

OR, Paradoja de Simpsons, Metanálisis

1. Defina los siguientes términos:

a) Inferencia inductiva

Respuesta

La inferencia inductiva consiste en desde lo particular establecer una proyección a lo general. Se comienza estudiando un caso o casos particulares y luego, bajo metodología, generalizar estos hallazgos a un nivel macro. En estadística, al realizar un muestreo de la población (casos particulares) se pretende inferir sobre la población (generalización), a esto podemos llamarlo *inferencia inductiva*.

b) Inferencia deductiva

Respuesta

La inferencia deductiva comienza sujeta a una teoría, y basada en ésta, se espera un comportamiento, se predicen consecuencias o características de un caso particular. Es decir, lo ya generalizado se pretende aplicar a uno o más casos particulares. En Zoología, al establecer que *todos los reptiles son de sangre fría* (generalización) es posible inferir que las tortugas (casos particulares) al ser reptiles, son de sangre fría, a esto podemos llamarlo *inferencia deductiva*.

c) Hipótesis nula

Respuesta

En el marco de una investigación o análisis, generalmente se suelen realizar pruebas de hipótesis, para recopilar evidencia y rechazar o no rechazar la hipótesis nula, la que consiste en ser una afirmación sobre la población de estudio que se pretende rechazar. La hipótesis nula se basa generalmente en algún conocimiento previo, por ejemplo:

Estudios anteriores no han establecido diferencias entre la efectividad de doxiciclina y azitromicina en el tratamiento de infección urogenital por Chlamydia trachomatis.

La hipótesis nula de interés en el contexto anterior, podría establecerse como sigue:

H_{nul} : La efectividad de doxiciclina es igual a la efectividad de azitromicina

Evidentemente es necesario especificar métrica (s) para medir la *efectividad* y así poder recopilar evidencia y obtener el resultado de la prueba de hipótesis en cuestión.

d) Falsabilidad (o refutabilidad) de una hipótesis

Respuesta

La falsabilidad de una hipótesis consiste en su cualidad propia y natural de ser puesta en duda bajo evidencia. No es muy útil estudiar hipótesis que no son falseables (que no

permita la duda), pero analizar hipótesis potencialmente falsificables, puede ser muy útil.

2. Dé al menos dos ejemplos de:

a) Estudios observacionales

Respuesta

1) Con el fin de estudiar la proporción de mujeres que realizan la PSU, del total de personas que rindieron la PSU se observa la proporción de mujeres que rindieron la prueba.

2) Con el fin de estudiar alguna relación entre fumar y el cancer de pulmón, se obtuvo una lista de pacientes con cáncer de pulmón y se calcula cuántos de estos fuman y cuántos no fuman.

b) Estudios experimentales

Respuesta

1) Con el fin de analizar el efecto de distintos abonos en el crecimiento de hortensias, se tienen 3 grupos de hortensias y a cada grupo se le aplica abono de tipo 1, 2 y 3. Se le mide el crecimiento a cada grupo de hortensias.

2) Se cree que la cantidad de correos que se le envían a los clientes puede afectar en su comportamiento financiero. Se selecciona una muestra de clientes y a un grupo se le envían diariamente 2 correos con información de los productos del banco y 1 extra, mientras que a los otros clientes se les envía sólo 2. Se pretende analizar si existirá algún cambio en el comportamiento de los clientes que recibían 3 correos diarios.

c) Estudios transversales (Cross-sectional)

Respuesta

1) Se pretende obtener la prevalencia de VIH en mujeres embarazadas. Se toma una muestra de las mujeres embarazadas inscritas en el curso Maternidad prenatal y se determina cuántas de ellas poseen VIH.

2) Se pretende obtener una estimación de la población que actualmente utiliza mascarillas dada la pandemia del Covid-19 con el fin de determinar la prevalencia del uso de métodos de barrera en la población y así evitar contagios. Se tomaron imágenes satelitales de lugares con alta concurrencia de personas y analizando dichas imágenes se obtiene una estimación de cuántas personas en éstas imágenes se encontraban usando mascarilla y cuántas no.

d) Estudios Caso-Control

Respuesta

1) Se pretende analizar si la ingesta de fitoestrógenos posee algún efecto en el cáncer de mama. Se seleccionan aleatoriamente mujeres con cáncer de mama recientemente diagnosticado y se separa en dos grupos: caso-control. Sólo al grupo control se le suministran fitoestrógenos. A ambos grupos se les miden las mismas características.

