

## 5. Contrastes Estadísticos de 1 y 2 Muestras

### Análisis de caso: Las ventas de pan tostado.

Nuestra empresa produce y comercializa productos de panadería y bollería industrial para toda la península. En particular en la Comunidad Valenciana, Islas Baleares y Murcia (Zona 1 Valencia), y en ambas Castillas (Zona 2 Centro) lo hace a través de 200 empresas de distribución (100 en cada zona) que representan otras tantas subáreas geográficas distintas. Vamos a introducir un nuevo tipo de pan tostado y queremos evaluar el impacto de la introducción sobre las ventas totales de pan tostado de la competencia. Nuestro propósito es introducir el nuevo producto durante 1 año en una de ambas zonas, previa campaña publicitaria, y evaluar si se ha producido un cambio significativo en las ventas totales de pan tostado de la competencia. Pero antes vamos a comprobar si existen diferencias significativas entre ambas zonas en la variable "ventas de la competencia de pan tostado" antes de la introducción de nuestro nuevo producto.

#### 1. Comprobando si dos zonas difieren significativamente en ventas.

**Pruebas t para muestras independientes: t de Student y t de Welch.  
El test de Levene de igualdad de varianzas.**

Instrucciones. Programa 3D.

```

/INPUT
  TITLE IS '3D. T-TWOGROUP. INTRODUCCIÓN DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE = 'C:\E1\3D001.DAT'.
  VARIABLES = 4.
  CASE=20.
  FORMAT IS FREE.
  # La cuestion que se quiere averiguar es si existen diferencias
  # significativas en ventas entre la zona 1 y la zona 2.
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1.
  #Distribu: es el nº de distribuidor (Representa una subárea geográfica definida).
  #Zona es la Zona geográfica (1=Valencia), (2=Centro).
  #Numclien es el nº de clientes de nuestro distribuidor en esa subárea (en miles).
  #Ventas1 representa las ventas en millones del último año de toda la competencia.
/GROUP
  VARIABLE = ZONA.
  CODES(ZONA)= 1, 2.
  NAMES(ZONA)=VALENCIA, CENTRO.
/TWOGROUP
  VARIABLE=VENTAS1.
/PRINT
  LINESIZE=80.
/PLOT
  FILE='C:\E1\3D001.PLT'.
/END
    
```

Los datos están en este archivo.

Para pruebas t "entre grupos" (muestras independientes), aquí se especifica la variable que clasifica los casos en 2 grupos.

Especifica la o las variables en que se va a comparar a los grupos.

Especifica el nº de caracteres por línea: 80 para monitor.

Especifica el archivo donde se guardarán los gráficos en alta resolución. Luego se abren e imprimen con PLOT desde el menú principal.

Datos:

1	1	10	1234	11	2	9	1134
2	1	7	1567	12	2	8	1589
3	1	13	1889	13	2	12	1799
4	1	6	998	14	2	8	1100
5	1	21	2399	15	2	20	2311
6	1	8	1800	16	2	7	1701
7	1	12	1709	17	2	4	804
8	1	4	799	18	2	12	1745
9	1	34	4782	19	2	7	1005
10	1	5	971	20	2	2	105

(Sigue en la Columna adjunta)

Los datos se presentan aquí a doble columna, pero, en el BMDP hay que introducirlos en una única columna

Resultados:

BMDP Program Output File: C:\E1\3D001.OUT. BMDP3D - T-TESTS Release: 7.  
 Prof. Dr. José Luis Meliá Navarro

NUMBER OF CASES READ. . . . . 20

GROUPING VARIABLE. . . ZONA	CATEGORY	FREQUENCY
	VALENCIA	10
	CENTRO	10

DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO.	TOTAL FREQ.	MEAN	STANDARD DEV.	ST.ERR OF MEAN	COEFF OF VAR
3 NUMCLIE	20	10.450	7.3662	1.6471	.70490
4 VENTAS1	20	1572.1	937.08	209.54	.59609

La media de ventas en las 20 subáreas es 1572 millones/año

VARIABLE NO.	NAME	S M A L L E S T			L A R G E S T			RANGE
		VALUE	Z-SCR	CASE	VALUE	Z-SCR	CASE	
3	NUMCLIE	2.0000	-1.15	20	34.000	3.20	9	32.000
4	VENTAS1	105.00	-1.57	20	4782.0	3.43	9	4677.0

Se detecta un valor muy alto en ventas con una puntuación típica muy elevada.

VENTAS1 VARIABLE NUMBER 4  
 \*\*\*\*\*

GROUP	1 VALENCIA	2 CENTRO
H		X
H H		X X
HHHHH	H	X X X X
M-----M		
I AN H= 1 CASES A		I AN X= 1 CASES A
N (N= 10) X		N (N= 10) X

En la Zona Valencia una subárea tiene unas ventas substancialmente más altas que el resto. Este dato puede considerarse un outlier y distorsionará la comparación de medias entre las dos zonas.

GROUP	VALENCIA	CENTRO	TEST STATISTICS	P-VALUE	DF
MEAN	1814.8000	1329.3000	LEVENE F FOR VARIABILITY	0.59 0.4524	1, 18
STD DEV	1153.3635	626.5875	POOLED T	1.17 0.2574	18
S.E.M.	364.7256	198.1443	SEPARATE T	1.17 0.2618	13.9
SAMPLE SIZE	10	10			
MAXIMUM	4782.0000	2311.0000			
MINIMUM	799.0000	105.0000			
Z MAX	2.57	1.57			
Z MIN	-0.88	-1.95			
CASE (MAX)	9	15			
CASE (MIN)	8	20			

No hay dif. signific. en variabilidad.

La zona Valencia presenta una media y una Desviación Típica mucho mayor, en parte debido al valor del outlier.

No hay dif. Sig. En las medias de ventas1.

Copying Plots from workfile to C:\E1\3D001.PLT

Los resultados obtenidos muestran que, aunque las zonas 1 y 2 presentan medias de ventas claramente diferenciadas, no hay diferencias significativas entre ambas zonas. La presencia de algún valor muy elevado recomienda volver realizar el análisis mediante un procedimiento robusto.

En este análisis hemos realizado una prueba t entre dos grupos o entre muestras independientes, . Esta prueba t se presenta en dos versiones. "Pooled T" es la t clásica de Student que asume que las varianzas de ambos grupos en la VD son iguales en la población y cuyo error típico de la diferencia en el denominador de la fórmula de t contiene un estimador conjunto de la varianza en ambos grupos. "Separate T" es la prueba t de Welch que no hace el supuesto adicional de igualdad de varianzas entre grupos. Esta prueba tenderá a ser demasiado conservadora si los grupos tienen efectivamente la misma varianza en la población. El test de Levene de igualdad de varianzas ayuda a tomar la decisión sobre que tipo de prueba t interpretar. Este test permite decir si los dos grupos difieren significativamente en su varianza. Si resulta significativo (p<0'05) es razonable inclinarse por la t de Welch (separate t). Si no resulta significativo podría pensarse en considerar la prueba t de Student, sin embargo, no poder rechazar la hipótesis nula de igualdad de varianzas no equivale a probar la igualdad de varianzas. Puede utilizarse la t de Student si se sabe que las varianzas de ambos grupos en la población es igual.

