11

Comparación de proporciones

Neus Canal Díaz

11.1. Introducción

En la investigación biomédica se encuentran con frecuencia datos o variables de tipo cualitativo (nominal u ordinal), mediante las cuales se clasifican grupos de sujetos o individuos en dos o más categorías excluyentes entre ellas. Imaginemos que queremos evaluar la presencia de desnutrición según el tratamiento sustitutivo empleado, bien Hemodiálisis (HD) o Diálisis Peritoneal (DP). En este caso, querremos evaluar la presencia o ausencia de un determinado evento (desnutrición) en función del tratamiento o terapia administrada. Cuando pretendemos comparar grupos de sujetos con respecto a una variable categórica, los resultados se presentan mediante tablas de frecuencias de doble entrada, conocidas con el nombre de **Tablas de contingencia**. El método estadístico a utilizar dependerá del número de proporciones a comparar, es decir, del número de categorías de la variable que se quiere comparar.

El caso más sencillo y habitual de comparar dos variables cualitativas, es aquel en que ambas variables tienen dos posibles categorías de respuesta (estas variables reciben el nombre de variables dicotómicas), reduciéndose la tabla de contingencia a una tabla 2x2. Generalmente una de ellas corresponde al tratamiento y la otra al resultado (curación, éxito, muerte, entre otras).

		Característica A (tratamiento o terapia)		
		Tratamiento A	Tratamiento B	Total
Característica B (desenlace)	Presencia	a	b	a+d
	Ausencia	c	d	c+d
	Total	a+c	b+d	n

Tabla 7. Tabla general de contingencia para dos variables dicotómicas (tabla 2x2)

Las pruebas estadísticas aplicables en la comparación de proporciones, ya sean dos o más, también difieren según se trate de comparar medidas realizadas en grupos independientes o de datos apareados (medidas realizadas en un mismo grupo de individuos en dos momentos distintos del tiempo). En el caso de comparar dos proporciones independientes, las pruebas más utilizadas son la prueba Z de comparación de proporciones y la prueba de Ji-cuadrado. En el caso de tratarse de datos apareados puede utilizarse la prueba de McNemar. En la comparación de más de dos proporciones independientes es posible aplicar la prueba de la Ji-cuadrado (veremos en qué casos es aplicable), mientras que en el caso de datos apareados se utiliza la prueba Q de Cochran.

11.2. Comparación de proporciones para datos independientes

Cuando queremos comparar una respuesta que se mide como una proporción entre dos o más niveles necesitamos pruebas que nos indiquen si hay diferencias entre estas proporciones, es decir, si se distribuyen homogéneamente entre los niveles de la variable o por el contrario, existen diferencias. Por lo tanto, la hipótesis experimental es que las proporciones de ocurrencia de determinado evento medido en muestras independientes son diferentes. Por ejemplo, la comparación de medidas de respuesta tipo curación, fracaso y/o evolución en distintos tratamientos corresponden a este caso.

11.2.1. Comparación de dos proporciones independientes

En caso de comparar una variable ordinal o nominal en función de dos categorías, estamos queriendo comparar una variable con dos categorías con otra con dos categorías. Imaginemos que se trata de comparar la proporción de pacientes que presentan desnutrición en función del tipo de tratamiento sustitutivo empleado. Presentar desnutrición es una variable binomial puesto que sólo admite dos valores: presenta desnutrición o no presenta desnutrición y la variable tratamiento sustitutivo, en este caso engloba HD y DP, es decir, dos categorías. Nos encontramos pues, ante un contraste de proporciones donde, intuitivamente podemos pensar que si ambos tratamientos fueran iguales, presentarían la misma proporción de pacientes desnutridos (H_0), mientras que si no lo fueran, la diferencia entre las dos proporciones no incluiría el cero (H_1).

Para el contraste de dos proporciones se empleará la prueba Z, que mediante la aproximación a la distribución normal, calculará el estadístico de contraste para la diferencia de proporciones. Por otra parte, otro procedimiento generalizado sería la utilización de la prueba Ji-Cuadrado para tablas de frecuencias de 2x2.

