

# Métodos Cuantitativos

## Módulo 2: El método científico

### 2.01 Método científico: el ciclo empírico

El **ciclo empírico** contiene el proceso de formulación de hipótesis acerca de cómo funcionan las cosas y de contrastación de esas hipótesis con datos empíricos de una forma *sistemática y rigurosa*. Esto es lo que caracteriza el enfoque *hipotético-deductivo* de la ciencia. Familiarizarse con las diferentes fases de este ciclo realmente ayudará a mantener en mente una visión general, especialmente cuando lleguemos a cosas más específicas como el diseño experimental y el muestreo.

Así que comenzaremos con la **fase de observación**, por supuesto. Aquí es donde ocurre la magia. Es aquí donde la observación da origen a una idea para una nueva hipótesis de observación. Puede tratarse de un patrón interesante, un evento inesperado, cualquier cosa que encontremos interesante y que queramos explicar. Cómo hicimos esa observación es algo que realmente no interesa. Puede ser observación personal, una experiencia que alguien compartió con nosotros, incluso una observación imaginaria que tiene lugar apenas en tu cabeza.

Por supuesto las observaciones por lo general vienen de resultados previos de investigación, que son obtenidos sistemáticamente, pero en principio todo vale.

Entonces tomemos un ejemplo de una observación personal hecha por mí. Yo tengo una suegra horrible. He hablado con algunos amigos y amigas y ellos también se quejan de sus suegras. Así que esto me parece un patrón interesante. ¡Un patrón entre el tipo de persona y su habilidad para que otros gusten de ella! Entonces la fase de observación se trata observar una relación en una o más instancias específicas.

En la **fase de inducción** esta relación, observada en instancias específicas, se convierte en una regla general. Ese es el significado de la inducción: tomar un planteamiento que es verdad en casos específicos e inferir que ese planteamiento es verdad en todos los casos, siempre. Por ejemplo, de la observación de que mis amigos y yo tenemos suegras horribles yo puedo inducir la regla general de que todas las suegras son horribles.

Por supuesto esta regla, o **hipótesis**, no es necesariamente verdad, podría estar errada. De eso se trata el resto del ciclo: probar nuestra hipótesis. *En la fase de inducción el razonamiento inductivo es utilizado para transformar observaciones específicas en reglas generales o hipótesis.*

En la **fase de deducción** deducimos que la relación especificada en la regla general podría también contener instancias nuevas y específicas. De nuestra hipótesis deducimos una expectativa explícita o una predicción acerca de nuevas observaciones. Por ejemplo, si todas las suegras son de hecho horribles, entonces, cuando les pido a diez de mis colegas que califiquen a sus suegras como “agradables”, “neutrales” u “horribles”, todos ellos deben escoger la categoría “horrible”.

DOCUMENTO DE TRABAJO – NO REPRODUCIR

Ahora bien, para hacer semejante predicción, necesitamos definir un diseño de investigación. necesitamos hacer definiciones de conceptos relevantes, instrumentos de medición, procedimientos, la muestra de la cual vamos a recoger los datos, etc.

En la **fase de prueba** interpretamos los resultados en términos de nuestras hipótesis. Si la predicción fue confirmada esto apenas nos provee *apoyo provisional* para nuestra hipótesis. Esto no significa que hayamos probado definitivamente la hipótesis. Esto, debido a que siempre es posible que en el futuro encontremos a alguien que simplemente ame a su suegra.

Si la predicción fuese refutada, esto no quiere decir que debemos rechazar la hipótesis de una vez. En muchos casos hay explicaciones plausibles para nuestro fracaso al tratar de confirmar la hipótesis. Si esas explicaciones tienen que ver con el diseño de la investigación, la hipótesis se preserva y se investiga nuevamente, pero con un diseño mejor. En otros casos la hipótesis es ajustada, con base en los resultados. La hipótesis es rechazada y descartada sólo en casos muy raros. *En la fase de evaluación los resultados son interpretados en términos de la hipótesis, lo que significa que es provisionalmente demostrada, ajustada o rechazada.*

Las observaciones recogidas en la fase de prueba pueden servir como nuevas observaciones específicas en la fase de observación. Esta es la razón por la cual el proceso es descrito como un ciclo. Nuevos datos empíricos obtenidos en la fase de prueba pueden dar origen a nuevas consideraciones que conducen a nuevos procesos. Y de esto es que se trata la ciencia empírica: tratamos de perfeccionar las mejores hipótesis y construir nuestra comprensión del mundo a medida que describimos el ciclo una y otra vez.

## 2.02 El método científico (des)confirmación

Vamos a ver cómo podríamos interpretar resultados que confirman o refutan nuestras predicciones y si deberíamos confirmar o rechazar nuestras hipótesis de acuerdo con ellos. Vamos a considerar la hipótesis de que todas las suegras son horribles. Formulé esta hipótesis basándome en observaciones personales.

Para probar esta hipótesis, he creado un diseño de investigación. seleccioné medir la “horribilidad” utilizando una escala de opciones: “agradable”, “neutral” y “horrible”. También decidí recoger datos de diez colegas de mi departamento. Con el diseño de investigación en su sitio, formulé la siguiente predicción: si la hipótesis “todas las suegras son horribles” es verdadera, entonces todos mis 10 colegas deberían escoger “horrible” para describir a sus suegras.