**Pruebas t recortadas.  
Test no-paramétrico de la suma de rangos de Wilcoxon (Mann-Whitney).**

**Instrucciones.**

```

/INPUT
  TITLE IS '3D. T-TWOGROUP. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE = 'C:\E1\3D001.DAT'.
  VARIABLES = 4.
  CASE=20.
  FORMAT IS FREE.
  # La cuestion que se quiere averiguar es si existen diferencias
  # significativas en ventas entre la zona 1 y la zona 2.
  # Ahora se eliminaran outliers y se utilizara un test no-paramétrico.
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIE, VENTAS1.
/GROUP
  VARIABLE = ZONA.
  CODES(ZONA) = 1, 2.
  NAMES(ZONA) = VALENCIA, CENTRO.
/TWOGROUP
  VARIABLE = VENTAS1.
  ROBUST.
  NONPAR.
/PRINT
  LINESIZE=80.
/PLOT
  FILE='C:\E1\3D002.PLT'.
/END
    
```

Solicita pruebas t recortadas (trimmed) excluyendo el valor más alto y más bajo de cada uno de los grupos.

Solicita el test de suma de rangos de Wilcoxon (Mann-Whitney) que compara los dos grupos sin asumir normalidad en las distribuciones de los grupos (aunque las distribuciones deben ser semejantes en su varianza y forma).

**Output seleccionado.**

BMDP Program Output File: C:\E1\3D002.OUT BMDP3D - T-TESTS

```

GROUP  1 VALENCIA                2 CENTRO

                X
                X X
H          H H                    X X
HHHHHHH      H          X XX X X
M-----M M-----M
I AN H=  1 CASES A I AN X=  1 CASES A
N   (N=  10)   X N   (N=  10)   X
    
```

GROUP	VALENCIA	CENTRO	TEST STATISTICS	P-VALUE	DF
MEAN	1814.8000	1329.3000	LEVENE F FOR		
TRIM MEAN	1570.8750	1359.6250	VARIABILITY	0.59 0.4524	1, 18
STD DEV	1153.3635	626.5875			
S.E.M.	364.7256	198.1443	POOLED T	1.17 0.2574	18
SAMPLE SIZE	10	10	SEPARATE T	1.17 0.2618	13.9
MAXIMUM	4782.0000	2311.0000	TRIM POOL.T	0.77 0.4567	14
MINIMUM	799.0000	105.0000	TRIM SEP. T	0.77 0.4576	13.1
Z MAX	2.57	1.57			
Z MIN	-0.88	-1.95	MANN-WHIT.	60.0 0.4497	
CASE (MAX)	9	15	(RANK SUMS	115.0	95.0)
CASE (MIN)	8	20			
2ND MAX	2399.0000	1799.0000			
2ND MIN	971.0000	804.0000			

La información sobre los grupos excluidos los valores máximos y mínimos de cada grupo (medias recortadas y segundos máximos y mínimos) muestra el efecto de estos sobre las medias (comparar con medias no recortadas en la línea superior. Las pruebas t recortadas confirman con mayor claridad la ausencia de diferencias significativas entre ambos grupos (p=0'45 aprox.)

El test no-paramétrico de Wilcoxon, basándose exclusivamente en las posiciones de rango de los valores, también muestra claramente que no puede rechazarse la hipótesis nula de diferencia entre los dos grupos.

2. Comprobando si dos zonas difieren significativamente en dos (ó más) variables simultáneamente: ventas y número de clientes.

Antes de introducir el nuevo producto vamos a realizar un análisis más para estudiar la comparabilidad de las dos zonas geográficas. Vamos a contrastar si ambas zonas difieren significativamente en las variables ventas de la competencia durante el año anterior a la introducción del producto (ventas1) y número de clientes de nuestras distribuidoras (numclien). También averiguaremos la relación entre estas variables en cada grupo. Dado que una subárea (distribuidor n° 9) de la zona Valencia presenta un valor atípicamente alto en "ventas1" que puede distorsionar los resultados vamos a excluirlo de los siguientes análisis, salvedad que debería atenderse particularmente.

**Comparando dos grupos en dos o más características cuantitativas simultáneamente: La  $T^2$  Multivariada de Hotelling y la distancia  $D^2$  de Mahalanobis.**

Instrucciones.

```

/INPUT
  TITLE IS '3D. T-TWOGROUP. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE = 'C:\E1\3D001.DAT'.
  VARIABLES = 4.
  CASE=20.
  FORMAT IS FREE.
  # La cuestion que se quiere averiguar es si existen diferencias
  # significativas en ventas y num. clientes entre la zona 1 y la zona 2.
  # Ahora se eliminara un caso con outliers
  # y se utilizara un test multivariado.
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1.
/TRANS
  IF (VENTAS1 EQ 4782) THEN (USE = 0).
/GROUP
  VARIABLE = ZONA.
  CODES(ZONA)= 1, 2.
  NAMES(ZONA)=VALENCIA, CENTRO.
/TWOGROUP
  VARIABLE=VENTAS1, NUMCLIEN.
  CORR.
  HOTEL.
/PRINT
  LINESIZE=80.
/PLOT
  FILE='C:\E1\3D003.PLT'.
/END

```

Esta instrucción suprime del análisis cualquier caso que en la variable ventas1 tenga un valor de 4782

Se especifican las dos variables dependientes en que compararemos.

Se solicitan las correlaciones entre variables en cada grupo.

Con esta instrucción se solicita el contraste entre ambos grupos en las dos variables dependientes (ventas1 y numclien) simultáneamente mediante la  $T^2$  de Hotelling y la  $D^2$  de Mahalanobis.

Output seleccionado.

BMDP Program Output File: C:\E1\3D003.OUT BMDP3D - T-TESTS  
 NUMBER OF CASES READ. . . . . 20  
 CASES WITH USE SET TO ZERO . . . . . 1  
 REMAINING NUMBER OF CASES . . . . . 19

El test multivariado de Hotelling T2 y la D2 de Mahalanobis producen un mismo valor de probabilidad para la hipótesis nula de que ambos grupos tienen igual media en la población en todas las variables contrastadas simultáneamente. En este caso no puede rechazarse la hipótesis nula de igualdad de medias.