11.2.1.1. Prueba Z

Esta prueba se basa en la aproximación normal de la distribución binomial. Queremos comparar dos proporciones, p_1 y p_2, observadas en dos grupos distintos de tamaños n_1 y n_2, respectivamente. Esta prueba es utilizable cuando los tamaños muestrales n_1 y n_2 son grandes, para poder aplicar el Teorema Central del Límite. El estadístico de contraste se calcula como:

$$Z = \frac{p_1 - p_2}{EED} = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

El estadístico Z sigue una distribución Normal (0, 1). El intervalo de confianza se obtiene mediante la fórmula $(p_1-p_2) \pm Z_\alpha * EED$, donde EED corresponde al error estándar de la diferencia de proporciones tal como se calcula en la fórmula anterior.

En esta prueba se utiliza la distribución normal como aproximación de la solución exacta de intervalos de confianza para proporciones, adecuada siempre que n sea mayor o igual a 30 y las frecuencias absolutas y las esperadas sean superiores a 4. El hecho de poder utilizar la distribución normal, nos permite asociar un intervalo de confianza a la diferencia de proporciones.

11.2.1.2. Prueba de la Ji-Cuadrado (X2)

La prueba de la Ji-Cuadrado es una de las pruebas mas frecuentemente utilizadas para el contraste de variables cualitativas, aplicándose para comparar si dos características cualitativas están relacionadas entre sí, si varias muestras de carácter cualitativo proceden de igual población o si los datos observados siguen una determinada distribución teórica.

Para su cálculo se calculan las frecuencias esperadas (las que deberían haberse observado si la hipótesis de independencia fuese cierta), para compararlas con las observadas en la realidad. Se calcula el valor del estadístico χ^2 como:

$$\chi^2 = \sum \frac{\left|O_{ij} - E_{ij}\right|^2}{E_{ij}} \sim \chi^2_{(f-1)\,(c-1)}$$

donde:
- O_{ij} corresponden a las frecuencias observadas dentro de la casilla de la fila i y columna j.
- E_{ij} corresponden a las frecuencias esperadas o teóricas
- f es el número de filas y c el número de columnas.
- (f -1)*(c-1) corresponden a los grados de libertad de la distribución del estadístico de contraste.

El primer paso consiste en construir la tabla de contingencia asociada a las dos variables a analizar. A partir de ella se calculan las frecuencias esperadas en cada casilla bajo la suposición de que ambas variables sean independientes.

En el caso más sencillo de una tabla 2x2, las frecuencias esperadas se calcularían, basándonos en la Tabla 7 como:

$$E_{11} = \frac{(a+b)*(a+c)}{n} \qquad E_{21} = \frac{(c+d)*(a+c)}{n}$$

$$E_{12} = \frac{(a+b)*(b+d)}{n} \qquad E_{22} = \frac{(c+d)*(b+d)}{n}$$

A partir de estas fórmulas el estadístico de contraste χ^2 puede simplificarse y obtenerse de manera más sencilla a partir de la fórmula:

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

En el caso de una tabla de contingencia de f filas y c columnas, las frecuencias esperadas se pueden obtener de manera similar, como se describe en la siguiente tabla f x c:

	A_1	A_2	...	A_c	Total
Y_1	$E_{11}=\frac{f_{1\cdot}*f_{\cdot 1}}{f_{\cdot\cdot}}$	$E_{12}=\frac{f_{1\cdot}*f_{\cdot 2}}{f_{\cdot\cdot}}$		$E_{1c}=\frac{f_{1\cdot}*f_{\cdot c}}{f_{\cdot\cdot}}$	$f_{1.}$
Y_2	$E_{21}=\frac{f_{2\cdot}*f_{\cdot 1}}{f_{\cdot\cdot}}$	$E_{22}=\frac{f_{2\cdot}*f_{\cdot 2}}{f_{\cdot\cdot}}$		$E_{2c}=\frac{f_{2\cdot}*f_{\cdot c}}{f_{\cdot\cdot}}$	$f_{2.}$
...					
Y_f	$E_{f1}=\frac{f_{f\cdot}*f_{\cdot 1}}{f_{\cdot\cdot}}$	$E_{f2}=\frac{f_{f\cdot}*f_{\cdot 2}}{f_{\cdot\cdot}}$		$E_{fc}=\frac{f_{f\cdot}*f_{\cdot c}}{f_{\cdot\cdot}}$	$f_{f.}$
Total	$f_{.1}$	$f_{.2}$		$f_{.c}$	$f_{..}$

