UNED

[TEMA 4]
ANÁLISIS DE DATOS PARA DISEÑOS DE DOS GRUPOS. MUESTRAS RELACIONADAS

ÍNDICE

4.1. – INTRODUCCIÓN.

Ya sabemos con lo que nos vamos a enfrentar en este tema, puesto que vimos en el anterior en qué consisten las muestras relacionadas y somos conscientes de su amplia utilización en la investigación en Psicología.

Las muestras relacionadas tienen una ventaja sobre las independientes que consiste en que nos ayudan a reducir la varianza de error, de manera que cuanto mayor sea la relación entre ambas muestras, menor será la varianza de la distribución muestral de las diferencias, obteniendo por lo tanto un estadístico de contraste mayor.

Comenzaremos por ilustrar cómo se compone la distribución muestral de las diferencias para muestras relacionadas (punto 4.3.1), para tratar, mediante dos ejemplos, el contraste para dos medias relacionadas cuando asumimos que conocemos la varianza poblacional de las diferencias (punto 4.3.2) y cuando asumimos, como es más habitual, que la desconocemos (punto 4.3.3). Pasaremos después al contraste de dos proporciones en muestras relacionadas (punto 4.4), para finalizar con el resumen (punto 4.5) y los ejercicios de autocomprobación (punto 4.6).

Los objetivos, aplicados ahora al caso de dos muestras relacionadas, son los mismos que en el tema anterior.

4.2 OBJETIVOS DEL TEMA.

- ✓ Plantear las hipótesis en función de los objetivos de la investigación
- ✓ Distinguir entre contraste unilateral y bilateral
- ✓ Seleccionar el estadístico de contraste más adecuado a las hipótesis planteadas
- ✓ Conocer la distribución muestral del estadístico seleccionado
- ✓ Realizar los cálculos oportunos para someter a contrastación empírica las hipótesis planteadas
- ✓ Relacionar el intervalo de confianza con el estadístico de contraste
- ✓ Interpretar el nivel p-crítico
- ✓ Determinar e interpretar el o los valores críticos de la distribución muestral
- ✓ Tomar una decisión respecto a las hipótesis planteadas
- ✓ Conocer, comprender e interpretar la magnitud del efecto

4.3 CONTRASTE DE HIPÓTESIS SOBRE DOS MEDIAS EN MUESTRAS RELACIONADAS.

Aunque contamos con dos medias como en el tema anterior, la distribución muestral es diferente, por lo que comenzaremos este apartado con un breve ejemplo, cuyo fin es exclusivamente didáctico y que nos servirá para comprender cómo se compone la distribución muestral de las diferencias.

4.3.1 Distribución muestral para dos medias relacionadas.

Supongamos que tenemos una población compuesta por 4 sujetos a los que medimos la variable dependiente antes y después de una terapia, siendo sus puntuaciones y sus diferencias las que podemos ver en la Tabla 4.1

Tabla 4.1
Ejemplo. Datos de cuatro sujetos antes y después de la terapia.

	Sujeto 1	Sujeto 2	Sujeto 3	Sujeto 4
Antes	8	6	4	10
Después	3	5	6	8
Diferencia	8 – 3 = 5	6 - 5= 1	4 – 6 = -2	10 – 8 = 2

La población sobre la que vamos a trabajar está formada por las diferencias entre las puntuaciones “antes” y “después”, o sea, por los valores, una vez ordenados: {-2; 1; 2; 5}. La media y varianza de la población de diferencias es:

$$\mu_d = \frac{6}{4} = 1{,}5, \qquad \sigma_d^2 = \frac{34}{4} - 1{,}5^2 = 6{,}25$$

Sobre la población de diferencias, tomamos todas las muestras de un determinado tamaño y las medias de todas ellas formarán la distribución muestral de las diferencias. Por ejemplo, para muestras de tamaño n=2, los resultados de todas las muestras con reposición pueden verse en la Tabla 4.2:

Tabla 4.2
Todas las posibles medias aritméticas para muestras de tamaño n=2

	-2	**1**	**2**	**5**
-2	-2	-0,5	0	1,5
1	-0,5	1	1,5	3
2	0	1,5	2	3,5
5	1,5	3	3,5	5

Por lo que la distribución muestral de las diferencias está formada por los valores:

{-2; -0,5; -0,5; 0; 0; 1; 1,5; 1,5; 1,5; 1,5; 2; 3; 3; 3,5; 3,5; 5}

La media y varianza de la distribución muestral podemos calcularla con los datos de la Tabla 4.1, o bien a través de los parámetros poblacionales como hemos visto en casos anteriores:

$$\mu_d = \frac{6}{4} = 1{,}5 \qquad \sigma^2_{\bar{Y}_1-\bar{Y}_2} = \frac{\sigma_d^2}{n} = \frac{6{,}25}{2} = 3{,}125$$

Podemos apreciar cómo el contraste de hipótesis sobre dos medias relacionadas es muy parecido al de una sola media. Aunque ahora partimos de dos muestras, al estar estas relacionadas, finalmente tenemos una sola variable (las diferencias entre cada par de puntuaciones). Para apreciar más claramente esta similitud, vamos a comparar el estadístico de contraste en ambos casos cuando la varianza poblacional es conocida (véase la Tabla 4.3.).