2) Con el fin de estudiar cuáles son los efectos de una dieta vegana en gatos, se seleccionan gatos que poseen una dieta de tipo convencional y se separa en dos grupos: caso-control. Sólo al grupo control se le modifica la dieta convencional a una dieta de tipo vegana nutritiva. A ambos grupos se les miden las mismas características.

3. Explique, con sus palabras, la diferencia entre un estudio prospectivo y uno retrospectivo.

Respuesta

La principal diferencia entre estudio prospectivo y retrospectivo radica en base a qué o cómo fueron seleccionados los sujetos del estudio, si en base a variables de entrada (incidentes) o de salida (resultado). Si la selección/separación de los sujetos se basa en una variable que pudiera sugerir o determinar un resultado o padecimiento, dicha variable es de entrada y el estudio es de tipo prospectivo (por ejemplo, seleccionar y separar personas dependiendo de si fuman o no fuman para estudiar el desarrollo de cáncer de pulmón, la variable fumar es de entrada pues ésta podría aparentemente causar o incidir en el cáncer de pulmón que sería variable de salida). Si la selección/separación de los sujetos se basa en si posee o no un evento o padecimiento (variable de salida) y se estudian otras variables que inciden en esta enfermedad (variables de entrada) el estudio es retrospectivo (por ejemplo estudiar pacientes que poseen depresión (variable de salida) y suministrarle a un grupo un placebo y a otro grupo un medicamento (variables de entrada) para analizar la efectividad de dicho medicamento EN la depresión).

4. Dé un ejemplo de un estudio prospectivo. ¿Cómo realizaría el estudio desde el punto de vista retrospectivo?

Respuesta

Prospectivo: Se interesa realizar un análisis de posibles variables que afecten en la insatisfacción de los clientes en su experiencia de cliente. Se separan los casos dependiendo de si adquirió su producto por Delivery o en modo presencial, pues se cree que la modalidad de entrega del producto podría afectar en el nivel de insatisfacción del cliente. Y se espera hasta que el cliente realice la evaluación de su experiencia de cliente.

Retrospectivo: Se separan los casos en clientes satisfechos e insatisfechos en su experiencia de cliente, y en cada grupo se le pregunta si el cliente adquirió su producto vía Delivery o presencial.

5. Dé un ejemplo de estudio de cohorte ¿Cómo realizaría el estudio desde el punto de vista de estudio caso-control?

Respuesta

Cohorte: Se quiere estudiar si jugar videojuegos podría producir trastornos de ansiedad. Se tienen dos grupos de personas, la cohorte que incluye personas que juegan videojuegos y el grupo no expuesto, donde no juegan videojuegos. Posteriormente se mide el desarrollo

de trastornos de ansiedad.

Caso-control: Se pretende analizar si jugar videojuegos se relacionaría con un agravamiento en los trastornos de ansiedad. Se tiene una muestra de pacientes con trastornos de ansiedad (con características similares) y se separa en dos grupos, uno de los grupos se somete a jugar videojuegos. Se analiza si existe algún cambio o efecto en el trastorno de ansiedad de los pacientes.

6. Explique por qué el muestreo aleatorio asegura estimadores de parámetros poblacionales insesgados, pero que esto no garantiza que una muestra particular sea *balanceada*.

Respuesta

Para explicarlo, propondré un ejemplo. Supongamos que la población de estudio es la siguiente:

2 8 4 7 1 3 9 5 3 4 2 5 3 2 8 9 2

A través del muestreo, puede ser posible extraer algunas muestras de la población (utilicemos MAS como tamaño muestral $n=4$):

Muestra A : 8 4 3 7 $\bar{x}_{\text{Muestra A}} = 5.5$

Muestra B : 2 5 1 7 $\bar{x}_{\text{Muestra B}} = 3.75$

Muestra C : 2 2 2 2 $\bar{x}_{\text{Muestra C}} = 2$

Muestra D : 4 7 5 2 $\bar{x}_{\text{Muestra D}} = 4.5$

Suponga que sólo se pudo extraer la **Muestra C** obtenida con MAS ($n=4$). Es posible notar que esta muestra no está balanceada, y de hecho de las 4 muestras es aquella que posee mayor distancia respecto al parámetro real μ , a esto nos referimos con que un MAS no asegura una muestra balanceada, podemos haber obtenido justo un muestreo que no representa un comportamiento similar a la población. Adicionalmente, una buena opción al tener distintas muestras, sería utilizar un promedio ponderado.