Tabla 8. Cálculo de las frecuencias esperadas en una tabla de contingencia

Para obtener el valor de la Ji-cuadrado las frecuencias observadas se comparan con los valores observados. Así, cuando mayor sea la diferencia entre los valores esperados y los observados mayor será el valor del estadístico, existiendo en este caso asociación entre las variables comparadas. El hecho de que las diferencias se eleven al cuadrado convierte cualquier diferencia en positiva, lo que indica si existe o no relación entre los factores pero no en que sentido se produce tal asociación.

Bajo la hipótesis nula de independencia, el estadístico Ji-cuadrado se distribuye según una distribución Ji-cuadrado con (f-1)*(c-1) grados de libertad. En el caso de tablas 2x2 los grados de libertad son 1.

Cuando el tamaño muestral no es demasiado grande, puede introducirse algún sesgo en los cálculos, ya que estos contrastes aproximan una distribución discreta por una continua. En caso de que más del 20% de las frecuencias esperadas sean menores de 5 o bien alguna celda tenga valores esperados inferiores a 2, se utiliza una corrección para eliminar este sesgo, conocida como la corrección de Yates para continuidad, aplicable en el caso de tablas 2x2. La **corrección de Yates** da un resultado más conservador y, siguiendo la notación utilizada en la Tabla 7, se calcularía como:

$$\chi^2 = \frac{n\left(|ad - bc| - \frac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Ejemplo. Imaginemos que se desea estudiar si existe relación entre el cintigrama renal (CR) y la presencia de reflujo vesicoureteral (RVU) en niños con primera pielonefritis aguda. Para ello se seleccionaron 127 niños a los que se les realizó los cintigramas renales y se evaluó la presencia de RVU. Para analizar la relación entre la positividad de RVU entre los niños con CR normal y CR alterado se utilizó la prueba de la Ji-Cuadrado.

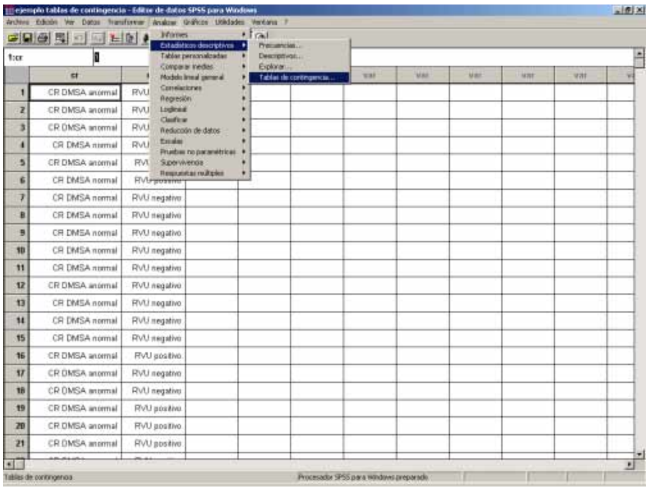
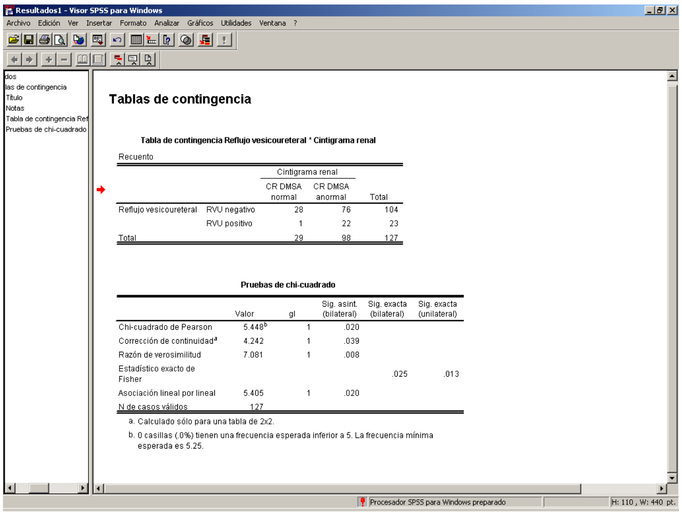