Tabla 4.3

Estadísticos de contraste para una media y dos relacionadas

Una media	Dos medias relacionadas
$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$ Donde: $\bar{X}$ = media de la muestra. σ^2= varianza de la población. μ_0 = media que postula la hipótesis nula	$$Z = \frac{\bar{D} - \mu_d}{\sqrt{\frac{\sigma_d^2}{n}}}$$ Donde: $\bar{D}$ = media de las diferencias de ambas muestras. σ_d^2 = varianza de la población de diferencias. μ_d = media de diferencias que postula la hipótesis nula. Habitualmente igual a cero

4.3.2 Conocida la varianza poblacional de las diferencias.

Ejemplo 4.1. Un psicólogo escolar está interesado en estudiar si la presión de los padres para el rendimiento escolar es igual en chicos y en chicas. Toma una muestra aleatoria de 36 parejas de hermanos (chico y chica), y mediante un test que proporciona medidas en una escala de intervalo mide la variable "presión para el rendimiento escolar" en todos los individuos. La media para los chicos (Grupo 1) fue igual a 21 y para el grupo de chicas fue igual a 19. Suponemos que conocemos la varianza de las diferencias y que es igual a 64. A un nivel de confianza del 99%. ¿Es igual la presión para el rendimiento escolar en chicos y chicas?

Condiciones y supuestos. Las muestras de chicos y chicas son relacionadas dado que están compuestas por parejas de hermanos. Conocemos la varianza de las diferencias en la población, la variable dependiente está medida a un nivel de intervalo, y no sabemos si la población de diferencias se distribuye normalmente, pero la muestra supera las 30 observaciones. En general, los supuestos son:

- Variable dependiente con un nivel de medida de intervalo o razón.
- Población de diferencias normalmente distribuida o n ≥ 30.
- Varianza poblacional de las diferencias conocida.

Formular las hipótesis. El psicólogo no tiene una hipótesis previa sobre las diferencias debidas al género en la variable "presión para el rendimiento escolar", por lo que planteamos un contraste bilateral.

$$\begin{array}{l} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{array} \quad \text{o bien,} \quad \begin{array}{l} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{array}$$

Elección del estadístico de contrate y su distribución muestral. El estadístico de contraste se distribuye normalmente, según la expresión de la Ecuación 4.1:

$$Z = \frac{\overline{D} - \mu_d}{\sqrt{\frac{\sigma_d^2}{n}}} = \frac{(21-19)-0}{\sqrt{\frac{64}{36}}} = 1{,}5 \qquad \text{(Ecuación 4.1)}$$

Mediante la tabla de curva normal del apéndice deducimos que la probabilidad de encontrar valores superiores a una puntuación típica de 1,5 es igual a: 0,0668. Al ser el contraste bilateral, el nivel p-crítico es: $2 \times 0{,}0668 = 0{,}1336$.

Establecer la regla de decisión en función del nivel de confianza. Al nivel de confianza del 99%, los valores críticos que delimitan la zona en la que mantenemos H_0 son: z = ±2,58. La Figura 4.1 muestra los datos del problema.

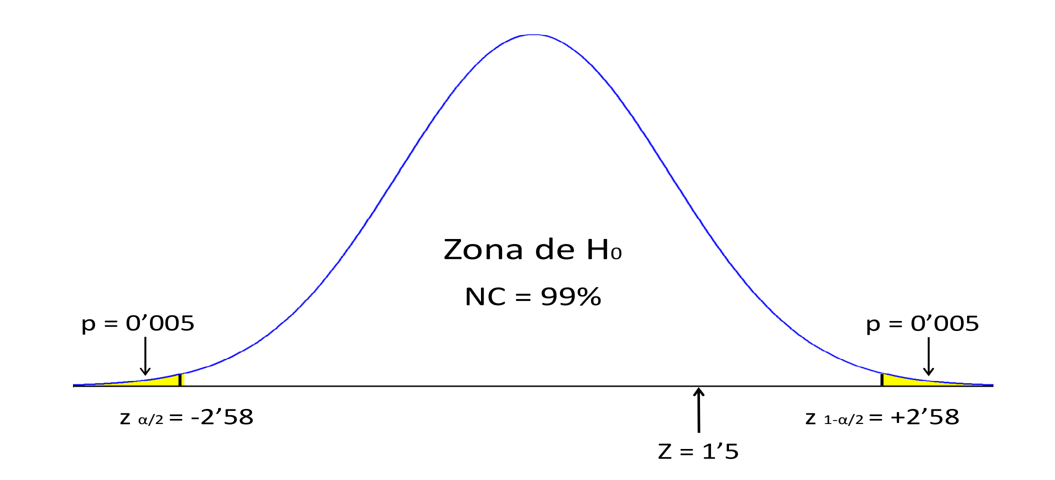

Figura 4.1.*Distribución muestral postulada por la hipótesis nula*

Conclusión. Al nivel de confianza del 99% no existen diferencias significativas entre las medias de chicos y chicas, puesto que el estadístico de contraste se encuentra comprendido en el intervalo que definen los valores críticos, por lo que mantenemos la hipótesis nula.

