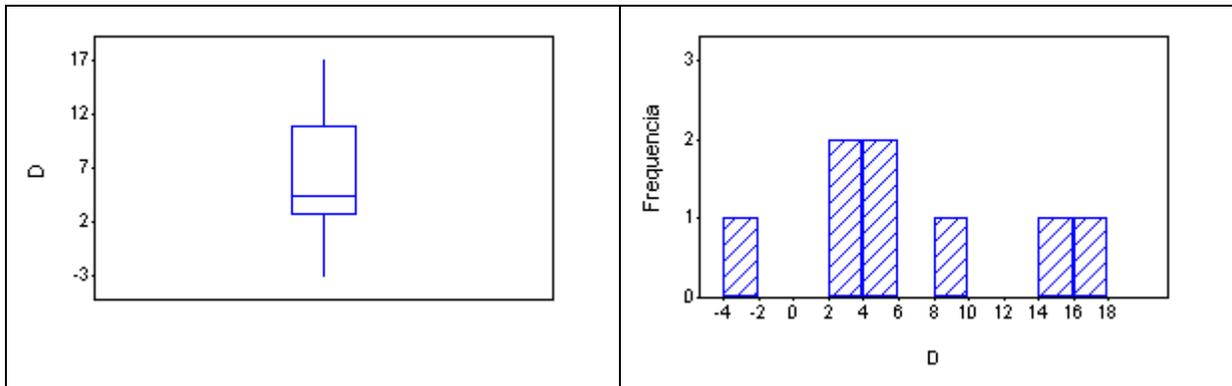
Definimos una nueva variable, la diferencia entre las concentraciones de un paciente.

$$D_i = X_i - Y_i.$$

$D_i > 0 \Rightarrow$ 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 esta variable observamos que en la mayoría de los pacientes la concentración de CO_2 se redujo al realizar la dieta

PLANTEO GENERAL

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

Modelo, con $\mu_D = \mu_X - \mu_Y$

suponemos que las diferencias de las observaciones se distribuyen Normalmente. No realizamos ningún supuesto sobre la distribución de cada variable, solamente sobre la diferencia. Luego

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

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

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.

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

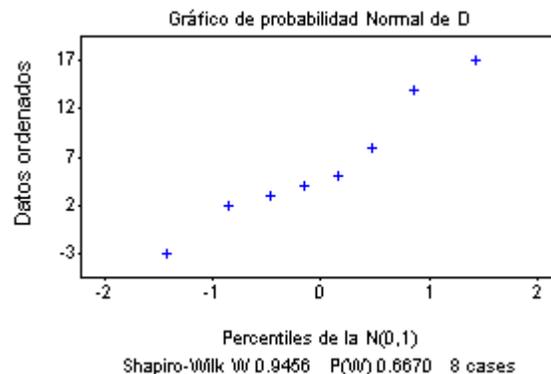
Estadístico del test $T = \frac{(\bar{X} - \bar{Y}) - \delta}{S_D / \sqrt{n}} \sim t_{n-1}$ bajo H_0

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 > t_{n-1, \alpha/2}$	$p = 2 P(T > T_{obs})$
b) $H_a: \mu_X - \mu_Y > \delta$	$T > t_{n-1, \alpha}$	$p = P(T > T_{obs})$
c) $H_a: \mu_X - \mu_Y < \delta$	$T < -t_{n-1, \alpha}$	$p = P(T < T_{obs})$

Ejemplo- continuación: 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 $\alpha = 0.05$.



Se satisface el supuesto de Normalidad del test t (p-valor = 0.6670).

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% ($p < 0.05$).

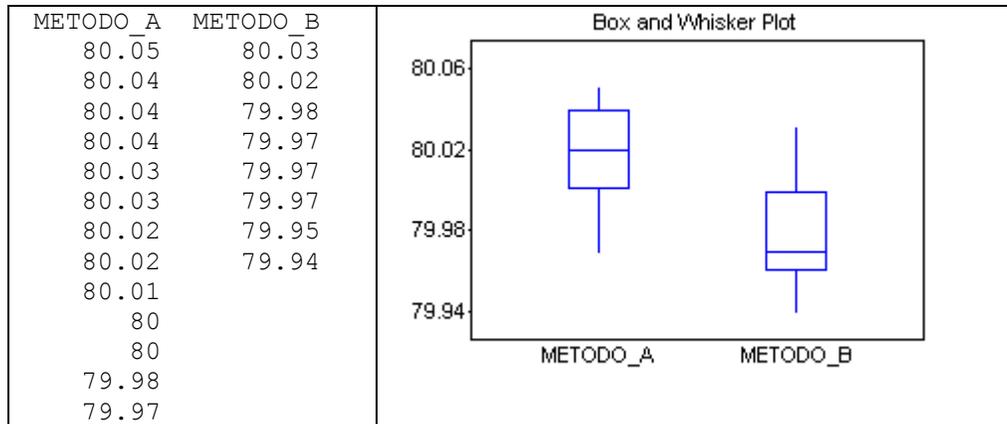
El cuadro siguiente muestra los resultados del test t anterior, dados por el Statistix:
Statistics -> One, Two, Multi-sample Tests -> Paired T Test

