

Rincón iconoclasta

Parte III de III. Los números: la significancia clínica o los estimadores de intervalo

Sergio Sobrino Cossio,* Arturo Reding Bernal,** Lucía Monserrat Pérez Navarro,*** Juan Carlos López Alvarenga**

Esta parte da por terminado el tema del protocolo de investigación. En las dos primeras se destacó la importancia de construir una buena hipótesis y de elegir un diseño apropiado que responda a las necesidades de la investigación.

Una hipótesis estructuralmente bien planteada contiene:

- El objetivo del estudio; es decir, lo que el investigador quiere conocer y probar. No es una receta de cocina; sin embargo, es aquí donde se descubre si el trabajo tiene como propósito analizar la prevalencia de una enfermedad, la causalidad o la fuerza de asociación de las variables, el riesgo o pronóstico, el diagnóstico, la eficacia o la efectividad de un tratamiento, el costo-beneficio, la calidad de la atención médica, las variables fármaco-económicas, la genómica, etc. De esto dependerá la evaluación crítica metodológica del trabajo.

- Las variables (predictoras y de desenlace) y su escala de medición (nominales, ordinales o numéricas).
- El diseño de investigación de acuerdo con la dirección del análisis (prospectivo o retrospectivo), que permite evaluar si es conveniente realizar un estudio de casos-controles, de cohortes o clínico controlado.
- Finalmente, los métodos, técnicas y procedimientos para la investigación; es decir, se eligen los diferentes métodos estadísticos según las escalas de medición de las variables y de lo que se quiere conocer: el riesgo, la diferencia de promedios o las proporciones.

ESTUDIOS DE CASOS Y CONTROLES¹

Son estudios observacionales en donde no hay intervención del investigador. A los sujetos que sufren el fenómeno o enfermedad se les denomina casos, y se comparan con otros individuos sanos a los que se les designa controles.

Efecto  Exposición

Con este tipo de estudios se valora si existe asociación entre la enfermedad (efecto) y los factores de riesgo (exposición). Este diseño es retrospectivo, por lo que en lugar de estimar el riesgo absoluto (estudios prospectivos), permite calcular de manera aproximada el riesgo relativo, denominado razón de momios (RM, *odds ratio*, en inglés), al concentrar los datos numéricos en una tabla de 2 x 2 o de cuatro casillas (Cuadro 1), donde se representa la frecuencia de exposición y la frecuencia de enfermedad. Se determina la razón de momios (*odds ratio* o RM) para conocer en los enfermos cuántas veces más se incrementa

* Departamento de Endoscopia del Instituto Nacional de Cancerología.

** Coordinación de Recursos en Bioinformática y Estadística. Dirección de Investigación del Hospital General de México.

*** Práctica privada, Xalapa, Veracruz.

Correspondencia: Dr. Juan Carlos López Alvarenga. Core de Bioestadística del Hospital General de México. Dr. Balmis núm. 148, col. Doctores, CP 06726, México, DF. Correo electrónico: jalvaren@sfbgenetics.org, jclalvar@yahoo.com

Este artículo debe citarse como: Sobrino-Cossio S, Reding-Bernal A, Pérez-Navarro LM, López-Alvarenga JC. Parte III de III. Los números: la significancia clínica o los estimadores de intervalo. Dermatol Rev Mex 2010;54(4):228-235.

www.nietoeditores.com.mx

Cuadro 1. Estudio de casos y controles

	Casos	Controles
Expuestos	a	b
No expuestos	c	d

a: casos expuestos; b: controles expuestos; c: casos no expuestos; d: controles no expuestos.

el riesgo en los individuos expuestos. Por ejemplo, si la razón de momios de pacientes con cáncer gástrico es de 6 en los que tienen infección crónica por *Helicobacter pylori*, puede decirse que estos individuos (Hp+) están en un riesgo seis veces mayor de padecer cáncer de estómago que los que no la tienen (Hp-).¹⁻³

La razón de momios (RM) se obtiene al multiplicar a por d y dividir el producto entre el resultado de b por c ($RM = a \cdot d / b \cdot c$). La fuerza de asociación es la probabilidad de desarrollar un efecto específico después de la exposición *versus* no exposición. El término “riesgo” implica la probabilidad de consecuencias adversas. La probabilidad de que personas sanas, pero expuestas a ciertos factores, contraigan la enfermedad, se conoce como riesgo.¹⁻³