MULTIVARIATE STATISTICS FOR GROUP VALENCIA VS. GROUP CENTRO  
 THERE ARE 9 CASES, 9 OF THEM COMPLETE IN GROUP VALENCIA  
 THERE ARE 10 CASES, 10 OF THEM COMPLETE IN GROUP CENTRO  
 NULL HYPOTHESIS IS THAT BOTH GROUPS HAVE EQUAL MEANS  
 FOR ALL VARIABLES

MAHALANOBIS D SQUARE 0.1189  
 HOTELLING T SQUARE 0.5632  
 F VALUE 0.2650 P-VALUE 0.7705  
 DEGREES OF FREEDOM 2, 16

CORRELATION MATRIX FOR GROUP 1 VALENCIA  
 VENTAS1 NUMCLIEN  
 4 3  
 VENTAS1 4 1.0000  
 NUMCLIEN 3 0.8827 1.0000

La correlación entre las variables ventas1 y numclien son aproximadamente iguales en ambos grupos.

CORRELATION MATRIX FOR GROUP 2 CENTRO  
 VENTAS1 NUMCLIEN  
 4 3  
 VENTAS1 4 1.0000  
 NUMCLIEN 3 0.8739 1.0000

VENTAS1 VARIABLE NUMBER 4  
 GROUP 1 VALENCIA 2 CENTRO  
 H H H X X X  
 HH H HHH H X XXX XXX X  
 M-----M M-----M  
 I AN H= 1 CASES A I AN X= 1 CASES A  
 N (N= 9) X N (N= 10) X

Los histogramas y los valores del contraste t entre medias varían respecto a análisis anteriores debido a la exclusión de un caso. No puede rechazarse la hipot. De igualdad de medias para ventas1.

GROUP	VALENCIA	CENTRO	TEST STATISTICS	P-VALUE	DF
MEAN	1485.1111	1329.3000	LEVENE F FOR VARIABILITY	0.25 0.6263	1, 17
STD DEV	523.1752	626.5875	POOLED T	0.58 0.5666	17
S.E.M.	174.3917	198.1443	SEPARATE T	0.59 0.5628	16.9
SAMPLE SIZE	9	10			
MAXIMUM	2399.0000	2311.0000			
MINIMUM	799.0000	105.0000			
Z MAX	1.75	1.57			
Z MIN	-1.31	-1.95			
CASE (MAX)	5	15			
CASE (MIN)	8	20			

NUMCLIEN VARIABLE NUMBER 3  
 GROUP 1 VALENCIA 2 CENTRO  
 H XX X  
 HHHH H HH X X XXX X X  
 M-----M M-----M  
 I AN H= 1 CASES A I AN X= 1 CASES A  
 N (N= 9) X N (N= 10) X

GROUP	VALENCIA	CENTRO	TEST STATISTICS	P-VALUE	DF
MEAN	9.5556	8.9000	LEVENE F FOR VARIABILITY	0.10 0.7590	1, 17
STD DEV	5.2705	4.9766	POOLED T	0.28 0.7837	17
S.E.M.	1.7568	1.5737	SEPARATE T	0.28 0.7845	16.5
SAMPLE SIZE	9	10			
MAXIMUM	21.0000	20.0000			
MINIMUM	4.0000	2.0000			
Z MAX	2.17	2.23			
Z MIN	-1.05	-1.39			
CASE (MAX)	5	15			
CASE (MIN)	8	20			

En la variable numclien tampoco puede rechazarse la hipótesis nula de igualdad de medias en la población.

### 3. Comparando las ventas de dos años consecutivos.

**Caso.**

Analizados los datos iniciales, decidimos introducir el nuevo producto en los distribuidores de la zona 1 (Valencia) (grupo experimental) durante un año. Después de introducir el nuevo producto en los distribuidores de la Zona 1 (Valencia) durante un año, reflejamos en la variable Ventas2 las ventas de la competencia en ese ejercicio en las 20 subáreas geográficas. Para simplicidad al año anterior a la introducción del nuevo producto le llamaremos año 1; a ese año corresponden los datos de ventas1, que actúan como un pretest. Para simplicidad, al año durante el cual hemos introducido el nuevo pan tostado en la zona 1, le llamaremos año 2; la variable ventas2 corresponde a las ventas de la competencia en el año 2 y constituye por tanto un postest.

**La prueba de t para muestras dependientes.**

Vamos a comparar las ventas de pan tostado de la competencia, utilizando la muestra de 20 subáreas geográficas seleccionadas para el estudio, del año 1 (ventas1) y del año 2 (ventas2). Para comparar los resultados de una misma muestra en dos ocasiones (o también para dos muestras apareadas caso a caso en la misma variable) utilizamos un tipo de prueba t conocida como prueba t para muestras dependientes (matched t). Dado que en este análisis cada una de las subáreas geográficas es el referente de comparación de sí misma en dos momentos temporales, podemos mantener la subárea 9 que tenía un valor inusualmente alto.

**Input.**

```

/INPUT
  TITLE IS '3D. T-MATCHED. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE ='C:\E1\3D002.DAT'.
  VARIABLES = 5.
  CASE=20.
  FORMAT IS FREE.
  # Se quiere averiguar es si existen diferencias
  # significativas entre ventas1 y ventas2 en las 20 subáreas.
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/MATCHED
  VARIABLES = VENTAS1, VENTAS2.
/PRINT
  LINESIZE=80.
/PLOT
  FILE='C:\E1\3D004.PLT'.
/END
    
```

Solicita la prueba t para muestras dependientes

**Datos.**

1	1	10	1234	1002
2	1	7	1567	1103
3	1	13	1889	1615
4	1	6	998	705
5	1	21	2399	1875
6	1	8	1800	1498
7	1	12	1709	1371
8	1	4	799	406
9	1	34	4782	4112
10	1	5	971	567
11	2	9	1134	1130
12	2	8	1589	1487
13	2	12	1799	1698
14	2	8	1100	1154
15	2	20	2311	2105
16	2	7	1701	1666
17	2	4	804	899
18	2	12	1745	1707
19	2	7	1005	1157
20	2	2	105	231

La última columna es la nueva variable "ventas2" que refleja las ventas totales de la competencia en cada una de las 20 subáreas seleccionadas para el estudio el año 2.

Output seleccionado.

BMDP Program Output File: C:\E1\3D004.OUT BMDP3D - T-TESTS

DESCRIPTIVE STATISTICS OF DATA

VARIABLE	TOTAL	MEAN	STANDARD DEV.	ST.ERR OF MEAN	COEFF OF VAR
NO. NAME	FREQ.				
4 VENTAS1	20	1572.1	937.08	209.54	.59609
5 VENTAS2	20	1374.4	814.48	182.12	.59261

Las ventas medias de la competencia el 2º año han sido menores.