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. Mediante 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.

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

Ejemplo: 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.



Podemos observar simetría dentro de la caja y una leve asimetría hacia los valores grandes en las colas para los datos correspondientes al método A. En los datos del método B hay una leve asimetría hacia los valores menores 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 > z_{\alpha/2}$	$p = 2 P(Z > Z_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$Z > z_{\alpha}$	$p = P(Z > Z_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$Z < -z_{\alpha}$	$p = P(Z < Z_{\text{obs}})$

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 > t_{n+m-2, \alpha/2}$	$p = 2 P(T > T_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$T > t_{n+m-2, \alpha}$	$p = P(T > T_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$T < -t_{n+m-2, \alpha}$	$p = P(T < T_{\text{obs}})$

Observaciones

- La hipótesis nula más utilizada es $H_0: \mu_X - \mu_Y = 0$.
- Si 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.
- Si 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).

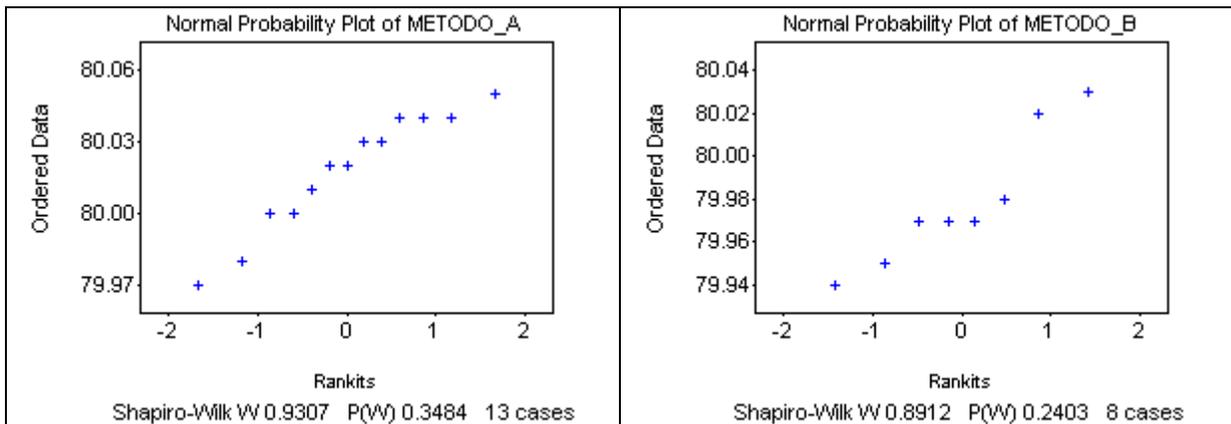
Volvamos al ejemplo del calor de fusión del hielo. Sean

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 además 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 ambos mayores que 0.20, no se rechaza el supuesto de Normalidad de los 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

$$s_p^2 = \frac{12s_x^2 + 7s_y^2}{19} = 0.00074235 \quad \text{y} \quad s_p = 0.02724617$$

y el valor observado del estadístico del test es

$$T_{\text{obs}} = \frac{(\bar{x} - \bar{y}) - 0}{s_p \sqrt{\frac{1}{13} + \frac{1}{8}}} = \frac{(80.018 - 79.979) - 0}{0.027 \sqrt{\frac{1}{13} + \frac{1}{8}}} = 3.18$$

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

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: $F = \frac{S_X^2}{S_Y^2} > f_{n-1, m-1, \alpha/2}$ ó $F = \frac{S_X^2}{S_Y^2} < f_{n-1, m-1, 1-\alpha/2}$

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.