7. Explique qué es un meta-análisis y dé ejemplos en que la utilización de esta técnica podría ser de utilidad.

Respuesta

El Metanálisis es de utilidad al momento de poseer una colección de n estudios del mismo tópico y se pretende unificar la información de estos n estudios para obtener un resultado. Muy útil cuando los estudios apuntan a resultados contrapuestos, metanálisis reúne evidencia en pos de una hipótesis. Un metanálisis realizado correctamente puede entregar resultados más sólidos que los estudios de manera singleton.

Ejemplo 1: En el ámbito científico existe una disyuntiva sobre si la Actividad física durante el embarazo puede prevenir el desarrollo de la Depresión Post Parto. Se determinaron 21 estudios que testearon distintos tipos de Actividad física durante el embarazo y su efecto en la depresión o síntomas de la depresión post parto. Los estudios obtienen distintas conclusiones por lo que, con el fin de obtener una conclusión sólida, se realiza un análisis sistemático: Metanálisis.

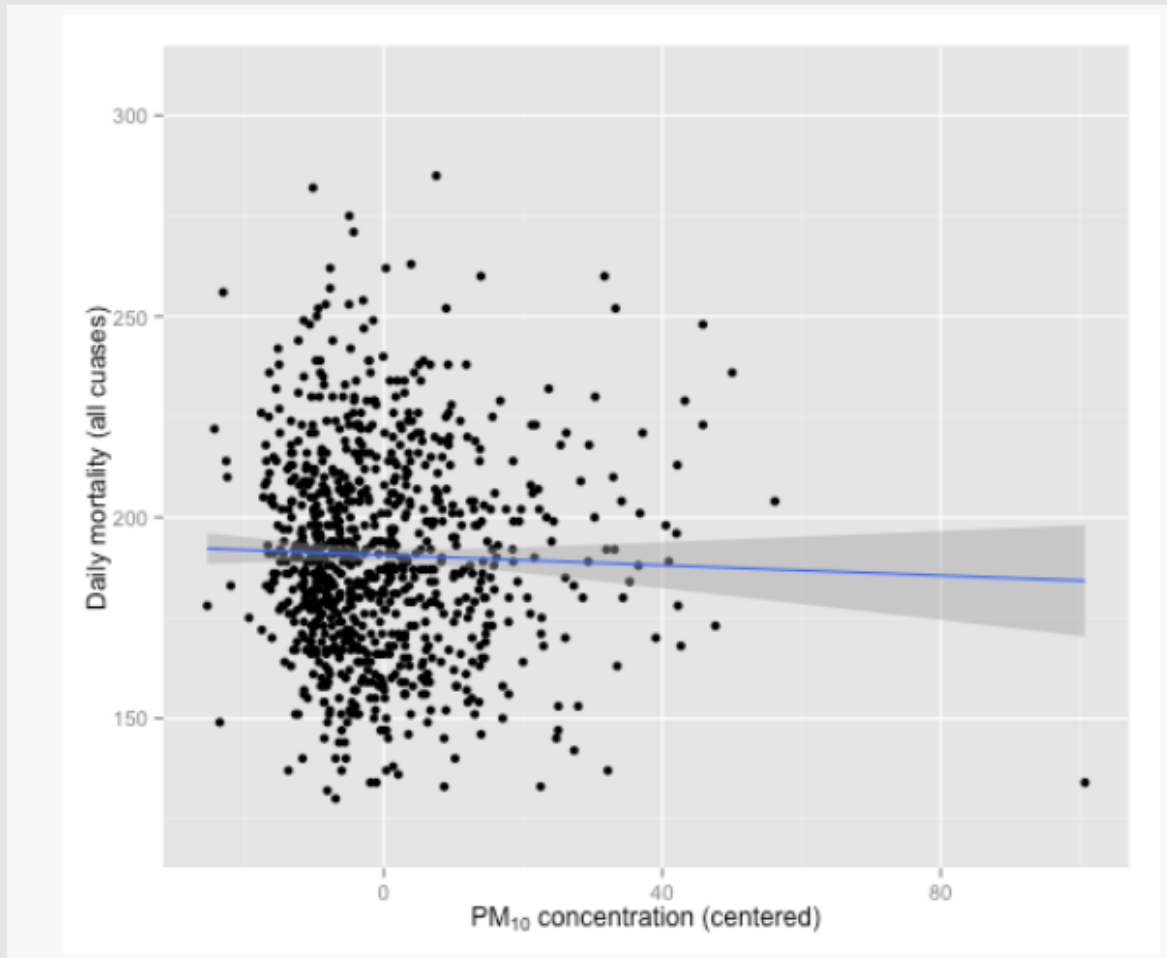
Ejemplo 2: Se realizaron distintos estudios de tipo caso-control para estudiar el efecto

de la Penicilamina en el tratamiento de cirrosis. Algunos estudios son concluyentes y algunos no. Se pretende obtener un resultado global incorporando la información de los estudios.