Interpretar el resultado en función del contexto de la investigación. Según los datos que manejamos, no existen diferencias en cuanto a la presión para el rendimiento escolar entre chicos y chicas.

Intervalo de confianza. Viene dado por la expresión de la Ecuación 4.2:

$$\overline{D} \pm z\alpha/_2 \sqrt{\frac{\sigma_d^2}{n}} \qquad \text{(Ecuación 4.2)}$$

Que con nuestros datos vale:

$$2 \pm 2{,}58\sqrt{\frac{64}{36}} \longrightarrow (-1{,}44;\ 5{,}44)$$

Observamos que el intervalo de confianza contiene al cero que es lo que postula la hipótesis nula.

4.3.3 Desconocida la varianza poblacional de las diferencias.

Al igual que comentábamos en temas anteriores, lo más habitual en un caso práctico es que desconozcamos la varianza de la población, siendo el proceso muy parecido al que acabamos de ver. Simplemente tenemos que sustituir la varianza poblacional de las diferencias (σ_d^2) por su estimador, que es la cuasivarianza de las diferencias en la muestra, y que se distribuye según t de Student con n - 1 grados de libertad.

Ejemplo 4.2. Un psicólogo que trabaja en una empresa imparte un curso sobre asertividad. El objetivo del curso consiste en fomentar esta habilidad en los directivos que forman parte de su departamento. Antes del curso mide la asertividad mediante un test que proporciona medidas en una escala de intervalo, y en el que las puntuaciones altas indican un comportamiento asertivo. Al finalizar el curso el psicólogo aplica de nuevo el test de asertividad a los asistentes. Las puntuaciones antes y después del curso fueron las siguientes:

Sujeto	1	2	3	4	5	6	7	8	9	10
Antes	18	24	25	24	27	30	24	31	24	28
Después	24	23	34	22	34	40	35	31	27	30

Con un nivel de confianza del 95%. Suponiendo que en la población la distribución de las diferencias es normal, ¿Podemos decir que el curso realizado por el psicólogo ha incrementado la asertividad de los directivos?

Antes de realizar el contraste de hipótesis vamos a calcular las diferencias y los cuadrados de las mismas entre las condiciones "Antes" y "Después" para cada par de sujetos. Estos cálculos, necesarios para calcular la media y la cuasivarianza insesgada, aparecen en la Tabla 4.4.

Tabla 4.4

Sujeto	1	2	3	4	5	6	7	8	9	10
Antes	18	24	25	24	27	30	24	31	24	28
Después	24	23	34	22	34	40	35	31	27	30
d_i	-6	1	-9	2	-7	-10	-11	0	-3	-2
d_i^2	36	1	81	4	49	100	121	0	9	4

La media y la varianza insesgada de las puntuaciones diferencia valen:

$$\overline{D} = \frac{\sum d_i}{n} = -4{,}5, \qquad S_d^2 = \frac{\sum d_i^2}{n} - (\overline{D})^2 = \frac{405}{10} - (-4{,}5)^2 = 20{,}25$$

$$\hat{S}_d^2 = S_d^2 \times \frac{n}{n-1} = 20{,}25 \times \frac{10}{9} = 22{,}5$$

Condiciones y supuestos. Según el enunciado del problema, la variable dependiente está medida a un nivel de intervalo. Hemos de suponer que la población de las diferencias sigue una distribución normal porque la muestra es pequeña y no conocemos su varianza. En general se tendrá que cumplir:

- Variable dependiente con un nivel de medida de intervalo o razón.
- Población de diferencias que se distribuye normalmente, o bien $n \geq 30$.
- Varianza poblacional de las diferencias desconocida.

Asumiendo que se cumplen estos supuestos, pasamos a plantear las hipótesis.

Hipótesis. El psicólogo tiene la idea de que su curso incrementará las puntuaciones en asertividad, o sea, que la media en esta variable será menor antes del curso (grupo 1) que después del curso (grupo 2), por lo que podemos plantear un contraste unilateral izquierdo de la siguiente forma.

$$\begin{matrix} H_0: \mu_1 - \mu_2 \geq 0 \\ H_1: \mu_1 - \mu_2 < 0 \end{matrix} \quad \text{o bien,} \quad \begin{matrix} H_0: \mu_1 \geq \mu_2 \\ H_1: \mu_1 < \mu_2 \end{matrix}$$