En 1994, la Agencia Internacional para la Investigación del Cáncer (OMS) consideró a *Helicobacter pylori* (Hp) un representante del grupo I (exposición a carcinógeno). Se determinó que existe evidencia suficiente en humanos para demostrar que la infección incrementa el riesgo de cáncer. Los estudios que revelaron la asociación de Hp con el cáncer gástrico fueron de casos y controles (CC) con razón de momios (RM, odds ratio [OR] en inglés) entre 3 y 6. En este ejemplo, se observa riesgo aumentado, ya que la razón de momios fue mayor de 1 (Cuadro 2).

ESTUDIOS DE COHORTE³

En ellos, un grupo de sujetos con factor de riesgo (FR) (por ejemplo: fumar, infección por *Helicobacter pylori*)

Cuadro 2. Riesgo de cáncer gástrico por infección de *Helicobacter pylori* según varios autores

Autor	Diseño	Casos (n)	Controles (n)	RM o RR
Parsonnet ⁴	CC	109	109	+ 3.6
Nomura ⁵	CC	109	109	+ 6.0
Forman ⁶	CC	29	116	+ 2.7
Hanson ⁷	CC	112	123	+ 2.6
Kokola ⁸	CC	36	22	+ 6.0

o factor pronóstico (por ejemplo: pacientes en terapia intensiva con criterios de Ranson o Baltazar para pancreatitis) se compara con otro de individuos que carecen de estos factores, el cual es seguido en el tiempo (medición longitudinal) para ver si desarrolla el efecto (desenlace; por ejemplo, muerte, complicaciones, etc.) y se analizan las diferencias.



Se evalúa la asociación entre FR (exposición) y la enfermedad (efecto). Este diseño de estudio es prospectivo, por lo que se calcula el riesgo relativo (RR) al concentrar los datos numéricos en una tabla de 2 x 2 o de cuatro casillas (Cuadro 3), representando la frecuencia de exposición y la frecuencia de enfermedad. Se calcula el RR para saber cuántas veces más se incrementa el riesgo en los sujetos expuestos.^{1,3}

Cuadro 3. Estudio de cohorte

	Enfermedad	
	Sí	No
Expuestos	a	b
No expuestos	c	d

El riesgo relativo (RR) es la razón de la probabilidad de que un evento ocurra entre el grupo expuesto *versus* el no expuesto ($RR = P \text{ expuesto} / P \text{ no expuesto}$) y se calcula: $RR = (a/a + b) / (c/c + d)$.⁹

- RR 1 = no hay asociación entre la existencia del factor de riesgo y el evento.
- $RR < 1$ = hay menos riesgo de que ocurra el evento en el grupo expuesto a un factor que en el no expuesto.
- $RR > 1$ = es mayor el riesgo de que ocurra el evento en el grupo expuesto a un factor que en el no expuesto.

En el Cuadro 4 se compara el riesgo de que suceda un evento entre ambos grupos. Si este riesgo es de 45/450 (0.1) en mujeres y de 200/500 (0.4) en hombres, el RR de muerte es de 0.25 (0.1/0.4); es decir, las mujeres tienen menor riesgo (ser mujer es un factor protector) de muerte que los hombres. Otra forma de analizarlo es obtener el inverso de RR: $1/0.25 = 4$, y restarle 1, que equivale al

Cuadro 4. Ejemplo de cálculo del riesgo

	<i>Efecto</i>	<i>No efecto</i>	<i>Total de sujetos</i>	<i>Tasa de eventos</i>
Expuestos	45 (a)	405 (b)	a + b = 450	a/(a + b) = 0.1 (10%)
No expuestos	200 (c)	300 (d)	c + d = 500	c/(c + d) = 0.4 (40%)
Total	a + c = 245	b + d = 705	a + b + c + d = 950	

valor neutro; esto indica que las mujeres tienen tres veces más protección que los hombres (75%).^{2,10} Continuando con este ejemplo, en el Cuadro 5 se muestra el cálculo de distintos tipos de riesgo que pueden obtenerse a partir de un estudio de cohorte.