Figura 41. Comparación de dos proporciones en SPSS, prueba de la Ji-Cuadrado

En los resultados de la Figura 42 se puede observar, en primer lugar, una tabla de contingencia 2x2, con el número de casos en cada una de las celdas de la tabla. El valor de significación de la Ji-Cuadrado es de 0,020 (inferior a 0,05), indicando que existe una relación estadísticamente significativa entre el cintigrama renal y la presencia de reflujo vesicoureteral. En nuestro ejemplo, al tratarse una tabla 2x2 puede aplicarse la corrección de Yates, siendo en este caso el valor de significación 0,039, indicando también la existencia de una relación estadísticamente significativa. El programa SPSS nos proporciona además, en el caso de las tablas 2x2, el estadístico exacto de Fisher. Este estadístico se utiliza cuando más del 20% de las frecuencias esperadas son inferiores a 5. El cumplimiento de los criterios de aplicabilidad aparece a continuación de la tabla con los distintos estadísticos y la significación estadística correspondiente.



Tablas de contingencia

Tabla de contingencia Reflujo vesicoureteral * Cintigrama renal

Recuento

		Cintigrama renal		
		CR DMSA normal	CR DMSA anormal	Total
Reflujo vesicoureteral	RVU negativo	28	76	104
	RVU positivo	1	22	23
Total		29	98	127

Pruebas de chi-cuadrado

	Valor	gl	Sig. asint. (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	5.448[b]	1	.020		
Corrección de continuidad[a]	4.242	1	.039		
Razón de verosimilitud	7.081	1	.008		
Estadístico exacto de Fisher				.025	.013
Asociación lineal por lineal	5.405	1	.020		
N de casos válidos	127				

a. Calculado sólo para una tabla de 2x2.

b. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 5.25.

Figura 42. Resultados en la comparación de dos proporciones en SPSS, prueba de la Ji-Cuadrado

Para terminar este apartado, es importante mencionar que la prueba χ^2 es sólo aproximada. Existe, sin embargo, un test *exacto* para las tablas 2x2. Se denomina **test de Fisher** y es el que se explica a continuación.

11.2.1.3. Test Exacto de Fisher

El test exacto de Fisher permite analizar la asociación entre dos variables dicotómicas cuando no se cumplen las condiciones necesarias para la aplicación del test de la Ji-cuadrado. Como ya mencionamos anteriormente, para aplicar la prueba de la Ji-cuadrado se exige que el 80% de las celdas presenten frecuencias esperadas superiores a 5. Así, en las tablas 2x2 es necesario que se verifique en todas sus celdas, aunque en la práctica se permite que una de ellas se muestre ligeramente por debajo. El test de Fisher se aplica también cuando alguno de los valores esperados es inferior a 2.

Esta prueba se basa en el cálculo de la probabilidad exacta de las frecuencias observadas. Evalúa la probabilidad asociada a cada una de las tablas 2x2 que se pueden formar manteniendo los mismos totales de filas y columnas que los de la tabla observada. La probabilidad exacta de observar un conjunto concreto de frecuencias a, b, c y d en una tabla 2x2, cuando se asume independencia y los totales de filas y columnas se consideran fijos, viene dada por una distribución hipergeométrica:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

Esta probabilidad se calcula para todas las tablas de contingencia que puedan formarse con los mismos totales que en la tabla observada, utilizándolos para calcular el valor de la p asociado al test de Fisher. El valor de p puede calcularse sumando aquellas probabilidades inferiores a la probabilidad de la tabla observada. Si el valor de p es pequeño ($p<0,05$) se debe rechazar la hipótesis nula de independencia, asumiendo que ambas variables están asociadas estadísticamente.