También podemos referirnos en las hipótesis directamente a la población de diferencias:

$$\begin{matrix} H_0: \mu_d \geq 0 \\ H_1: \mu_d < 0 \end{matrix}$$

Elección del estadístico de contraste y su distribución muestral. El estadístico de contraste se distribuye según la t de Student con n-1 grados de libertad, y lo calculamos según la expresión:

$$T = \frac{\overline{D} - \mu_d}{\sqrt{\frac{\hat{S}_d^2}{n}}} = \frac{-4{,}5 - 0}{\sqrt{\frac{22{,}5}{10}}} = -3 \qquad (4.3)$$

Para averiguar el nivel crítico de forma aproximada, buscamos en la tabla t de Student del apéndice entre que valores se encuentra el estadístico de contraste, en este caso para 9 grados de libertad observamos: $-3{,}25 < -3 < -2{,}821$, por lo que el nivel p-crítico se encontrará entre los valores: $0{,}005 < p < 0{,}01$.

Establecer la regla de decisión en función del nivel de confianza. Los grados de libertad del ejemplo son: n – 1 = 9 Acudiendo a la tabla t de Student, el valor crítico que delimita cuando mantenemos o rechazamos la hipótesis nula, a un nivel de confianza del 95%, es -1,833. En la Figura 4.2 representamos los datos del ejemplo.

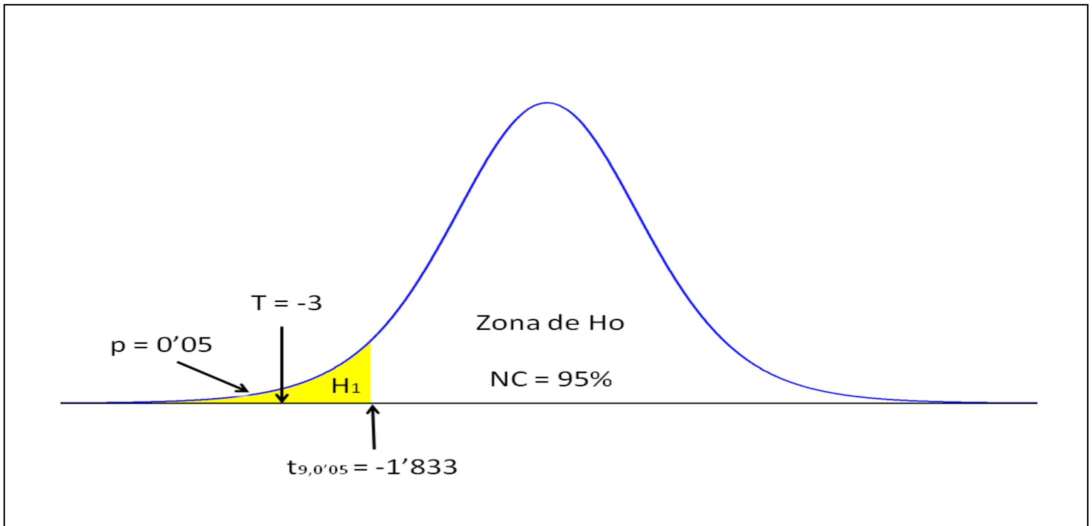

Figura 4.2. *Distribución muestral de diferencias según especifica la hipótesis nula.*

Conclusión. A un nivel de confianza del 95%, rechazamos la hipótesis nula puesto que el estadístico de contraste es más extremo que el valor crítico (máxima diferencia que cabe esperar por simple azar), por lo que concluimos que la media en asertividad de los directivos es inferior antes que después del curso. En cualquier caso, los datos obtenidos superan con creces el nivel de confianza que de antemano había fijado el psicólogo tal y como queda de manifiesto en el valor del nivel crítico.

Interpretar el resultado en función del contexto de la investigación. El curso realizado por el psicólogo ha obtenido los resultados esperados, demostrando su utilidad para fomentar la asertividad en los directivos de su departamento.

Intervalo de confianza. Podemos calcularlo mediante la Ecuación 4.4:

$$\overline{D} \pm t_{\frac{\alpha}{2},\ n-1}\sqrt{\frac{\hat{S}_d^2}{n}} \qquad \text{(Ecuación 4.4)}$$

Que aplicándola a nuestros datos:

$$-4{,}5 \pm 2{,}262\sqrt{\frac{22{,}5}{10}} \longrightarrow (-1{,}107;\ -7{,}893)$$

Con lo que interpretaríamos, a un nivel de confianza del 95% que la media en asertividad es menor antes del curso (ambos límites son negativos y no contienen el valor cero plateado en la hipótesis nula), observándose un aumento después del mismo, que oscila entre 1,107 y 7,893 puntos en la puntuación media del test.

4.5 CONTRASTE DE HIPÓTESIS SOBRE DOS PROPORCIONES EN MUESTRAS RELACIONADAS.

Emplearemos este contraste de hipótesis cuando tengamos una muestra con una variable dependiente dicotómica o dicotomizada que medimos en dos momentos temporales distintos, que en general llamaremos “Antes” y “Después”. Veamos un ejemplo.