ENSAYO CLÍNICO²

Los estudios aleatorios, doble-ciego y placebo-control son diseños experimentales que tienen la mayor fuerza y confiabilidad para establecer causalidad o fuerza de asociación, riesgo y respuesta al tratamiento. Requiere una intervención (variable independiente), seguimiento en el tiempo (longitudinal) y comparación de resultados. Así, puede controlarse y compararse la situación, lo que aumenta la solidez de los resultados del estudio (validez interna) para generalizarlos a la población (validez externa).

¿Cómo es que la adición de una maniobra experimental puede explicar los resultados comparativos? Al formar dos o más grupos paralelos (independientes) de sujetos semejantes que comparten características o están afectados por la misma enfermedad, la asignación aleatoria a un grupo (de estudio o placebo) disminuye el error y

la influencia de los confusores; así, un sujeto tiene igual probabilidad de pertenecer a uno u otro grupo, lo que disminuye la tendencia del investigador. Este diseño se aproxima al experimento biológico verdadero. El cegamiento no permite que el investigador (ciego), el sujeto, o ambos (doble ciego), conozcan a qué grupo pertenece y, por tanto, minimiza el sesgo.

- Es indispensable señalar criterios diagnósticos, espectro de la enfermedad, comorbilidad y criterios de selección (inclusión, exclusión y eliminación), y explicar la forma de la asignación de pacientes a los grupos.
- Para la maniobra se necesita una descripción detallada del fármaco (fórmula, dosis, esquema, etc.), maniobra educacional o intervención que haya sido aplicada, la cual se compara contra el mejor tratamiento o se determina por qué se usó un placebo. También hay que consignar eventos adversos, ya sean o no graves, su inicio y término.
- Se deben comentar los criterios objetivos o variables que se miden (escalas: nominales, ordinales y cuantitativas) y los instrumentos de medición, así como detallar el análisis estadístico. Con ello se evalúa la reproducibilidad y la exactitud.

Cuadro 5. Cálculo de riesgos de un estudio de cohorte

<i>Ecuaciones</i>	<i>Variable</i>
Tasa de eventos de riesgo relativo = efecto maniobra (Y)/efecto control (X)	< 0: reducción del riesgo relativo (-) 0.75, o (-) 75% (RRR)
Reducción del riesgo relativo (RRR) = 1 - (0.15/0.20 = 0.75) = 0.25 = 1 - RR	
Tasa de eventos maniobra (X) - control (Y) control (Y)	
$\frac{1}{\text{maniobra (X) - control (EGC)}}$	< 0: número necesario para tratar (NNT) (-) 3.33
(a/a + b)/(c/c + d)	Riesgo relativo (RR) 0.25
a/(a + b) - c/(c + d)	Riesgo atribuible (RA) (-) 0.39 o (-) 39%
1 - RR (o 1 - RM)	Fracción preventiva 0.75 o 75%

ESTADÍSTICA PARA LAS VARIABLES CATEGÓRICAS: JI AL CUADRADO (CHI CUADRADA, CHI SQUARE)¹¹⁻¹⁴

Ejemplo clínico: se evaluó la asociación entre el reflujo gastroesofágico diurno y nocturno con la exposición esofágica al ácido en 24 horas en 59 pacientes que habían tenido pirosis cuatro veces a la semana en los últimos seis meses (Cuadro 6).

Cuadro 6. Frecuencias observadas: reflujo gastroesofágico (diurno o nocturno) y acidez esofágica (24 horas)*

Acidez esofágica (24 horas)	RGE sí	RGE no	Frecuencia marginal
Mayor	22	15	37
Menor	12	10	22
Frecuencia marginal	34	25	59

* Observaciones en dos grupos independientes.

El ejemplo trata de relacionar la vinculación entre estas dos variables nominales (categóricas): reflujo gastroesofágico nocturno (sí o no) y acidez esofágica en 24 horas (sí o no).