VARIABLE	S M A L L E S T			L A R G E S T		
NO. NAME	VALUE	Z-SCR	CASE	VALUE	Z-SCR	CASE
4 VENTAS1	105.00	-1.57	20	4782.0	3.43	9
5 VENTAS2	231.00	-1.40	20	4112.0	3.36	9

VENTAS1 VS. VENTAS2 (VAR. NO. 4 VS. 5)

VENTAS1		VENTAS2		VENTAS1		VENTAS2	
H				MEAN	1572.0502	1374.4000	
H		X					
HH H		X X		STD DEV	937.0807	814.4783	
HH H		XXX		S.E.M.	209.5376	182.1229	
HH H H		XX XXXX		SAMPLE SIZE	20	20	
H HHHHHH	H	XXXXXXXX	X	MAXIMUM	4782.0000	4112.0000	
M-----M	-----M	M-----M	-----M	MINIMUM	105.0000	231.0000	
I AN H= 1 CASES A	I AN X= 1 CASES A			Z MAX	3.43	3.36	
N (N= 20) X	N (N= 20) X			Z MIN	-1.57	-1.40	
				CASE (MAX)	9	9	
				CASE (MIN)	20	20	

Los histogramas muestran la distribución de ventas1 y ventas2. Ventas2 muestra una media menor y también menor variabilidad (ver desv. típica).

VENTAS1 - VENTAS2 (VAR. NO. 4 - 5)

VENTAS1 - VENTAS2		TEST STATISTICS		P-VALUE	DF
MEAN	197.6500	MATCHED T	3.83	0.0011	19
STD DEV	230.7848				
S.E.M.	51.6050				
SAMPLE SIZE	20				
MAXIMUM	670.0000	CORRELATION	0.9750	0.0000	18
MINIMUM	-152.0000				
Z MAX	2.05				
Z MIN	-1.52				
CASE (MAX)	9				
CASE (MIN)	19				

El histograma muestra la diferencia ventas1 menos ventas2 calculada para cada caso. La diferencia media es de 197.65.

Hay diferencias significativas entre las ventas del primer y segundo año (el segundo año la competencia ha vendido 197.65 menos en promedio por subárea geográfica). Ventas1 y Ventas2 presentan un Coef. Corr. de Pearson de 0'97, estadísticamente significativo, mostrando una alta relación entre los resultados de ambos periodos.

**Pruebas t recortadas para muestras dependientes.  
Tests no-paramétricos: de los signos y de Wilcoxon.**

Como en el caso de las pruebas t para muestras independientes, en las pruebas t para muestras dependientes también es posible calcular pruebas t recortadas y el test de Wilcoxon. Además, en este caso se puede calcular también el test no paramétrico de los signos. Consecuentemente con la utilización de tests no paramétricos, la correlación entre las dos mediciones se obtiene mediante el coeficiente de correlación de Spearman.

Input.

```

/INPUT
  TITLE IS '3D. T-MATCHED. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE = 'C:\E1\3D002.DAT'.
  VARIABLES = 5. CASE=20. FORMAT IS FREE.
  # Se quiere averiguar es si existen diferencias
  # significativas entre ventas1 y ventas2 en las 20 subáreas
  # utilizando pruebas t recortadas, test de signos y de Wilcoxon
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/MATCHED
  VARIABLES = VENTAS1, VENTAS2.
  ROBUST.
  NONPAR.
/PRINT
  LINESIZE=80.
  CASE=0.
/END

```

Solicita pruebas t recortadas suprimiendo el máximo y el mínimo en cada medición.
Solicita los tests no-paramétricos: de los signos y de Wilcoxon.
Evita que nos muestre en el output los 10 primeros casos, como hace 3D por defecto.



Output seleccionado.

BMDP Program Output File: C:\E1\3D005.OUT BMDP3D - T-TESTS

VENTAS1		VENTAS2		VENTAS1		VENTAS2	
H				MEAN	1572.0502	1374.4000	
H		X		TRIM MEAN	1475.2224	1285.8334	
HH H		X X		STD DEV	937.0807	814.4783	
HH H		XXX		S.E.M.	209.5376	182.1229	
HH H H		XX XXXX		SAMPLE SIZE	20	20	
H HHHHHH	H	XXXXXXXX	X	MAXIMUM	4782.0000	4112.0000	
M-----M		M-----M		MINIMUM	105.0000	231.0000	
I AN H=	1 CASES	A I AN X=	1 CASES	Z MAX	3.43	3.36	
N (N=	20)	X N (N=	20)	Z MIN	-1.57	-1.40	
				CASE (MAX)	9	9	
				CASE (MIN)	20	20	
				2ND MAX	2399.0000	2105.0000	
				2ND MIN	799.0000	406.0000	

VENTAS1 - VENTAS2		TEST STATISTICS		P-VALUE	DF
MEAN	197.6500	MATCHED T	3.83	0.0011	19
TRIM MEAN	190.8333	TRIMMED T	3.56	0.0024	17
STD DEV	230.7848	SIGN TEST*		0.0118	
S.E.M.	51.6050	WILCOXON**	26.0	0.0032	
SAMPLE SIZE	20				
MAXIMUM	670.0000	CORRELATION	0.9750	0.0000	18
MINIMUM	-152.0000	SPEARMAN R	0.9338	0.0000	18
Z MAX	2.05				
Z MIN	-1.52	* VENTAS1 > VENTAS2 IN	16		
CASE (MAX)	9	CASES OF	20	WITH NONZERO DIFS.	
CASE (MIN)	19	** TOTAL OF RANKS WITH LESS			
2ND MAX	524.0000	FREQUENT SIGN =	26.0		
2ND MIN	-126.0000				

El output ofrece ahora las medias recortadas, el valor y probabilidad bajo la hipótesis nula de la t recortada (trimmed t y su p-value) y el valor de probabilidad (p-value) del test de signos y del de Wilcoxon. Los tres tests, así como la prueba t (matched test), revelan que puede rechazarse la hipótesis nula de igualdad de medias entre ambos años. Como muestran las medias el segundo año la competencia ha vendido un promedio de 197'65 menos por subárea que en el año previo a la introducción del producto. La correlación de Pearson (0'97) y la Spearman (0'93) muestran que el descenso en ventas es aproximadamente lineal (aproximadamente constante o proporcional para las subáreas). Ese descenso lineal en ventas apunta a factores generales que han actuado sobre las 20 subáreas, por lo que quizás no pueda atribuirse a la introducción del pan tostado en la zona 1.

4. Comparando las medias de ventas de la competencia en dos áreas (grupo "control" y grupo "experimental").

Caso.

Los resultados anteriores sugieren que ha habido un descenso significativo en las ventas de la competencia en el conjunto de las 20 subáreas geográficas bajo estudio. Quizás este descenso sea generalizado y se deba a razones coyunturales ajenas a nuestro trabajo. Es necesario contrastar si las subáreas del área 1 (Valencia) en la que hemos introducido el nuevo producto difieren significativamente en ventas2 de las del área 2 (Centro) donde no lo hemos introducido.