Ejemplo 4.3. Un empresario, antes de introducir en el mercado un determinado producto “X”, toma una muestra aleatoria de 500 sujetos de la población a la que quiere dirigirse y les pregunta si comprarían o no dicho producto. A continuación les muestra las posibles ventajas que aporta el producto “X” y les vuelve a preguntar si lo comprarían. Mostramos los resultados en la Tabla 3.5. ¿Podemos afirmar al nivel de confianza del 99% que, la demostración del producto “X” ha sido eficaz?

En la Tabla 4.5 aparecen los resultados del experimento.

Tabla 4.5
Número de personas que compraría “X” antes y después de la demostración.

	¿Compraría “X” antes de la demostración?			
		SI	NO	
¿Compraría “X” después?	SI	a = 60	**b = 60**	120
	NO	**c = 40**	d = 340	380
		100	400	500

En este contraste nos interesan los sujetos que han cambiado de opinión tras la presentación de “X” (casillas “b” y “c” de la Tabla 4.6), en general, aquellos sujetos cuya puntuación es diferente en los dos momentos en los que medimos la variable dependiente. Una vez seleccionados dichos sujetos, nos preguntamos si el número de cambios es el mismo en las dos direcciones (hipótesis nula) o si, por el contrario, la mayor parte de los sujetos que cambian de opinión lo hacen en una dirección determinada (hipótesis alternativa). Supongamos que en nuestro ejemplo han cambiado de opinión 100 sujetos pero 50 de ellos que inicialmente no comprarían “X” si lo harían después de la demostración mientras que también 50 sujetos cambian de opinión pero de forma opuesta. En este caso el número de personas que compraría “X” es el mismo antes y después, por lo que no habría sido eficaz la demostración llevada a cabo por el empresario.

En la Tabla 4.5 apreciamos que 60 sujetos han cambiado de opinión en la dirección que pretendía el empresario puesto que declararon que no comprarían “X” antes de la demostración y sí después. Por otro lado, tenemos 40 sujetos que han cambiado de opinión en la dirección contraria. Se trata de comprobar si el primer grupo de sujetos supera estadísticamente al segundo. En total son 100 $(b + c)$ los sujetos que cambian de opinión.

Condiciones y supuestos. Contamos con 100 observaciones independientes y una variable dicotómica, donde definimos el éxito como: “No antes, Sí después” y el fracaso como “Sí antes, No después”. En general tendremos:

- Variable dependiente dicotómica o dicotomizada
- Muestra con “*b+c”* observaciones independientes, donde *b+c > 25*.

Hipótesis. La hipótesis nula especifica que la proporción de éxitos es igual o menor que la de fracasos, y la alternativa que la proporción de éxitos es superior a la de fracasos.

$$H_0: \pi_b \leq \pi_c$$
$$H_1: \pi_b > \pi_c$$

Estadístico de contraste y su distribución muestral. Podemos utilizar dos estadísticos de contraste diferentes. El primero de ellos se distribuye normalmente y viene expresado por la Ecuación 4.5:

$$Z = \frac{b-c}{\sqrt{b+c}} = \frac{60-40}{\sqrt{60+40}} = 2 \qquad \text{(Ecuación 4.5)}$$

Contraste que podríamos haber realizado con los conocimientos adquiridos en Tema 2. Si consideramos que tenemos una única muestra de 100 observaciones, podemos plantear las siguientes hipótesis:

$$H_0: \pi_{\text{Éxitos}} \leq 0{,}5$$
$$H_1: \pi_{\text{Éxitos}} > 0{,}5$$

Siendo el estadístico de contraste igual a:

$$Z = \frac{p - \pi_{\text{Éxitos}}}{\sqrt{\frac{\pi_{\text{Éxitos}}(1-\pi_{\text{Éxitos}})}{n}}} = \frac{\frac{60}{100} - 0{,}5}{\sqrt{\frac{0{,}5 \times 0{,}5}{100}}} = 2 \xrightarrow{\textit{nivel crítico}} p = 0{,}0228$$

También podemos aplicar el test de McNemar, donde el estadístico de contraste se distribuye según Chi cuadrado con un grado de libertad, siendo igual a:

$$X^2 = \frac{(b-c)^2}{b+c} \qquad (4.6)$$

Con los datos del Ejemplo 4.3:

$$X^2 = \frac{(b-c)^2}{b+c} = \frac{(60-40)^2}{60+40} = 4 \xrightarrow{\textit{nivel crítico}} p = 0{,}0455$$

Observamos que este estadístico es el cuadrado del anterior, por lo que se distribuirá según Chi cuadrado con un grado de libertad, como podríamos deducir de los conocimientos adquiridos en la asignatura Introducción al Análisis de Datos. También apreciamos que el nivel p-crítico es mayor para Chi cuadrado que para Z, lo que sucederá siempre que el contraste sea unilateral, siendo en este caso más difícil rechazar la hipótesis alternativa con el test de McNemar. Si el contraste fuera bilateral el nivel crítico sería el mismo para ambos estadísticos.