La ji al cuadrado (χ^2) es una estadística para comparar proporciones. Las variables nominales (categóricas) obtenidas de una muestra con n observaciones independientes se ordenan en una tabla de frecuencias o contingencia (renglones y columnas). En cada celda se coloca el número de unidades, combinando categorías de las dos variables. Es una tabla bidireccional en donde las variables se relacionan con la hipótesis de trabajo y se prueba si existe independencia o relación entre ellas (criterio de independencia). Es importante que la categoría o clase sea mutuamente excluyente; por ejemplo, un paciente sólo puede categorizarse como con reflujo nocturno positivo o negativo. La distribución de frecuencias origina frecuencias observadas entre renglones y columnas (cuatro celdas: 1,1, 1,2, 2,1 y 2,2) y frecuencias marginales (número de observaciones en cada columna y en cada renglón).

Si se multiplican los valores marginales (37, 22, 34 y 25) y se dividen entre el número total de observaciones ($n = 59$), se obtiene la frecuencia esperada.

- Celda 1,1 = $E_{1,1} = (37)(34)/(59) = 21.32$
- Celda 1,2 = $E_{1,2} = (37)(25)/(59) = 15.67$
- Celda 2,1 = $E_{2,1} = (22)(34)/(59) = 12.67$

- Celda 2,2 = $E_{2,2} = (22)(25)/(59) = 9.32$

Recuérdense las frecuencias:

observadas: celda 1,1 (22), celda 1,2 (15), celda 2,1 (12) y celda 2,2 (10).

esperadas: celda 1,1 (21.32), celda 1,2 (15.67), celda 2,1 (12.67) y celda 2,2 (9.32).

En cada celda se resta: valor observado - valor esperado (ejemplo 22 - 21.32), y la resultante se eleva al cuadrado (0.46); después se divide el numerador entre el valor esperado de cada celda (0.46/21.32).

- Celda 1,1 = $E_{1,1} = (22 - 21.32)^2/(21.32)$
- Celda 1,2 = $E_{1,2} = (15 - 15.67)^2/(15.67)$
- Celda 2,1 = $E_{2,1} = (12 - 12.67)^2/(12.67)$
- Celda 2,2 = $E_{2,2} = (10 - 9.32)^2/(9.32)$

Se alcanza un valor $\chi^2 = 0.136$ como resultado de la suma de la celda 1,1 (0.022) + celda 1,2 (0.029) + celda 2,1 (0.036) + celda 2,2 (0.049).

El valor crítico se obtiene de acuerdo con las tablas de distribución χ^2 (1 GL, unidireccional o una cola, $\alpha = 0.05$ es de 3.84). El resultado fue χ^2 calculada = 0.136 < χ^2 crítica 3.84, lo que no permite rechazar la H0. Por tanto, la interpretación es que no existe relación entre el reflujo gastroesofágico diurno/nocturno y la acidez esofágica de 24 horas en pacientes que han tenido pirosis cuatro veces a la semana en los últimos seis meses (χ^2 calculada = 0.136 < χ^2 crítica 3.84, 1 GL, $\alpha = 0.05$).

Todas las frecuencias esperadas deben ser mayores o iguales a 2, y no más de 20% de las celdas deben tener frecuencias esperadas menores de 5. En caso de no cumplirse esto, hay que realizar la prueba exacta de Fisher o la corrección de Yates para reducir la discrepancia entre la estadística ji al cuadrado calculada y la distribución teórica.

Una vez que se entiende el porqué de este método, pueden usarse herramientas de búsqueda electrónica gratuitas de *software* estadístico. Se solicita chi square test calculador, y se ingresan en la tabla de 2 x 2 las observaciones numéricas, el nivel de significancia y los grados de libertad, y automáticamente el programa calcula el valor de la ji al cuadrado y la significancia estadística.

En resumen:

1. Se desarrolla la hipótesis de relación entre variables.
2. Se formula la hipótesis estadística.

- Nula (H0): no existe relación entre el reflujo gastroesofágico nocturno y la acidez esofágica de 24 horas en pacientes que han tenido pirosis cuatro veces a la semana en los últimos seis meses, $A = B$.
- Alternativa (H1): sí existe relación entre el reflujo gastroesofágico nocturno y la acidez esofágica de 24 horas en pacientes que han tenido pirosis cuatro veces a la semana en los últimos seis meses, $A \neq B$.