11.2.2. Comparación de tres o más proporciones independientes

Una de las ventajas de la Ji-cuadrado en comparación con la Z, es que puede ser utilizada para comparar más de dos proporciones. La fórmula es la presentada en el aparatado 11.2.1.Comparación de dos proporciones independientes, teniendo presente que para determinar el grado de significación debe utilizarse los grados de libertad adecuados a la tabla de contingencia.

En el caso de que la prueba de la Ji-cuadrado no sea aplicable (más del 20% de las casillas con una frecuencia esperada inferior a 5 y ninguna inferior a 2) sólo existe una única alternativa que consiste en agrupar categorías de la variable para aumentar así el número de casos de cada casilla, siempre que tenga sentido clínico hacerlo.

11.3. Comparación de dos proporciones para datos apareados

En un estudio se plantea evaluar si una determinada variación en la posología de un fármaco es capaz de presentar la misma adherencia al tratamiento, respecto a la técnica habitual. En este caso, el método de comparación de proporciones deberá considerar que estas muestras están apareadas, puesto que serán los mismos sujetos los que primero probarán una posología y luego otra. Se trata de estudios "pre-post" en los que la respuesta se evalúa al cabo de un tiempo en el cual se ha variado alguna de las condiciones basales.

11.3.1. Comparación de dos proporciones apareadas

La comparación de dos proporciones en muestras apareadas, se estudiará mediante las pruebas asociadas a la tabla de contingencia. La comparación entre dos proporciones apareadas puede realizarse utilizando el estadístico de contraste Z; sin embargo, la prueba más conocida corresponde a la prueba de McNemar.

11.3.1.1. Test de McNemar

Este test se utiliza cuando se trata de comparar dos proporciones observadas en el mismo grupo de individuos en dos ocasiones distintas de tiempo (antes y después de algún estímulo). Se pretende comparar si se produce algún cambio significativo entre ambas mediciones. Clasificamos un grupo de individuos entre dos categorías mutuamente excluyentes, indicadas por + (positivo) y – (negativo). Pasado un estímulo o intervención es posible que alguno de estos individuos cambie de categoría, de manera que la tabla de frecuencias que se obtendría sería la siguiente:

		Después		
		Positivo	Negativo	Total
Antes	Positivo	a	b	a+b
	Negtivo	c	d	c+d
	Total	a+c	b+d	n

Tabla 9. Tabla general de contingencia para dos proporciones observadas en un mismo grupo en dos ocasiones distintas de tiempo

La proporción de individuos con la característica positiva antes sería $p1 = \frac{a+b}{n}$ y después sería $p2 = \frac{a+c}{n}$

Nos interesa contrastar si la diferencia entre estas dos proporciones es cero (hipótesis nula) frente a que $p1$ y $p2$ sean diferentes ($p1 - p2 = \frac{b-c}{n} \neq 0$). Para ello, nos podemos centrar en las celdas b y c que son las que muestran discordancia entre las dos mediciones, contrastando si el número de individuos que tras la intervención han dejado de presentar la característica + (b) es el mismo que el número de individuos que tras la intervención han realizado el cambio inverso (c), es decir han dejado de presentar la característica -. El error estándar para la diferencia entre dos proporciones es:

$EED = \frac{1}{n}\sqrt{b+c-\frac{(b-c)^2}{n}}$ que bajo la hipótesis nula(b-c=0) se reduce a $EED = \frac{1}{n}\sqrt{b+c}$

El estadístico de contraste que sigue una distribución Normal (0,1) se calcula como:

$$z = \frac{p1 - p2}{EED} = \frac{\frac{b-c}{n}}{\frac{1}{n}\sqrt{b+c}} = \frac{b-c}{\sqrt{b+c}}$$