**Prueba t para muestras independientes con t recortada y test de Wilcoxon.**

Al tratarse de un análisis entre 2 grupos corresponde utilizar una prueba t para muestras independientes, obtendremos también la t recortada y el test no paramétrico de Wilcoxon para prevenir dificultades debidas a los valores extremos o a la forma de las distribuciones.

La forma de plantear este análisis, ahora sobre ventas2, ya ha sido estudiado.

Input

```

/INPUT
  TITLE IS '3D. T-TWOGROUP. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE ='C:\E1\3D002.DAT'. VARIABLES = 5. CASE=20. FORMAT IS FREE.
  # La cuestion que se quiere averiguar es si existen diferencias
  # significativas en ventas2 entre la zona 1 y la zona 2.
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/GROUP
  VARIABLE = ZONA. CODES(ZONA)= 1, 2. NAMES(ZONA)=VALENCIA, CENTRO.
/TWOGROUP
  VARIABLE=VENTAS2. ROBUST. NONPAR.
/PRINT
  LINESIZE=80. CASE=0.
/PLOT
  FILE='C:\E1\3D006.PLT'.
/END
    
```

Output seleccionado:

BMDP Program Output File: C:\E1\3D006.OUT BMDP3D - T-TESTS

DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO.	TOTAL FREQ.	MEAN	STANDARD DEV.	ST. ERR OF MEAN	COEFF OF VAR
4 VENTAS1	20	1572.1	937.08	209.54	.59609
5 VENTAS2	20	1374.4	814.48	182.12	.59261

VENTAS2 VARIABLE NUMBER 5

GROUP 1 VALENCIA 2 CENTRO

```

              X X
            H H H
          HHHH H      H      X XXXXX
M-----M M-----M
I AN H= 1 CASES A I AN X= 1 CASES A
N (N= 10) X N (N= 10) X
    
```

GROUP	VALENCIA	CENTRO	TEST STATISTICS	P-VALUE	DF
MEAN	1425.3999	1323.4000	LEVENE F FOR		
TRIM MEAN	1216.9999	1362.2500	VARIABILITY	1.05 0.3186	1, 18
STD DEV	1056.3223	528.0735			
S.E.M.	334.0384	166.9915	POOLED T	0.27 0.7879	18
SAMPLE SIZE	10	10	SEPARATE T	0.27 0.7890	13.2
MAXIMUM	4112.0000	2105.0000	TRIM POOL.T	-0.60 0.5603	14
MINIMUM	406.0000	231.0000	TRIM SEP. T	-0.60 0.5616	12.2
Z MAX	2.54	1.48			
Z MIN	-0.97	-2.07	MANN-WHIT.	43.0 0.5967	
CASE (MAX)	9	15	(RANK SUMS	98.0 112.0)	
CASE (MIN)	8	20			
2ND MAX	1875.0000	1707.0000			
2ND MIN	567.0000	899.0000			

No hay diferencias significativas en ventas2 entre las 10 subáreas de la zona 1 y las 10 subáreas de la zona 2.

## 5. Comparando las medias del cambio en ventas de la competencia en dos áreas.

### Caso.

El resultado anterior puede parecer desalentador.. Por un lado hemos visto que ha habido una disminución significativa general en las ventas de la competencia en el 2º año; pero, por otra parte, acabamos de ver que no hay diferencias significativas en las ventas de la competencia entre las 10 subáreas de la zona 1 donde hemos introducido el nuevo producto y las 10 subáreas de la zona 2 donde no lo hemos introducido. ¿Significa esto que la introducción del nuevo producto no ha producido ningún efecto sobre las ventas de la competencia?

Un modo de evaluar esta cuestión de una forma más precisa es estudiar el cambio en ventas entre el año 1 y el año 2. Para ello vamos a construir una variable transformada que denominaremos "cambio" igual a  $\text{ventas1} - \text{ventas2}$ . Esta variable refleja la magnitud y sentido del cambio que se ha producido en las ventas de la competencia para cada distribuidor. Si ahora comparamos las 10 subáreas del área 1 (grupo "experimental") con las 10 subáreas del área 2 (grupo "control") en la variable cambio podremos ver si hay diferencias significativas entre ambos grupos en el modo en que han variado las ventas de la competencia.

**Prueba t para muestras independientes sobre una variable tipo "cambio" ( $\text{tiempo1} - \text{tiempo2}$ ).**

Input.

```

/INPUT
  TITLE IS '3D. T-TWOGROUP. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE ='C:\E1\3D002.DAT'. VARIABLES = 5. CASE=20. FORMAT IS FREE.
  # La cuestion que se quiere averiguar es si existen diferencias
  # significativas en 'cambio' de ventas entre la zona 1 y la zona 2.
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/TRANS
  CAMBIO = VENTAS1 - VENTAS2.
/GROUP
  VARIABLE = ZONA.
  CODES(ZONA)= 1, 2.
  NAMES(ZONA)=VALENCIA, CENTRO.
/TWOGROUP
  VARIABLE=CAMBIO. ROBUST. NONPAR.
/PRINT
  LINESIZE=80. CASE=0.
/PLOT
  FILE='C:\E1\3D007.PLT'.
/END

```

Se define la variable cambio como diferencia entre  $\text{ventas1}$  y  $\text{ventas2}$ . En este tipo de variables puede definirse la resta en cualquiera de las direcciones pero esto debe tenerse en cuenta después al interpretar las diferencias.

Aquí se solicita comparar los dos grupos en la nueva variable cambio.

Output seleccionado:

BMDP Program Output File: C:\E1\3D007.OUT BMDP3D - T-TESTS

GROUPING VARIABLE. . . ZONA

CATEGORY	FREQUENCY
VALENCIA	10
CENTRO	10

DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO. NAME	TOTAL FREQ.	MEAN	STANDARD DEV.	ST.ERR OF MEAN	COEFF OF VAR
6 CAMBIO	20	197.65	230.78	51.605	1.1676

VARIABLE NO. NAME	S M A L L E S T VALUE	Z-SCR	CASE	L A R G E S T VALUE	Z-SCR	CASE	RANGE
6 CAMBIO	-152.00	-1.52	19	670.00	2.05	9	822.00

CAMBIO	VARIABLE NUMBER	6
GROUP 1	VALENCIA	
GROUP 2	CENTRO	

M	H	H	H	H	H	H	H	H	M	M	X	X	X	X	X	X	M
I	A	N	H=	1	CASES	A	I	A	N	X=	1	CASES	A				
N	(N=	10)				X	N	(N=	10)				X				

Los histogramas muestran un comportamiento muy distinto entre las dos zonas en la variable cambio. Mientras en la zona 1 la variable cambio toma datos con signo positivo (disminuciones de ventas) en la zona 2 los datos muestran signos negativos (aumento o mantenimientos de ventas de la competencia).