8. Dé un ejemplo, en donde pueda haber efecto de una variable confundente (explícite cuál variable sería la dependiente, independiente y confundente) ¿cómo podría definir probabilísticamente una variable confundente?

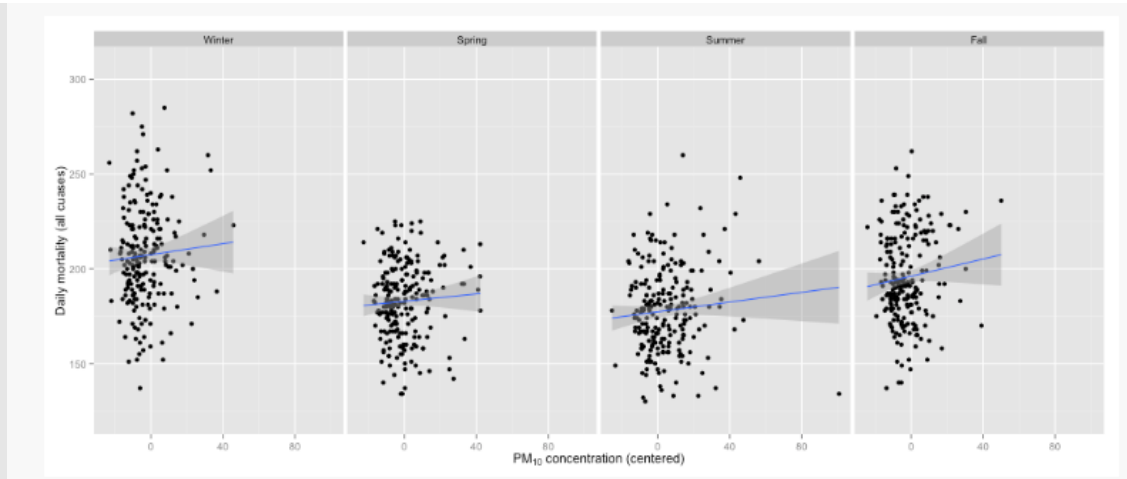
Respuesta

Se pretende estudiar si las cantidades de Partículas en suspensión (PM_{10}) en el aire pudiesen significar un aumento en la variable mortalidad diaria. Se obtuvo el siguiente gráfico:



Se concluyó que al momento de realizar una regresión lineal, se obtiene que la pendiente es negativa, es decir, una relación inversa (A mayor concentración de partículas en suspensión en el aire, menor mortalidad).

Pero, si incorporamos la información de la variable estación, y realizamos el mismo gráfico pero por estación del año, observamos lo siguiente:



En todos los casos, se tiene una pendiente positiva. Si al añadir una variable se obtienen conclusiones contrapuestas, estamos en presencia de la **Paradoja de Simpsons**.

Variable dependiente: Mortalidad diaria

Variable independiente: Concentración de PM₁₀ Partículas en suspensión

Variable confundente: Estación del año (se encuentra anidada en la posible variable causal)

En este caso, sea E la variable dependiente (Mortalidad diaria), C la variable independiente (Concentración de PM₁₀ y F la variable confundente (Estación del año), matemáticamente la paradoja de Simpsons puede entenderse matemáticamente como:

$$P(E|C) > P(E|C^c)$$

$$P(E|C, F) < P(E|C^c, F)$$

$$P(E|C, F^c) < P(E|C^c, F^c)$$

El evento dependiente (E) sería más probable al condicionar por el evento independiente (C) en una muestra dada. Pero al mismo tiempo, si se condiciona por un nuevo evento (F), se observa una disminución en la probabilidad del evento dependiente (E) en cada subpoblación (sujeta a F) de la muestra.

9. En el contexto de variables confundentes ¿Qué conclusiones obtiene si $OR > 1$ antes de estratificar y $OR_{MH} < 1$ post-estratificación?

Respuesta

Este resultado puede entenderse como que al considerar sólo la variable independiente de interés, se concluye que la variable dependiente aumentaría las chances ($OR > 1$) del padecimiento en el grupo positivo respecto al grupo negativo. Sin embargo, al incorporar la información de una nueva característica (asociada con las variables anteriores) y realizar la estratificación (con el método de Mantel-Haenszel) se obtiene que la variable independiente no sugiere realmente un aumento en las chances del padecimiento en el grupo positivo respecto al grupo negativo. Esto podría entenderse como un claro ejemplo de la Paradoja de Simpsons. Al incorporar la información de la nueva variable (confundente), y estratificar por ésta, se barre el efecto que ésta variable sugería implícitamente en el ejercicio, y que antes (al no considerarla, se asociaba a un efecto meramente de la variable independiente de estudio).