3. Reglas de decisiones: distribución y estadística ji al cuadrado. Por lo general, se utiliza un nivel de significancia de 0.05, un grado de libertad, y se selecciona la zona crítica ($x_{2,crit} = 3.84$ para rechazar la hipótesis nula. Para 0.05 y 1 GL, la $x_{2,crit}$ es 3.84; para 0.10 y 1 GL es 2.7, y para 0.05 y 2 GL es 5.99.

Grados de libertad (GL) = (número de columnas - 1) (núm. de renglones - 1).

Cuando el valor obtenido de la ji al cuadrado es mayor a 3.84 para una tabla de contingencia de 2×2 se rechaza la hipótesis nula.

PARA VARIABLES NUMÉRICAS: t DE STUDENT¹⁵⁻²⁰

La distribución de probabilidad t de Student determina las diferencias entre dos medias muestrales en una población con distribución normal cuando el tamaño de la muestra es pequeño; además, permite construir intervalos de confianza (IC) para las diferencias de las medias cuando se desconoce la desviación típica de una población.

Si se desea conocer las diferencias de las medias o promedios entre dos grupos con observaciones numéricas, se usa la distribución t para contrastar la hipótesis de diferencias. La media de un grupo de estudio es un valor hipotético de la media poblacional.

W. Gosset publicó sus resultados en 1908 bajo el seudónimo de Student: "Cuando una observación procede de una distribución normal, las medias se distribuyen de manera normal, siempre y cuando se conozca su desviación estándar".

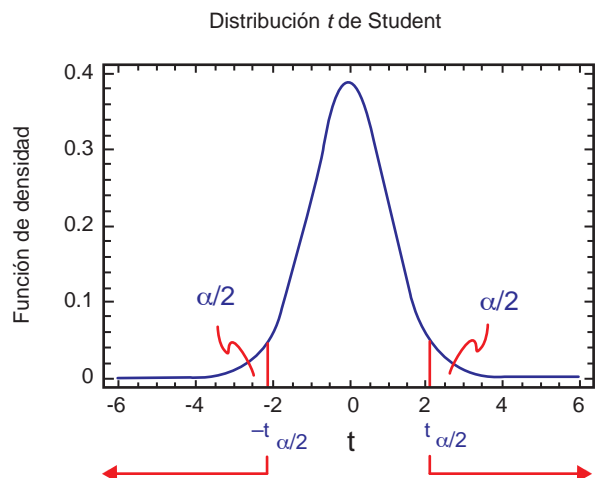
Puede utilizarse esta estadística cuando se desea saber: 1) si la media de un número de observaciones es igual a un valor estándar normal (estimador), 2) si la media cambia después de un tratamiento o intervención, y 3) si las medias de dos grupos independientes son iguales. El diseño para grupos independientes involucra una selección aleatoria de la muestra y asignación al azar de participantes del estudio a uno de los grupos.

Las características que definen una distribución muestral son la estadística de interés (media, σ), la selección aleatoria, el tamaño y la población de donde se extrajo la muestra.

Ésta es similar a la distribución Z , sobre todo si el tamaño de la muestra es mayor a 30. Es necesario considerar tres factores: 1) qué tan grande o tan pequeña es la diferencia entre las muestras, 2) la variabilidad de una muestra, y 3) el tamaño de la misma. Puede usarse cuando las observaciones tienen una distribución normal o gaussiana (campana, unimodal y simétrica con media igual a 0), lo que se determina al graficar (gráfica de probabilidad de Lillieford) las observaciones mediante un histograma que incluye la campana.

Es una familia de distribución de probabilidad caracterizada por grados de libertad (GL), que corresponde a $n - 1$.

La distribución t , como ya se dijo, es similar a la distribución Z o normal estándar ($\mu = 0$, $\sigma = 1$). Como rara vez se conoce la verdadera desviación estándar (σ), debido a que en la investigación médica generalmente se prefieren muestras y no poblaciones, se puede usar la distribución t . Si se conoce σ se puede utilizar la distribución Z , ya que sin importar la n , la distribución muestral de la media de una población con distribución normal tiene una forma gaussiana.