GROUP	VALENCIA	CENTRO	TEST STATISTICS	P-VALUE	DF
MEAN	389.4000	5.9000	LEVENE F FOR		
TRIM MEAN	373.9999	0.6250	VARIABILITY	0.12 0.7318	1, 18
STD DEV	133.4626	113.6754			
S.E.M.	42.2046	35.9473	POOLED T	6.92 0.0000	18
SAMPLE SIZE	10	10	SEPARATE T	6.92 0.0000	17.6
MAXIMUM	670.0000	206.0000	TRIM POOL.T	6.89 0.0000	14
MINIMUM	232.0000	-152.0000	TRIM SEP. T	6.89 0.0000	14.0
Z MAX	2.10	1.76			
Z MIN	-1.18	-1.39	MANN-WHIT.	100.0 0.0002	
CASE (MAX)	9	15	(RANK SUMS	155.0	55.0)
CASE (MIN)	1	19			
2ND MAX	524.0000	102.0000			
2ND MIN	274.0000	-126.0000			

Es muy interesante observar las medias de cambio. Mientras que en la zona Valencia las ventas de la competencia han disminuido un promedio de 389'4 por subárea, en la zona Centro se han mantenido casi iguales (un descenso de 5'9). Este efecto todavía es mayor si se consideran las medias recortadas. Sin embargo, como puede observarse en las desviaciones típicas no hay diferencias apenas en la variabilidad del cambio. El test de Levene corrobora esta ausencia de diferencias significativas en variabilidad. Todos los test que comparan ambas distribuciones muestran la existencia de diferencias significativas (con una p para la hipótesis nula muy baja) en el cambio registrado en ambas zonas. Por tanto puede decirse que las ventas de la competencia han disminuido significativamente más en las subáreas del área donde hemos introducido el nuevo producto.

6. Comparando las ventas de una muestra con la media de la población.

**Caso.**

Los resultados anteriores sugieren que se ha producido un descenso generalizado en las ventas de la competencia en las subáreas geográficas bajo estudio. Quizás esas 20 subáreas no puedan considerarse representativas del comportamiento del mercado peninsular en las 940 subáreas en que se divide este. En realidad hay tres áreas más en que opera la empresa: área 2= Cataluña; área 3=Norte; Área 4=Sur que fueron descartadas del estudio por diversas razones operativas.

A través de una Empresa especializada (p.e. datos Nielsen), sabemos que la media de ventas de toda la competencia en las 940 subáreas geográficas en el año 2 es precisamente 1218.

¿Hay diferencias significativas en ventas2 entre las subáreas de la muestra de la zona 2, que hemos tomado como control para el estudio, y la media de la población?

**Prueba t para una sola muestra: Contraste de la media de una muestra con la de la población.**

Input.

```

/INPUT
TITLE IS '3D. ONEGROUP. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
FILE = 'C:\E1\3D002.DAT'. VARIABLES = 5. CASE=20. FORMAT IS FREE.
# La cuestion que se quiere averiguar es si existen diferencias
# significativas en ventas2 entre la zona 2 y la media poblacional.
/VARIABLES
NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/TRANS
IF (ZONA EQ 1) THEN (USE = 0).
/ONEGROUP
VARIABLE = VENTAS2.
MEAN= (VENTAS2)1218.
/PRINT
LINE SIZE=80. CASE=0.
/END
    
```

Si la variable Zona es igual a 1, entonces NO utilices el caso en este análisis

En el párrafo /ONEGROUP se establece que la comparación se efectua en la variable ventas2 con la media poblacional 1218.

Output seleccionado.

```

BMDP Program Output File: C:\E1\3D008.OUT  BMDP3D - T-TESTS
NUMBER OF CASES READ. . . . . 20
CASES WITH USE SET TO ZERO . . . . . 10
REMAINING NUMBER OF CASES . . . . . 10
    
```

Resumen de casos: 10 casos, los de zona 1, eliminados por "use = 0"

DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO. NAME	TOTAL FREQ.	MEAN	STANDARD DEV.	ST. ERR OF MEAN	COEFF OF VAR
1 DISTRIBU	10	15.500	3.0277	.95743	.19533
2 ZONA	10	2.0000	0.0000	0.0000	0.0000
3 NUMCLIEN	10	8.9000	4.9766	1.5737	.55917
4 VENTAS1	10	1329.3	626.59	198.14	.47137
5 VENTAS2	10	1323.4	528.07	166.99	.39903

Los estadísticos descriptivos se refieren ahora solo a los 10 casos de Zona 2.

VARIABLE NO. NAME	S M A L L E S T VALUE	Z-SCR	CASE	L A R G E S T VALUE	Z-SCR	CASE	RANGE
1 DISTRIBU	11.000	-1.49	11	20.000	1.49	20	9.0000
2 ZONA	2.0000		11	2.0000		11	0.0000
3 NUMCLIEN	2.0000	-1.39	20	20.000	2.23	15	18.000
4 VENTAS1	105.00	-1.95	20	2311.0	1.57	15	2206.0
5 VENTAS2	231.00	-2.07	20	2105.0	1.48	15	1874.0

```

VENTAS2  VAR. 5  VS. MEAN= 1218.0000
*****
    
```

		TEST STATISTICS	P-VALUE	DF
	MEAN	1323.4000	0.63	9
	STD DEV	528.0735	0.5436	
	S.E.M.	166.9915		
	SAMPLE SIZE	10		
	MAXIMUM	2105.0000		
	MINIMUM	231.0000		
I AN H=	1 CASES A	Z MAX	1.48	
N (N=	10) X	Z MIN	-2.07	
	CASE (MAX)		15	
	CASE (MIN)		20	

La media de las 10 subáreas (=1323'4) no difiere significativamente de la media de la población (=1218).

7. Comparando las ventas de una muestra con un objetivo.

**Caso.**

La media de ventas de la competencia en las 10 subáreas del área de Valencia en el año 1 (antes de introducir el nuevo producto) fueron de 1814'8. Al planificar la introducción del nuevo producto, teniendo en cuenta ese punto de partida, la Dirección Comercial marco el objetivo de ganar mercado en esta área por introducción del nuevo producto hasta reducir las ventas promedio de la competencia a 1451'8 en el año 2, una reducción del 20% que supondría 362'96 ventas promedio menos para las 10 subáreas de Valencia. ¿Cómo podemos saber si las ventas de la competencia en el año 2 difieren significativamente de nuestro objetivo?