En el ejemplo que hemos desarrollado, el contraste es unilateral, lo que no plantearía ningún problema si utilizamos el estadístico Z, pero si empleamos el estadístico χ^2 sólo podremos plantear una **hipótesis estadística** bilateral, puesto que al estar elevado al cuadrado, este estadístico no nos informa de la dirección de las diferencias. No obstante, **el investigador puede interpretar los datos**, y por lo tanto la dirección de las diferencias, fijándose en cuál de las dos casillas "b" o "c" presenta una frecuencia mayor.

La diferencia entre los estadísticos Z y X^2, consiste en que el estadístico X^2 es insensible a la dirección del cambio producido. En nuestro ejemplo, el valor de ambos estadísticos es:

$$Z = \frac{b-c}{\sqrt{b+c}} = \frac{60-40}{\sqrt{60+40}} = +2$$

$$X^2 = \frac{(b-c)^2}{b+c} = \frac{(60-40)^2}{60+40} = +4$$

Observamos que ambos estadísticos son positivos, pero si los resultados fueran los contrarios a los obtenidos, es decir, si las casillas "b" y "c" estuvieran al revés respecto del ejemplo anterior, tendríamos **c > b**. Los datos serían:

Número de personas que compraría "X" antes y después de la demostración.					
	¿Compraría "X" antes de la demostración?				
		SI	NO		
¿Compraría "X" después?	SI	a = 60	**b = 40**	100
	NO	**c = 60**	d = 340	400
		120	380	500

Al calcular los estadísticos obtendríamos:

$$Z = \frac{b-c}{\sqrt{b+c}} = \frac{40-60}{\sqrt{60+40}} = -2$$

$$X^2 = \frac{(b-c)^2}{b+c} = \frac{(40-60)^2}{40+60} = +4$$

El estadístico Z ahora nos da negativo porque la frecuencia de "b" es menor que la de "c". Ha detectado que hemos cambiado los datos, pero el estadístico X^2 nos da el mismo valor a pesar de que las frecuencias están al revés. Es decir X^2 puede detectar que existe un cambio pero no en qué dirección, por eso no podemos plantear una hipótesis estadística unilateral con dicho estadístico.

Por otro lado, con χ^2 no tiene sentido rechazar la hipótesis nula si obtenemos un valor muy pequeño, lo que nos indicaría que el número de sujetos que cambian de opinión en ambas direcciones son muy parecidos. De hecho, como podemos apreciar en la Ecuación 4.6, χ^2 valdrá cero cuando "$b = c$" que es precisamente lo que postula la hipótesis nula. En la Figura 4.3 mostramos cuándo mantener o rechazar H_0 al emplear la prueba de McNemar. En temas posteriores veremos que sucede algo análogo en otras pruebas estadísticas, por ejemplo cuando aplicamos un Análisis de la Varianza.

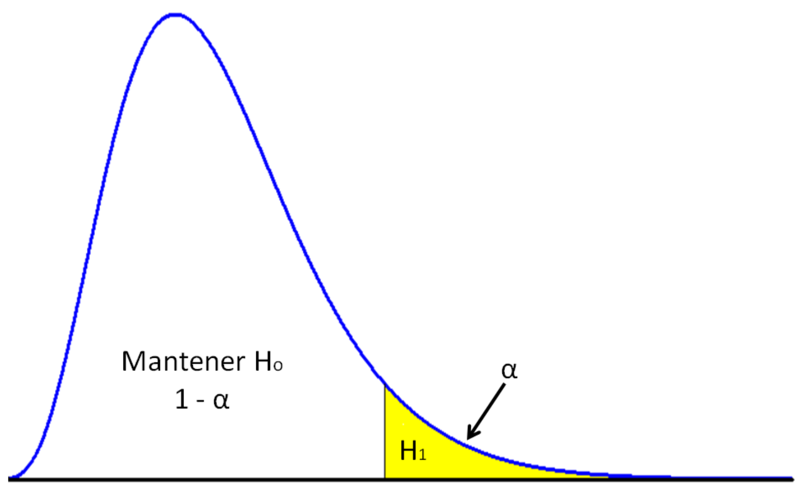

Figura 4.3. *Aplicación de la prueba de McNemar, en la que siempre queda α en la parte derecha.*

Establecer regla de decisión en función del nivel de confianza. Buscando en las tablas correspondientes los valores críticos al nivel de confianza del 99%, para los estadísticos Z y χ^2 con un grado de libertad tenemos que:

Estadístico Z: $2 < 2{,}33 \longrightarrow \mathit{Mantenemos}\ H_0$

Estadístico χ^2: $4 < 6{,}6349 \longrightarrow \mathit{Mantenemos}\ H_0$

Conclusión. Al nivel de confianza del 99% fijado por el empresario, no podemos rechazar la hipótesis nula, no obstante, como podemos apreciar al calcular el nivel crítico, los resultados sí son significativos al nivel de confianza del 95%.