La distribución t incluye los tres factores antes mencionados. Una de las peculiaridades es que la variancia de la distribución t es mayor que en la Z ; por ello, para calcularla no se utiliza σ , sino el error estándar de la media

(EEM) que resulta de dividir σ entre \sqrt{n} . La DEM combina el factor 2, variabilidad (σ), y el 3, tamaño de la muestra (n). La DEM tiene menor variación que σ . La variabilidad es escasa cuando los sujetos son más homogéneos o semejantes, y el método de medición es más preciso.

Cuando en la distribución t n es igual o menor a 30, la razón crítica no tiene distribución normal. Si se aumenta el tamaño de la muestra, se incrementan los grados de libertad, y la forma de la curva de t se aproxima a Z ; es decir, que la σ depende de n y GL .

La fórmula para calcular la distribución Z es:

$$\frac{\bar{X}_1 - \bar{X}_2}{\sigma}$$

La fórmula para calcular la distribución t es:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{EEM (\sigma/\sqrt{n})}$$

Pasos:

1. *Hipótesis estadística:* $H_0 = \bar{X}_1 - \bar{X}_2 = 0$, $H_1 = \bar{X}_1 - \bar{X}_2 \neq 0$

2. *Prueba estadística:* t o Z

3. *Nivel de significancia:* generalmente se utiliza un nivel de 0.05.

Los intervalos de confianza (IC) proporcionan más información, ya que el valor de p se basa en la contrastación de las hipótesis que asume que la H_0 es verdadera, la probabilidad de obtener un resultado extremo al que se observó, e indica si es consecuencia de la casualidad. Algunos de los inconvenientes son el error tipo I (aleatorio), o *alfa*, si se concluye que existe diferencia entre las medias cuando no la hay; y el error tipo II (selección), o *beta*, cuando se determina que no hay diferencias si las hay.

En la literatura se ha destacado el uso dicotómico del valor de p (>0.05 , o no significativo) cuando se contrasta una hipótesis. Ésta es una desventaja, ya que el resultado se simplifica en significativo o no significativo, lo que es un estimador impreciso de la realidad, dada la variabilidad en el muestreo. Es erróneo suponer que el objetivo de presentar resultados es alcanzar la significancia estadística, puesto que ésta no brinda mayor información, y tampoco aporta datos sobre la magnitud de la variación entre los grupos de estudio. Los intervalos de confianza no revelan un estimador puntual, sino un rango de valores (estimadores de intervalo) que son considerados plausibles y representan una aproximación de la muestra de la

población. Son importantes para determinar la magnitud del efecto en la variable de interés, y ofrecen un rango de valor alrededor de un promedio, en donde caen las observaciones. La amplitud del intervalo de confianza depende del error estándar de la media.

Un intervalo de confianza de 95% da la oportunidad de que los límites superior e inferior señalados incluyan las diferencias de medias de la población en estudio. Significa que en 95% de las veces, los intervalos de confianza incluirán las diferencias de promedio cuando se repite una serie de estudios idénticos de distintas muestras de una población, debido a que la n y el tamaño del error estándar de la media (EEM) influyen en ellos. Al aumentar n disminuye la amplitud de los intervalos.

La potencia de la prueba ($1 - \beta$) depende del tamaño de la muestra para detectar diferencias; generalmente se utiliza 0.9 o 0.8.

4. *Valor crítico para significancia estadística:* depende del nivel de significancia estadística y los grados de libertad.

5. *Cálculos:* puede construirse un intervalo de confianza del 95% para la proporción poblacional p .

$$\hat{p} - 1.96 \times \sqrt{\hat{p} \times (1-\hat{p})/n} \leq \hat{p} \leq \hat{p} + 1.96 \times \sqrt{\hat{p} \times (1-\hat{p})/n}$$

Ejemplo:

En un estudio de prevalencia de factores de riesgo de enfermedad por reflujo gastroesofágico que incluyó 412 hombres mayores de 45 años de edad, se encontró que 10.5% tenía esófago de Barrett. Un intervalo de confianza de 95% para la proporción de hombres con Barrett está dado por:

$$0.105 - 1.96 \times \sqrt{0.105 \times (1 - 0.105)/412} \leq p \leq 0.105 + 1.96 \times \sqrt{0.105 \times (1 - 0.105)/412}$$

Luego, la proporción de Barrett varía entre 0.075 y 0.135, con un intervalo de confianza de 95%.