**Prueba t para una sola muestra: Contraste con una media hipotética de la población.**

**Input.**

```

/INPUT
TITLE IS '3D. ONEGROUP. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
FILE='D:\E1\3D002.DAT'. VARIABLES = 5. CASE=20. FORMAT IS FREE.
# La cuestion que se quiere averiguar es si existen diferencias
# significativas en ventas2 entre la zona 1 y el objetivo para esa zona.
/VARIABLES
NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/TRANS
OMIT = 11 TO 20.
/ONEGROUP
VARIABLE = VENTAS2.
MEAN= (VENTAS2)1451.8.
/PRINT
LINESIZE=80. CASE=0.
/END
    
```

La instrucción OMIT permite excluir casos del análisis mencionándolos por su número de caso, aquí del 11 al 20.

Se establece que los casos de la zona 1 (los no excluidos por OMIT) se comparen con la media-objetivo de 1451,8. Esta media representa un descenso del 20% sobre las ventas promedio en esa área el año anterior.

**Output seleccionado:**

BMDP Program Output File: D:\E1\3D009.OUT BMDP3D - T-TESTS  
 DESCRIPTIVE STATISTICS OF DATA

VARIABLE NO.	NAME	TOTAL FREQ.	MEAN	STANDARD DEV.	ST. ERR OF MEAN	COEFF OF VAR
1	DISTRIBU	10	5.5000	3.0277	.95743	.55048
2	ZONA	10	1.0000	0.0000	0.0000	0.0000
3	NUMCLIEN	10	12.000	9.1894	2.9059	.76578
4	VENTAS1	10	1814.8	1153.4	364.73	.63553
5	VENTAS2	10	1425.4	1056.3	334.04	.74107

Estos estadísticos se refieren ahora a las 10 subáreas del área de Valencia.

VARIABLE NO.	NAME	S M A L L E S T VALUE	Z-SCR	CASE	L A R G E S T VALUE	Z-SCR	CASE	RANGE
1	DISTRIBU	1.0000	-1.49	1	10.000	1.49	10	9.0000
2	ZONA	1.0000		1	1.0000		1	0.0000
3	NUMCLIEN	4.0000	-0.87	8	34.000	2.39	9	30.000
4	VENTAS1	799.00	-0.88	8	4782.0	2.57	9	3983.0
5	VENTAS2	406.00	-0.97	8	4112.0	2.54	9	3706.0

NUMBER OF CASES READ. . . . . 20  
 CASES WITH USE SET TO ZERO . . . . . 10  
 REMAINING NUMBER OF CASES . . . . . 10

VENTAS2 VAR. 5 VS. MEAN= 1451.8000

TEST STATISTICS P-VALUE DF

```

-----
MEAN 1425.3999
STD DEV 1056.3223
S.E.M. 334.0384
SAMPLE SIZE 10
MAXIMUM 4112.0000
MINIMUM 406.0000
Z MAX 2.54
Z MIN -0.97
CASE (MAX) 9
CASE (MIN) 8
    
```

1-SAMPLE T -0.08 0.9387 9

La media empírica obtenida en ventas2 en las 10 subáreas es 1425'3999, algo menor que el objetivo de 1451'8 del que no difiere significativamente.

**Muy importante:**  
 Significación estadística y significación práctica son cosas muy distintas. La existencia de diferencias significativas es un concepto estadístico que expresa que es improbable que esas diferencias se deban meramente al azar. Pero no implica que las diferencias sean prácticamente importantes, grandes o relevantes económicamente. Del mismo modo, la ausencia de diferencias estadísticamente significativas no implica que éstas sean pequeñas o poco importantes en términos prácticos. Por tanto, es imprescindible evaluar separadamente el significado práctico de los resultados obtenidos.

8. Comparando las ventas de dos años a través de una variable cambio.

**Prueba t para una sola muestra contrastando si su media difiere significativamente de cero.**

Hemos contrastado anteriormente si existían diferencias significativas entre ventas1 y ventas 2 a través de una prueba t para muestras dependientes (matched t). Ese mismo contraste puede realizarse de otro modo utilizando la variable "cambio" que expresa la diferencia entre ventas1 y ventas2. En efecto, la prueba t para muestras dependientes equivale a una prueba t para una sola muestra sobre la variable diferencia contrastando si esta difiere significativamente de 0. Este contraste no aporta ninguna información adicional, pero ayuda a entender cual es el significado de una prueba t para muestras dependientes.

**Input.**

```

/INPUT
  TITLE IS '3D. T-ONEGROUP. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE ='C:\E1\3D002.DAT'. VARIABLES = 5. CASE=20. FORMAT IS FREE.
  # Comprobamos que una T-Matched equivale a una T-onegroup
  # sobre la variable 'cambio' definida ventas1 menos ventas2.
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/TRANS
  CAMBIO = VENTAS1 - VENTAS2.
/ONEGROUP
  VARIABLE=CAMBIO.
  MEAN = (CAMBIO)0. #Es innecesaria pues la asumirá por defecto.
/PRINT
  LINESIZE=80. CASE=0.
/END
    
```

En un test Onegroup es innecesario especificar que la media poblacional es 0: Este es el valor que se asumiría por defecto.

**Output seleccionado:**

```

BMDP Program Output File: C:\E1\3D012.OUT  BMDP3D - T-TESTS
NUMBER OF CASES READ. . . . . 20

CAMBIO  VAR.   6  VS. MEAN=  0.0000

                                TEST STATISTICS  P-VALUE  DF
                                1-SAMPLE T      3.83 0.0011  19
MEAN          197.6500
STD DEV       230.7848
S.E.M.        51.6050
SAMPLE SIZE   20
MAXIMUM       670.0000
MINIMUM       -152.0000
Z MAX         2.05
Z MIN         -1.52
CASE (MAX)    9
CASE (MIN)    19

H   HH  HH H
HHHHHH HHHHHH H
M-----M
I  AN H=   1 CASES  A
N   (N= 20)      X
    
```

La media de la población sobre la que se contrasta es 0

Hay diferencias significativas (p=0'0011) entre las ventas del primer y segundo año. El segundo año la competencia ha vendido en promedio 197'65 menos por subárea. El test t onegroup sobre cambio (con media de la población = 0) es idéntico al test t matched planteado entre ventas1 y ventas2

9. Comparación gráfica de las ventas en dos años consecutivos.

**Gráfico Quantil-Quantil para muestra dependientes.**

**Input**

```

/INPUT
TITLE IS '3D. T-MATCHED. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
FILE = 'C:\E1\3D002.DAT'. VARIABLES = 5. CASE=20. FORMAT IS FREE.
# Se quiere averiguar es si existen diferencias
# en la distribucion de ventas1 y ventas2 en las 20 subareas
# utilizando QQPLOT (Quantil-Quantil plot)
/VARIABLES
NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/MATCHED
VARIABLES = VENTAS1, VENTAS2.
QQPLOT.
/PRINT
LINESIZE=80. CASE=0.
/PLOT
FILE = 'C:\E1\3D011.PLT'. #Aqui se guardará el QQplot.
/END
    
```

Para un gráfico Quantil-Quantil para muestras dependientes (matched) se plantea la instrucción QQPLOT en el párrafo MATCHED.