Interpretar el resultado en el contexto de la investigación. Si bien podemos afirmar que los resultados son significativos (con un nivel de confianza del 95%), no superan el nivel de confianza fijado de antemano. De un total de 400 personas que no comprarían el producto "X" inicialmente, tan sólo 60 han cambiado de opinión, y por otro lado, 40 de ellas lo han hecho en la dirección contraria. El empresario del ejemplo debería considerar si es rentable la campaña publicitaria que ha diseñado, puesto que en total, tan sólo se incrementa en 20 personas el número de posibles clientes que ganaría tras la demostración del producto “X”.

4.5 RESUMEN DEL TEMA.

El esquema seguido vuelve a ser el mismo que en los dos anteriores, por lo que el resumen del Tema 2 en el que se fijan los pasos de todo contraste de hipótesis sigue siendo válido. Además, si introducíamos pocos conceptos nuevos en el tema anterior, en este caso son todavía menos. Hemos comenzado por el contraste sobre dos medias relacionadas donde cambia la manera de obtener la distribución muestral de las diferencias, que por otro lado es muy similar a cuando trabajábamos con una única muestra. Finalmente estudiamos dos estadísticos para el contraste sobre dos proporciones en muestras relacionadas.

En las Tablas 4.6 y 4.7 que presentamos a continuación ofrecemos un resumen de las técnicas desarrolladas a lo largo de este tema.

Tabla 4.6. *Contrastes de hipótesis para dos medias en muestras relacionadas.*

SUPUESTOS	- Observaciones independientes. - Nivel de medida de intervalo o razón. - Distribuciones normales en la población de diferencias o bien: $n \geq 30$.	
Varianza poblacional	**Conocida**	**Desconocida**
HIPÓTESIS	*Igual en los dos casos. En función de la hipótesis científica plantearemos un contrate:* Bilateral: $H_0: \mu_1 - \mu_2 = 0$, $H_1: \mu_1 - \mu_2 \neq 0$ Unilateral izquierdo: $H_0: \mu_1 - \mu_2 \geq 0$, $H_1: \mu_1 - \mu_2 < 0$ Unilateral derecho: $H_0: \mu_1 - \mu_2 \leq 0$, $H_1: \mu_1 - \mu_2 > 0$	
ESTADÍSTICO DE CONTRASTE	$Z = \dfrac{\overline{D} - \mu_d}{\sqrt{\dfrac{\sigma_d^2}{n}}}$	$T = \dfrac{\overline{D} - \mu_d}{\sqrt{\dfrac{\hat{S}_d^2}{n}}}$
DISTRIBUCIÓN MUESTRAL	Normal tipificada: $N(0,1)$	"t" de Student: $g.l. = n - 1$
REGLA DE DECISIÓN	(ver tablas siguientes)	

REGLA DE DECISIÓN

Distribución normal (varianza conocida).

	Bilateral	Unilateral Derecho	Unilateral Izquierdo
Mantener H_0	$z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}$	$Z < z_{1-\alpha}$	$Z > z_{\alpha}$
Rechazar H_0	$Z < z_{\frac{\alpha}{2}}$ ó $Z > z_{1-\frac{\alpha}{2}}$	$Z > z_{1-\propto}$	$Z < z_{\alpha}$

Distribución "t" de Student (varianza desconocida).

	Bilateral	Unilateral Derecho	Unilateral Izquierdo
Mantener H_0	$t_{g.l.,\frac{\alpha}{2}} < T < t_{g.l.,1-\frac{\alpha}{2}}$	$T < t_{g.l.,1-\alpha}$	$T > t_{g.l.,\alpha}$
Rechazar H_0	$T < t_{g.l.,\frac{\alpha}{2}}$ ó $T > t_{g.l.,1-\frac{\alpha}{2}}$	$T > t_{g.l.,\propto}$	$T < t_{g.l.,\alpha}$

<table>
<tr><td colspan="3">Tabla 4.7. Contrastes de hipótesis sobre dos proporciones en muestras relacionadas.</td></tr>
<tr><td>SUPUESTOS</td><td colspan="2">- Variable dependiente dicotómica o dicotomizada que medimos en la misma muestra en dos ocasiones.
- Muestra con “$b + c$” observaciones independientes, donde $b + c$ > 25 (b y c representan a los sujetos cuya puntuación es distinta en las dos ocasiones.</td></tr>
<tr><td>HIPÓTESIS</td><td colspan="2">Bilateral
$H_0: \pi_b = \pi_c$
$H_1: \pi_b \neq \pi_c$

Unilateral izquierdo
(sólo para estadístico Z)
$H_0: \pi_b \geq \pi_c$
$H_1: \pi_b < \pi_c$