6. *Conclusiones:* con la prueba de contrastación de hipótesis se reporta el nivel de significancia estadística de acuerdo con el valor de p .

COMPARACIÓN ENTRE TRES O MÁS MEDIAS: ANÁLISIS DE VARIANCIA O ANOVA²¹⁻²²

Se recomienda usar el análisis de variancia, ANOVA, el cual permite identificar las diferencias entre los grupos. Así, se tendrá una observación dentro de cada grupo, que se representa con la letra X , y otra, que es la media de todo el

grupo de estudio. Por tanto, si se tienen tres grupos, puede representarse: $(X - X.) = (X - X_j) + (X_j - X.)$. Esto quiere decir que se pueden ver las disimilitudes entre la media del grupo de este sujeto y la diferencia entre la media del grupo y la media global, es decir, la gran media.

En el siguiente ejemplo hipotético, se observa la edad de los pacientes con esófago en cascanueces, los cuales se estratificaron en tres grupos de acuerdo con su exposición esofágica al ácido (grupo 1), esófago hipersensible (grupo 2) y pirosis funcional (grupo 3), y se calcularon la media (\bar{X}) y la desviación estándar (DE).

	Grupo 1 n = 17	Grupo 2 n = 17	Grupo 3 n = 15
1	49	50	22
2	89	45	33
3	14	60	45
4	28	42	48
5	46	44	45
6	29	70	75
7	46	80	33
8	35	55	35
9	43	50	45
10	36	35	56
11	71	110	76
12	48	53	34
13	82	40	45
14	63	80	32
15	63	51	35
16	72	49	
17	42	55	
Promedio	50.35	57	43.93
DE	20.306	18.671	15.308

El sujeto 1 del grupo 1 tiene 49 años. La media de edad del grupo 1 es de 50.3 años, la del grupo 2 de 57 y la del grupo 3 de 43.3. La gran media es de 50.6 años.

Si se compara la media del sujeto 1 con la de su grupo, la diferencia es de - 1.3.

La estimación de la variancia de las medias es:

$$S^2 = \text{suma } (X - \bar{X})^2/n - 1 \text{ y suma } n_j (X - \bar{X}_j)^2/j - 1$$

En ANOVA se considera la variación de todas las observaciones, y se divide en la variación entre cada sujeto y la media de su grupo, y en la variación de la media de cada grupo y la media global.

$$(49 - 50.3) + (50.3 - 50.6) = - 1.0$$

Paso 1. No hay diferencia entre los grupos $\mu_1 = \mu_2 = \mu_3$ versus $\mu_1 = \mu_2 = \mu_3$.

Paso 2. Prueba F para dos variancias. La H0 indica que las variancias son iguales.

Paso 3. Nivel de significancia estadística, por ejemplo 0.05.

Paso 4. De acuerdo con las tablas asignadas para distribución: F con $j - 1 = 2$ grados de libertad (tres grupos menos uno) en el numerador y suma $(n_j - 1)$; es decir, $49 - 3 = 46$ GL para alfa 0.05. El punto crítico en la tabla es de 3.23.

Paso 5. Se calcula la gran media:

$$\bar{X} = (17)(50.3) + (17)(57) + (15)(43.9)/17 + 17 + 15 = 2482.6/49 = 50.6$$

Luego, el numerador del cuadrado de la media entre grupos es MS_A :

$$(17)(50.3 - 50.6)^2 + (17)(57 - 50.6)^2 + (15)(43.9 - 50.6)^2 =$$

$$1.53 + 696.32 + 673.35 = MS_A = 1371.2$$

Las variancias se usan para calcular la estimación mezclada del cuadrado del error estándar (MS_E):

$$\frac{16(20.3)^2 + 16(18.6)^2 + 14(15.3)^2}{16 + 16 + 14} =$$

$$\frac{6593.44 + 5535.36 + 3277.26}{46} =$$

$$\frac{15406.06}{46}$$

$$= 334.91$$

$$F = MS_A/MS_E = 1371.2/334.91 = 4.04$$

$$F = 4.04$$

El valor observado de la razón F fue 4.04, que es mayor que 3.23, valor crítico de la tabla; por tanto, se rechaza la hipótesis nula de medias iguales. Esto indica que hay diferencias entre los grupos de pacientes con esófago en cascanueces y reflujo gastroesofágico.