Es importante guardar el Plot mediante la instrucción FILE en el párrafo PLOT. Después puede abrirse, editarse e imprimirse desde PLOT en el menú principal BMDP

**Output Seleccionado (Gráfico).**

```

BMDP Program Output File: C:\E1\3D011.OUT  BMDP3D - T-TESTS
NUMBER OF CASES READ. . . . . 20

VENTAS1 VS. VENTAS2 (VAR. NO. 4 VS. 5)
-----
      VENTAS1          VENTAS2          VENTAS1          VENTAS2
      H              X              MEAN          1572.0502    1374.4000
      H              X X              STD DEV       937.0807    814.4783
      HH H           XXX              S.E.M.        209.5376    182.1229
      HH H H        XX XXXX              SAMPLE SIZE    20          20
      H HHHHHH      H XXXXXXXX X       MAXIMUM       4782.0000   4112.0000
M-----M M-----M MINIMUM       105.0000    231.0000
I AN H= 1 CASES A I AN X= 1 CASES A Z MAX          3.43      3.36
N (N= 20) X N (N= 20) X Z MIN         -1.57     -1.40
                        CASE (MAX)          9          9
                        CASE (MIN)         20         20

VENTAS1 - VENTAS2 (VAR. NO. 4 - 5)
-----
      VENTAS1 - VENTAS2          TEST STATISTICS  P-VALUE  DF
      MEAN          197.6500    MATCHED T      3.83 0.0011  19
      STD DEV       230.7848
      S.E.M.        51.6050
      H HH HH H      SAMPLE SIZE    20
      HHHHHH HHHHHH H MAXIMUM       670.0000    CORRELATION  0.9750 0.0000  18
M-----M MINIMUM       -152.0000
I AN H= 1 CASES A Z MAX          2.05
N (N= 20) X N (N= 20) X Z MIN         -1.52
                        CASE (MAX)          9
                        CASE (MIN)         19

      QUANTILE-QUANTILE PLOT
      (BASED ON 20 POINTS)
      Copying Plots from workfile to C:\E1\3D011.PLT ...
    
```

El programa efectua el QQPLOT basado en los 20 casos y lo guarda en el archivo en que se le ha especificado. El output estadístico restante ya lo hemos comentado anteriormente.

Un gráfico Quantil-Quantil traza los cuantiles de los datos de una distribución frente a los de otra. Por ejemplo, un punto representa el par que expresa la ubicación de la mediana de una distribución frente a la de la otra; y así con cada par de cuantiles. Un cuantil es un valor entre 0 y 1 que expresa que proporción de datos deja por debajo de sí un valor determinado en una distribución. Los gráficos Quantil-Quantil expresan globalmente el comportamiento de una distribución frente a otra. Si las dos son iguales todos los puntos caerán sobre una línea de 45° que representa x=y. Si los puntos se ubican en una paralela a esa línea entonces las distribuciones difieren por adición (suma o resta) de una constante. Una diferencia en ángulo expresa una diferencia multiplicativa. Ambas situaciones pueden darse combinadas.



10. Comparación gráfica de la diferencia inter-anual en ventas en dos áreas diferentes.

**Gráfico Quantil-Quantil para muestras independientes.**

Los gráficos QQ pueden plantearse también para comparar dos distribuciones de dos muestras independientes. En nuestro ejemplo vamos a comparar gráficamente las subáreas del área Valencia con las subáreas del área centro en la variable cambio.

Input.

```

/INPUT
  TITLE IS '3D. T-TWOGROUP. INTRODUCCION DE UN NUEVO TIPO DE PAN TOSTADO'.
  FILE ='C:\E1\3D002.DAT'. VARIABLES = 5. CASE=20. FORMAT IS FREE.
  # La cuestion que se quiere averiguar es si existen diferencias
  # en la distribucion de 'cambio' de ventas entre la zona 1 y la zona 2
  # mediante QQPLOTS (Quantil-Quantil plots) o gráficos QQ.
/VARIABLES
  NAMES = DISTRIBU, ZONA, NUMCLIEN, VENTAS1, VENTAS2.
/TRANS
  CAMBIO = VENTAS1 - VENTAS2.
/GROUP
  VARIABLE = ZONA.
  CODES(ZONA)= 1, 2.
  NAMES(ZONA)=VALENCIA, CENTRO.
/TWOGROUP
  VARIABLE=CAMBIO.
  QQPLOT.
/PRINT
  LINESIZE=80. CASE=0.
  NO MEAN. #Evita que el programa muestre medias de las variables (batch).
  NO EZSC. #Evita las puntuaciones tipicas para maximo y minimo.
  NO ECAS. #Evita el numero de caso del maximo y minimo.
/PLOT
  FILE='C:\E1\3D010.PLT'.
/END
    
```

Para obtener el gráfico QQ se introduce la instrucción QQPLOT en el párrafo TWOGROUP.

Para guardar el gráfico en un archivo de PLOTS.

Output Seleccionado (Gráfico).

BMDP Program Output File: C:\E1\3D010.OUT BMDP3D - T-TESTS

NUMBER OF CASES READ. . . . . 20

GROUPING VARIABLE. . . ZONA

	CATEGORY	FREQUENCY
	VALENCIA	10
	CENTRO	10

CAMBIO VARIABLE NUMBER 6

GROUP	1 VALENCIA	2 CENTRO
	HH H	X XX
	HHHHH H	XXXXXX X
M-----M		M-----M
I AN H=	1 CASES A	I AN X= 1 CASES A
N (N= 10)	X N	(N= 10) X

El output estadístico ya ha sido analizado anteriormente.

GROUP	VALENCIA	CENTRO	TEST STATISTICS	P-VALUE	DF
MEAN	389.4000	5.9000	LEVENE F FOR VARIABILITY	0.12 0.7318	1, 18
STD DEV	133.4626	113.6754	POOLED T	6.92 0.0000	18
S.E.M.	42.2046	35.9473	SEPARATE T	6.92 0.0000	17.6
SAMPLE SIZE	10	10			
MAXIMUM	670.0000	206.0000			
MINIMUM	232.0000	-152.0000			
Z MAX	2.10	1.76			
Z MIN	-1.18	-1.39			
CASE (MAX)	9	15			
CASE (MIN)	1	19			

El programa efectúa el QQPLOT basado en 10 puntos (10 para cada una de las zonas), y lo guarda en el archivo en que se le ha indicado en el párrafo /PLOT mediante la instrucción FILE. Para ver, editar e imprimir el gráfico se utiliza la opción PLOT desde el menú principal BMDP.

QUANTILE-QUANTILE PLOT (BASED ON 10 POINTS)

Copying Plots from workfile to C:\E1\3D010.PLT ...