Unilateral derecho
(sólo para estadístico Z)
$H_0: \pi_b \leq \pi_c$
$H_1: \pi_b > \pi_c$</td></tr>
<tr><td>ESTADÍSTICO DE CONTRASTE</td><td>$Z = \frac{b - c}{\sqrt{b + c}}$</td><td>$\chi^2 = \frac{(b - c)^2}{b + c}$</td></tr>
<tr><td>DISTRIBUCIÓN MUESTRAL</td><td>$N(0,1)$</td><td>χ^2 con un grado de libertad</td></tr>
<tr><td>REGLA DE DECISIÓN</td><td>Igual que en casos anteriores</td><td>$X^2 < \chi^2_{1,1-\alpha} \longrightarrow$ Mantener H_0
$X^2 > \chi^2_{1,1-\alpha} \longrightarrow$ Rechazar H_0</td></tr>
</table>

4.7 EJERCICIOS DE AUTOCOMPROBACIÓN.

1. En cuál de los siguientes diseños empleamos muestras relacionadas: a) cuando los mismos sujetos son sometidos a dos condiciones experimentales; b) cuando disponemos de parejas de hermanos gemelos que son asignados aleatoriamente a grupos distintos; c) las dos opciones anteriores son verdaderas.
2. La varianza de la distribución muestral es mayor: a) en un diseño de dos muestras independientes; b) en un diseño de dos muestras relacionadas; c) son iguales en ambos casos.
3. Si trabajamos con dos grupos siendo los tamaños de las muestras: $n_1 = 30$ y $n_2 = 35$, sabemos que: a) se trata de un contraste paramétrico; b) las muestras están relacionadas; c) las muestras son independientes.
4. En un contraste sobre dos proporciones en muestras relacionadas realizamos el contraste: a) con los sujetos que obtienen puntuaciones diferentes en dos ocasiones; b) con los sujetos que obtienen la misma puntuación en dos ocasiones; c) utilizamos todas las puntuaciones de los sujetos en las dos ocasiones.
5. En un contraste sobre dos proporciones en muestras relacionadas sólo podemos plantear tanto un contraste unilateral como bilateral: a) con el estadístico Z; b) mediante la prueba de McNemar; c) ninguna de las anteriores es correcta.

Un psicólogo escolar considera que una determinada técnica de relajación puede disminuir la ansiedad de los estudiantes para realizar los exámenes. Para comprobar esta hipótesis toma una muestra aleatoria de 31 estudiantes a los que mide su ansiedad mediante un test antes de un examen de matemáticas (Condición 1), posteriormente les enseña técnicas de relajación y vuelve a aplicar el test antes del siguiente examen de matemáticas (Condición 2) a los mismos alumnos. La media en ansiedad antes y después de aprender técnicas de relajación fueron, respectivamente: 12,18 y 10. Sabiendo que a mayor puntuación mayor ansiedad, que la cuasivarianza de las diferencias es igual a 25 y con un nivel de confianza del 95% conteste a las siguientes cuestiones:

6. La hipótesis alternativa es: a) $\mu_1 - \mu_2 \neq 0$; b) $\mu_1 - \mu_2 < 0$; c) $\mu_1 - \mu_2 > 0$.
7. El valor del estadístico de contrate es: a) 1,697; b) 2,428; c) -1,697.
8. El valor crítico es: a) 1,697; b) 2,428; c) -1,697.
9. Suponiendo que la hipótesis nula es cierta, la probabilidad de encontrar unos resultados como los obtenidos en la muestra es, aproximadamente: a) 0,25; b) 0,01; c) 0,005.
10. La conclusión del estudio es: a) mantener la hipótesis nula porque el estadístico de contraste es menor que el nivel crítico; b) rechazar la hipótesis nula porque el estadístico de contraste es mayor que el valor crítico; c) mantener la hipótesis nula porque $p < \alpha$.

4.9 SOLUCIONES A LOS EJERCICIOS DE AUTOCOMPROBACIÓN.

1. "c".

2. "a".

3. Si $\mathrm{n}_1 \neq \mathrm{n}_2$ la respuesta correcta siempre es "C". Con muestras relacionadas cada sujeto de una muestra tiene su pareja (o él mismo) en la otra muestra, por lo que trabajando con muestras relacionadas siempre se cumplirá que: $\mathrm{n}_1 = \mathrm{n}_2$

4. "a".

5. "a"

6. El psicólogo espera que la media en ansiedad sea menor tras enseñar a técnicas de relajación a los alumnos, por lo que la respuesta correcta es "c".

7. La respuesta correcta es "b".

$$T = \frac{\overline{D} - \mu_d}{\sqrt{\frac{\hat{S}_d^2}{n}}} = \frac{(12{,}18 - 10) - 0}{\sqrt{\frac{25}{31}}} = 2{,}428$$

8. Buscando en la tabla t de Student con 30 grados de libertad y un nivel de confianza del 95% comprobamos que la respuesta correcta es "a".

9. Consultado las tablas observamos que entre los valores que contiene para 30 g.l. el valor que más se aproxima al estadístico de contraste es 2,457, siendo la probabilidad de encontrar valores más extremos que dicho valor igual a 0,01, luego la respuesta es"b".

10. "b".