10. Data are tabulated separately for smokers and nonsmokers in Table 1.1. The exposure under investigation, “shipbuilding,” refers to employment in shipyards during World War II. By

using a separate tabulation, with the first half of the table for nonsmokers and the second half for smokers, we treat smoking as a potential confounder. A confounder is a factor, an exposure by itself, not under investigation but related to the disease (in this case, lung cancer) and the exposure (shipbuilding); previous studies have linked smoking to lung cancer, and construction workers are more likely to be smokers. The term exposure is used here to emphasize that employment in shipyards is a suspected risk factor; however, the term is also used in studies where the factor under investigation has beneficial effects.

TABLE 1.1

Smoking	Shipbuilding	Cases	Controls
No	Yes	11	35
	No	50	203
Yes	Yes	84	45
	No	313	270

- a) Calculate the odds ratio associated with employment in shipyards for nonsmokers.

Respuesta

El cálculo de OR es el siguiente:

$$OR = \frac{\frac{11}{35}}{\frac{50}{203}} = 1.276$$

El cual se interpreta de modo que (para el grupo de no fumadores) las chances de presentar cancer de pulmón al trabajar como astillero son 1.276 veces las chances de presentar cáncer de pulmón al no trabajar como astillero, es decir, se observaría una mayor chance de presentar cáncer de pulmón al ser astillero que al no serlo (para el grupo de no fumadores).

- b) Calculate the same odds ratio for smokers.

Respuesta

El cálculo de OR es el siguiente:

$$OR = \frac{\frac{84}{45}}{\frac{313}{270}} = 1.61$$

El cual se interpreta de modo que (para el grupo de fumadores) las chances de presentar cancer de pulmón al trabajar como astillero son 1.61 veces las chances de presentar cáncer de pulmón al no trabajar como astillero, es decir, se observaría una mayor chance de presentar cáncer de pulmón al ser astillero que al no serlo (para el grupo de fumadores). Notar que respecto al grupo de no fumadores la razón de chances es menor que al grupo de fumadores, lo cual tiene sentido pues existe un mayor riesgo o efecto implícito explicado al fumar que al no fumar.

- d) Assuming that the odds ratios for the two groups, nonsmokers and smokers, are equal (in other words, smoking is not an effect modifier), calculate the Mantel–Haenszel estimate of this common odds ratio.

Respuesta

El cálculo de $OR_{\text{Mantel-Haenszel}}$ es el siguiente:

$$OR_{\text{Mantel-Haenszel}} = \frac{\frac{11 \cdot 203}{11+35+50+203} + \frac{84 \cdot 270}{84+45+313+270}}{\frac{35 \cdot 50}{11+35+50+203} + \frac{45 \cdot 313}{84+45+313+270}} = 1.533$$

La razón de chances combinada de Mantel-Haenszel es de 1.533, con el supuesto de que fumar no sería una variable confundente, el cual pudiera interpretarse de modo que, considerando la información de si fuma o no fuma, la razón de chances de presentar cáncer de pulmón para astilleros es 1.533 veces la razón de chances de presentar cáncer de pulmón para no astilleros.

11. ¿Qué significado tiene un efecto aleatorio en el meta análisis?

Respuesta

Considerar un modelo de efectos aleatorios en un meta análisis puede entenderse como incorporar diferencias entre los estudios más allá que el proceso de aleatoriedad de los individuos de cada estudio, es decir, al utilizar un modelo de efectos aleatorios además de asumir que existe una variabilidad intra estudio (dentro de cada estudio, el proceso de selección de las unidades de estudio, variabilidad muestral), incorpora una variabilidad entre estudios (diferencias contextuales, por ejemplo, los estudios se realizaron en distintos países o regiones, en distintos años, u otras diferencias que puedan significar una variabilidad natural entre cada estudio).

12. ¿Puede un estudio de meta análisis combinar distintos tipos de estudio? e.g., combinar estudios retrospectivo con prospectivo.

Respuesta

Sí es posible considerar distintos tipos de estudio. Sin embargo, es menester notar que dependiendo de la pregunta de investigación a resolver con el meta análisis, algunos estudios pueden más informativos (y apropiados) que otros. Por ejemplo, si se pretende analizar la efectividad de un medicamento respecto a un placebo, sería naturalmente más apropiado centrarse en estudios de tipo caso-control.