También se puede considerar el estadístico de contraste: $\chi^2 = \frac{(b-c)^2}{(b+c)}$ que sigue una distribución Ji-cuadrado con 1 grado de libertad. Como en el caso de la χ^2, si las frecuencias son pequeñas puede utilizarse la corrección de Yates:

$$\chi^2 = \frac{(|b-c|-1)^2}{(b+c)}$$

11.3.1.2. Sensibilidad y especificidad

Una particularidad de la comparación de proporciones para datos apareados es la comparación de métodos diagnósticos. En este caso, la prueba de diagnóstico presenta unos valores "a priori" o "pre" y es posteriormente cuando se sabe en realidad si los resultados son correctos o no "a posteriori" o "post". Los términos de sensibilidad y especificidad, por tanto, se utilizan cuando se quieren comparar dos métodos diagnósticos, uno de ellos considerado de referencia (definido anteriormente como "post").

Se define la **sensibilidad** como la proporción de enfermos identificados correctamente por la prueba diagnóstica, o dicho de otra manera, la probabilidad de que para un individuo enfermo obtenga en la prueba realizada un resultado positivo. Por lo tanto la sensibilidad es la capacidad que tiene la prueba para detectar la enfermedad.

La **especificidad** es la proporción de no enfermos que son identificados correctamente por la prueba diagnóstica, es decir, la probabilidad de que para un individuo sano obtenga un resultado negativo. Por tanto se puede definir como la capacidad para detectar individuos sanos.

Según la notación utilizada en la siguiente tabla, la sensibilidad se calculará como el número de pacientes con valoración diagnóstica positiva según la prueba realizada y con diagnóstico final positivo (a) sobre el total de pacientes con diagnóstico final positivo (a+c). La especificidad se calculará como el número de pacientes con valoración diagnóstica negativa según la prueba realizada y diagnóstico final negativo (d) sobre el total de pacientes con diagnóstico final negativo (b+d).

		Diagnóstico final		Total
		Positivo	Negativo	
Prueba	Positivo	a	b	a+b
	Negtivo	c	d	c+d
	Total	a+c	b+d	N

Tabla 10. Tabla utilizada para el cálculo de la sensibilidad y especificidad de una prueba diagnóstica

Asociado también a estas pruebas, el **valor predictivo positivo** (VPP) se define como la probabilidad de que los pacientes que tienen la prueba positiva tengan la enfermedad, y el **valor predictivo negativo** (VPN) corresponde a la probabilidad de que los pacientes que tienen la prueba negativa no tengan la enfermedad. Por tanto, el cálculo de la sensibilidad, especificidad, VPP y VPN de las pruebas diagnósticas será el siguiente:

Sensibilidad (S)	S = a/(a+c), IC = ± 1,96 ÷(S(1-S)/(n), donde n=a+c
Especificidad (E)	E = d/(b+d), IC = ± 1,96 ÷(E(1-E)/n), donde n=b+d
Valor predictivo positivo (VPP)	VPP = a / (a+b) IC = ± 1,96 ÷(VPP(1-VPP)/n), donde n=a+b
Valor predictivo negativo (VPN)	VPN = d / (c+d) IC = ± 1,96 ÷(VPN(1-VPN)/n), donde n=c+d

Tabla 11. Calculo de la sensibilidad, especificidad, VPP y VPN de las pruebas diagnósticas

En el siguiente ejemplo tenemos una prueba diagnóstica que se desea comparar con la prueba de referencia utilizada hasta el momento en el diagnóstico de una determinada enfermedad. Los valores obtenidos, son los siguientes:

		Diagnóstico final		
		+ Positivo	- Negativo	Total
Prueba nueva	+ Positivo	76	8	84
	- Negtivo	36	32	68
	Total	112	40	152

Tabla 12. Tabla de frecuencias utilizada en el ejemplo de cálculo de sensibilidad, especificidad, VPP y VPN de una prueba diagnóstica

La sensibilidad de la prueba corresponderá a 76/112=0,68, mientras que la especificidad será de 32/40=0,8. De esta forma, podremos decir que la prueba nueva es capaz de predecir correctamente el 68% de los pacientes realmente positivos y el 80% de los negativos. Paralelamente, el valor predictivo positivo se calculará como 76/84=0,90 y el valor predictivo negativo como 32/68=0,47; lo cual, significa que si el resultado de la prueba es positivo, tenemos el 90% de probabilidad de que el paciente realmente sea positivo, mientras que si el resultado es negativo, sólo estaremos seguros en un 47% de que realmente su diagnóstico final es negativo.