CONCLUSIONES

La construcción de una hipótesis de trabajo esconde las variables de medición y permite que el investigador seleccione el diseño metodológico adecuado y los métodos estadísticos pertinentes para el objeto de estudio. El conocimiento de los posibles sesgos y confusores, y la validez interna y externa del estudio realzan la importancia de la investigación.

REFERENCIAS

1. Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring III JR. Case control studies. In: Greenberg RS, Daniels SR, editors. *Medical Epidemiology*. 4th ed. New York: Lange Medical Books-McGraw-Hill, 2005;p:147-161.
2. Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring III JR. Clinical Trials. In: Greenberg RS, Daniels SR, editors. *Medical Epidemiology*. 4th ed. New York: Lange Medical Books-McGraw-Hill, 2005;p:107-131.
3. Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring III JR. Cohort Studies. In: Greenberg RS, Daniels SR, editors. *Medical Epidemiology*. 4th ed. New York: Lange Medical Books-McGraw-Hill, 2005;p:132-146.
4. Hansson L, Engstrand L, Nyren O. *Hp* infection: independent risk indicator of gastric adenocarcinoma. *Gastroenterol* 1993;105:1098-1103.
5. Nomura A. *Hp* infection and gastric carcinoma among Japanese Americans in Hawaii. *N Engl J Med* 1991;325:1132-1136.
6. Forman D, Newell DG, Fullerton F. Association between infection with *Hp* and risk of gastric cancer: evidence from a prospective investigation. *BMJ* 1991;302:1302-1305.
7. Parsonnet J, Friedman G, Vandersteen D. *Hp* infection and the risk of gastric carcinoma. *N Engl J Med* 1991;325:1127-1131.
8. Kokkola A, Valle J, Haapiainen R. *Hp* infection in young patients with gastric carcinoma. *Scand J Gastroenterol* 1996;31:643-647.
9. Siström CL, Garvan CW. Proportions, odds, and risk. *Radiology* 2004;230(1):12-9. doi:10.1148/radiol.2301031028
10. Deeks J. When can odds ratios mislead? Odds ratios should be used only in case-control studies and logistic regression analyses [letter]. *BMJ* 1998;317(7166):1155-6;discussion 1156-7.
11. Matthews DE, Farewell VT. Approximate significance test for contingency tables. The χ^2 test for 2 x 2. Using and understanding medical statistics. 4th ed. Bern: Karger AG, 2007;p:28-34,42-45.
12. Dawson-Saunders B, Trapp RG. *Bioestadística Médica*. México: El Manual Moderno, 1993;p:165-186.
13. Dawson B, Trapp RG. *Basic and clinical biostatistics*. 4th ed. New York: Lange Medical Books-McGraw-Hill, 2004;p:149-152.
14. Essex-Sorlie D. *Medical biostatistics and epidemiology*. Norwalk: Appleton and Lange, 1995;p:227-238.
15. Matthews DE, Farewell VT. *Using and understanding medical statistics*. 4th ed. Bern: Karger AG, 2007.
16. Essex-Sorlie D. *Medical biostatistics and epidemiology*. Norwalk: Appleton and Lange, 1995;p:166-190.
17. Dawson B, Trapp RG. *Basic and clinical biostatistics*. 4th ed. New York: Lange Medical Books-McGraw-Hill, 2004;p:134-161.
18. Dawson-Saunders B, Trapp RG. *Bioestadística Médica*. México: El Manual Moderno, 1993;p:115-145.
19. Altman DG. Means and their differences. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with confidence*. 2nd ed. London: BMJ, 2000;p:28-36.
20. Gardner MJ, Altman DG. Confidence intervals rather P values. In: Altman DG, Machin D, Bryant TN, Gardner MJ, editors. *Statistics with confidence*. 2nd ed. London: BMJ, 2000;p:15-27.
21. Saunders B, Trapp RG. *Bioestadística Médica*. México: El Manual Moderno, 1993;p:143-163.
22. Dawson B, Trapp RG. *Basic and clinical biostatistics*. 4th ed. New York: Lange Medical Books-McGraw-Hill, 2004;p:163-189.