Consideremos nuevamente la última porción de la salida de Statistix para los datos de calor 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 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 (B) y (C).

Cuando las variables tienen distribución normal pero no es posible suponer que las varianzas son iguales es posible usar una modificación del test t que presentamos a continuación.

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

Modelo:

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

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

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 > z_{\alpha/2}$	$p = 2 P(Z > Z_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$Z > z_{\alpha}$	$p = P(Z > Z_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$Z < -z_{\alpha}$	$p = P(Z < Z_{\text{obs}})$

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 ν **grados de libertad aprox.** cuando $H_0: \mu_X - \mu_Y = \delta$ es verdadera. El parámetro ν es siempre menor que $n+m-2$ y se calcula como la parte entera de la siguiente expresión (poco amigable por cierto):

$$\nu = \text{parte entera} \left(\frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m} \right)^2}{\left(\frac{s_X^2}{n} \right)^2 + \left(\frac{s_Y^2}{m} \right)^2} - 2 \right) \quad (9)$$

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^* > t_{\nu, \alpha/2}$	$p = 2 P(T^* > T^*_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$T^* > t_{\nu, \alpha}$	$p = P(T^* > T^*_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$T^* < -t_{\nu, \alpha}$	$p = P(T^* < T^*_{\text{obs}})$

El valor $t_{\nu, \alpha/2}$ es el percentil de la distribución t con ν **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 ν 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

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

12 MUESTRAS GRANDES

12.1 Caso general

$$\text{Modelo: } \left. \begin{array}{l} X_1, \dots, X_n \text{ i.i.d} \\ Y_1, \dots, Y_m \text{ i.i.d} \end{array} \right\} \text{ independientes con } \begin{cases} E(X_i) = \mu_X & \text{Var}(X_i) = \sigma_X^2 \\ E(Y_i) = \mu_Y & \text{Var}(Y_i) = \sigma_Y^2 \end{cases}$$

Si n y m son suficientemente grandes

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

Por lo tanto un **Intervalo de Confianza de nivel aproximado $1-\alpha$** para $\mu_X - \mu_Y$ está dado por

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

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

$$\text{Estadístico del test } Z = \frac{(\bar{X} - \bar{Y}) - \delta}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \stackrel{a}{\sim} N(0,1) \text{ bajo } H_0$$

Región de rechazo de **nivel aproximado 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 > z_{\alpha/2}$	$p = 2 P(Z > Z_{\text{obs}})$
b) $H_a: \mu_X - \mu_Y > \delta$	$Z > z_{\alpha}$	$p = P(Z > Z_{\text{obs}})$
c) $H_a: \mu_X - \mu_Y < \delta$	$Z < -z_{\alpha}$	$p = P(Z < Z_{\text{obs}})$

12. 2 Tests e intervalos de confianza para la diferencia de proporciones

Modelo:
$$\left. \begin{array}{l} X_1, \dots, X_n \text{ i.i.d Bi}(1, p_X) \\ Y_1, \dots, Y_m \text{ i.i.d Bi}(1, p_Y) \end{array} \right\} \text{ independientes}$$

Si n y m son suficientemente grandes

$$\frac{(\bar{X} - \bar{Y}) - (p_X - p_Y)}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{\bar{Y}(1-\bar{Y})}{m}}} \stackrel{a}{\sim} N(0,1)$$

Por lo tanto un **Intervalo de Confianza de nivel aproximado $1-\alpha$** para $p_X - p_Y$ está dado por

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

Un **test** para la **hipótesis nula**: $H_0: p_X - p_Y = 0$ estará basado en el siguiente

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

Como bajo H_0 la varianza de las X 's es igual a la varianza de las Y 's para estimar la varianza común se utiliza un promedio pesado de los estimadores de las varianzas en cada grupo, esto es:

$$S_p^2 = \frac{n\bar{X}(1-\bar{X}) + m\bar{Y}(1-\bar{Y})}{n+m-2}$$

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

Tipo de Hipótesis alternativa	Región de Rechazo de nivel α	p-valor
a) $H_a: p_X - p_Y \neq 0$	$ Z > z_{\alpha/2}$	$p = 2 P(Z > Z_{\text{obs}})$
b) $H_a: p_X - p_Y > 0$	$Z > z_{\alpha}$	$p = P(Z > Z_{\text{obs}})$
c) $H_a: p_X - p_Y < 0$	$Z < -z_{\alpha}$	$p = P(Z < Z_{\text{obs}})$