Una prueba es sensible y específica si es capaz de identificar a todos los que están enfermos como tales y a todos los sanos como tales. Es difícil tener una prueba muy sensible y a la vez muy especifica, hay que tener información sobre los dos aspectos. La sensibilidad y especificidad son la información que necesitamos para decidir la validez de una prueba y nos ayudará a decidir si utilizar una prueba diagnostica o no (validez de criterio).

11.3.1.3. Curvas ROC

Los resultados de una prueba diagnostica no tienen por qué ser dicotómicos, pueden ser continuos. Muchas variables cuantitativas y continuas las transformamos en dicotómicas para poder tomar una decisión de tratar o no a alguien. Es por eso que deberemos conocer cual es el punto de corte que nos definirá el límite para iniciar o no iniciar el tratamiento.

Las Curvas ROC son un procedimiento estadístico que permiten seleccionar el punto de corte maximizando a la vez la sensibilidad y la especificidad. El análisis y la determinación de la curva de ROC se hacen mediante la determinación de la sensibilidad y especificidad, es decir, ver la fiabilidad en el máximo número posible de puntos de la prueba. Se realizan trazando un diagrama en el que la ordenada (eje y) es la sensibilidad y la abscisa (eje x) es el valor 1-especificidad, tal como se muestra en la Figura 43. Cuanto más sensible y especifica sea la prueba (representación: puntos más hacia arriba y más hacia la izquierda) más se alejará de la diagonal y mejor será el punto de corte seleccionado.

Una vez preparada la curva ROC se seleccionará como punto de corte aquel punto mas alejado de la diagonal, que corresponde al valor que presenta una mayor sensibilidad y especificidad a la vez.

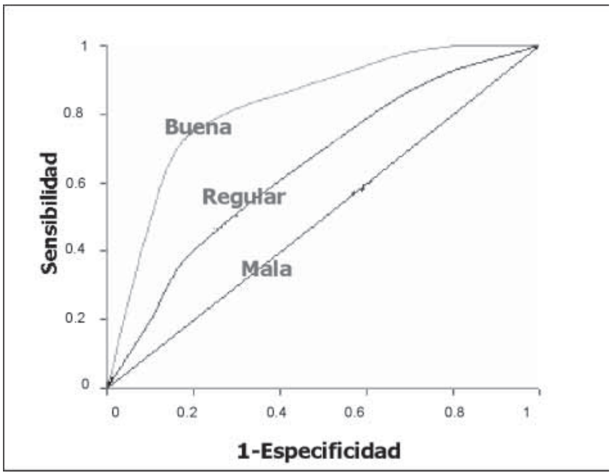


Figura 43. Tipo de curvas ROC

11.3.1.4. Indice Kappa

El índice Kappa se utiliza cuando se quieren comparar dos métodos pero no existe uno de referencia. Este índice mide el grado de concordancia entre ambos y se define como:

$$k = \frac{P_o - P_e}{1 - P_e}$$

donde P_0 es la proporción de acuerdos observados y P_e la proporción de acuerdos esperados en la hipótesis de independencia entre los métodos, es decir de acuerdos por azar. En el caso de variables ordinales, existe también el **índice Kappa ponderado**, que consiste en asignar unos pesos para cuantificar la importancia relativa a los desacuerdos. Las variables utilizadas normalmente corresponden a valoraciones del tipo: nada, leve, moderado, severo, que corresponde a una escala categórica ordinal.

Existen otras medidas de asociación para dos variables en escala ordinal que son parecidas a las medidas de asociación entre variables continuas (coeficientes de correlación), en interpretación y en cálculo. Las medidas de asociación más utilizadas, aunque no estén incluidas en este capítulo son: Gamma, Tau-b y Tau-c de Kendall y D de Somers. La explicación del índice Kappa lo veremos en más profundidad en el apartado 13 'Medidas de asociación'.

11.3.2. Comparación de tres o más proporciones apareadas

Para comparar tres o más proporciones relacionadas entre sí, usaremos la **Q de Cochran**. Este procedimiento es la generalización del test de McNemar para dos proporciones relacionadas. De hecho, si el número de proporciones es igual a dos, el estadístico de McNemar y el de Cochran coinciden.

Si tenemos J proporciones, P_{+j} representan las proporciones de aciertos de cada una de las muestras: $P_{+j} = \frac{T_{+j}}{n}$ (T_{+j} es la suma de aciertos de cada muestra). Para contrastar la hipótesis de igualdad entre J proporciones, se calcula el estadístico, que se distribuye según χ^2 con J-1 grados de libertad, como:

$$Q = \frac{J(J-1)\sum T_{+j} - (J-1)T^2}{JT - \sum T_{i+}^2}$$

11.4. Consideraciones importantes

Según a qué correspondan las proporciones que se desean comparar, se deberá seleccionar la prueba estadística adecuada. Supongamos que en una muestra se desea analizar el efecto de dos tratamientos dermatológicos para el acné A y B, presentando tres tipos de respuestas: curación, mejoría y fracaso. Tras un periodo de observación, se contabilizan los resultados de cada grupo. Para determinar si la probabilidad de obtener un resultado favorable es independiente de cuál sea el tratamiento se aplicará la prueba χ^2 (Ji-cuadrado), que contrasta la hipótesis nula de que tratamiento y curación son independientes. Imaginemos que queremos simplificar el estudio y consideramos que una mejoría en el acné todavía no es una curación y lo consideramos fracaso. De este modo, tendremos una tabla 2x2 analizable con la misma prueba, y además, podremos estimar la prueba Z, y mediante su intervalo de confianza, la diferencia obtenida entre el porcentaje de curaciones del grupo tratado con el fármaco A y el tratado con el fármaco B. Si el tamaño de muestra fuera pequeño o alguno de los efectivos menor a 5, utilizaríamos el test exacto de Fisher.

Cuando el tipo de estudio es de tipo "antes-después", es decir con datos apareados, las pruebas que se utilizan son menos intuitivas. En caso de ser dos momentos en el tiempo y dos respuestas estaríamos ante una tabla 2x2 y podríamos utilizar el test de McNemar. En comparaciones de más de dos proporciones se utilizaría su generalización, la prueba Q de Cochran.

El caso de las tablas 2x2 para proporciones apareadas, presenta particularidades cuando se están estudiando pruebas diagnósticas. En este supuesto, se juntan varios conceptos tales como la probabilidad "a priori", el teorema de

Bayes, la estadística bayesiana y, recientemente en este capítulo, los términos de sensibilidad y especificidad de las pruebas. Con estos conceptos destinados a conocer las probabilidad de reconocer los valores reales positivos y negativos, respectivamente, aparecen otros como son el valor predictivo positivo y el valor predictivo negativo, que estiman las probabilidades de conocer si realmente presentan diagnóstico positivo o negativo los individuos que han presentado el resultado en la prueba positiva o negativa, respectivamente. Como ya hemos comentado, la sensibilidad y la especificidad de una prueba nos ofrecen la información que necesitamos para decidir su validez en el diagnóstico de ciertas patologías. Para finalizar, se presenta el siguiente esquema con las pruebas a utilizar cuando se pretenden comparar proporciones:

Comparar 2 grupos independientes	Prueba Z Prueba de Fisher Prueba Ji-cuadrado
Comparar 2 grupos apareados	Prueba de McNemar
Comparar 3 ó más grupos independientes	Prueba Ji- cuadrado
Comparar 3 ó más grupos apareados	Cochran Q

Tabla 13. Resumen del capítulo