Ok, veamos primero la **confirmación**. Supongamos que los diez colegas calificaron a sus suegras como horribles. La predicción es *confirmada*, ¡pero eso no significa que la hipótesis haya sido probada! Es fácilmente concebible que en el futuro se demuestre que estamos equivocados. Si repitiéramos el estudio podríamos encontrar a una persona que simplemente adore a su suegra.

El punto es que la confirmación nunca es concluyente. La única cosa que podemos decir es que nuestra hipótesis está provisionalmente soportada. Cuando mayor soporte resulte de diferentes estudios, mayor aceptación gana una hipótesis, pero nunca podremos probarla. Déjenme insistir

en esto: *ningún enunciado científico puede nunca ser probado de una vez y para siempre*. Lo máximo que podemos hacer es producir evidencia abrumadora para soportar una hipótesis.

Ahora veamos la **refutación**. Supongan que sólo ocho de diez colegas calificaron a su suegra como horrible y dos como neutral. Obviamente en este caso nuestra predicción resultó ser *falsa*. En términos lógicos, los hallazgos empíricos que *contradicen* la hipótesis deberían conducir a su rechazo. Si nuestra hipótesis establece que todos los cisnes son blancos y después encontramos cisnes negros en Australia, podremos de forma muy concluyente rechazar nuestra hipótesis.

En la práctica, sin embargo, especialmente en las ciencias sociales, hay con frecuencia *explicaciones alternativas plausibles* para nuestro fallo al confirmar la hipótesis. Estas explicaciones son de hecho tan fáciles de encontrar que *rara vez rechazamos* la hipótesis de inmediato. En muchos casos esas explicaciones tienen que ver con problemas metodológicos. El diseño de investigación o el instrumento de medición no eran apropiados, o tal vez variables relevantes del contexto no fueron controladas, etc.

Volviendo a nuestro ejemplo de la suegra: pude haber cometido un error procedimental mientras recogía las respuestas de mis dos colegas que calificaron a sus suegras como neutrales. Tal vez se me olvidó decirles que sus respuestas eran confidenciales, haciéndolos sentirse incómodos de escoger la categoría más negativa.

Si hay explicaciones metodológicas plausibles para nuestro fracaso en confirmar la hipótesis, *preservamos* la hipótesis y más bien escogemos rechazar *los presupuestos auxiliares implícitos* que tienen que ver con el diseño de investigación y la medición. Investigamos la hipótesis original otra vez, sólo que con un mejor diseño de investigación.

Algunas veces los resultados dan origen a la **modificación** de la hipótesis. Supongan que las ocho colegas que tuvieron suegras horribles eran todas mujeres y los otros dos eran hombres. Tal vez las suegras son en efecto horribles, ¡pero sólo con sus nueras!

Si alteramos la hipótesis ligeramente añadiendo cláusulas adicionales (“sólo aplica para nueras”), entonces en sentido estricto estaríamos rechazando la hipótesis original. Por supuesto la nueva hipótesis es esencialmente la misma hipótesis original, sólo que no tan general y por lo tanto no tan fuerte. El rechazo o el ajuste radical de una hipótesis es en realidad muy escaso en ciencias sociales. Los progresos son hechos a través de incrementos muy pequeños, y no de pasos de gigantes.

### 2.03 El método científico: criterios

Seguimos el ciclo empírico para dar lugar a hipótesis que probar contrastándolas con observaciones. Pero una vez los resultados están listos, una confirmación no significa que la hipótesis a ha sido probada, así como una refutación no significa que automáticamente la rechazemos.

Entonces, ¿cómo decidimos si un estudio es convincente? Bueno, hay dos principales criterios de evaluación: **confiabilidad** y **validez**.

La **confiabilidad** está estrechamente relacionada con la *replicabilidad* un estudio es replicable si investigadores independientes entre sí son capaces de repetirlo. Un resultado de investigación es *confiable* si realmente repetimos el estudio y encontramos resultados consistentes.

La **validez** es más complicada. Un estudio es *válido* si la conclusión sobre la relación formulada en la hipótesis entre las propiedades refleja la realidad con exactitud. En síntesis, un estudio es válido si la conclusión basada en los resultados es “verdad”.

Supongan que yo formulo una hipótesis: la soledad causa sentimientos de depresión. Deduzco que, si disminuyo la soledad en adultos mayores dándoles un gato del cual cuidar, sus sentimientos de depresión también van a decrecer. Ahora supongan que yo realizo este estudio en un hogar de retiro y encuentro que la depresión de hecho decrece después de que los residentes cuidan de un gato. ¿Es válido este estudio? ¿Los resultados apoyan la conclusión de que la soledad causa depresión?

Bueno, dado que es todavía una pregunta muy general, vamos a considerar tres tipos más específicos de validez: **constructo**, **validez interna** y **validez externa**.

La **validez del constructo** es un prerrequisito importante e la validez interna y externa. Un estudio tiene una alta validez de constructo si las propiedades (o “constructos”) que aparecen en la hipótesis son medidas y manipuladas con exactitud. En otras palabras, nuestros métodos tienen alta validez de constructo si en realidad *miden* y *manipulan* las propiedades que se espera que midan y manipulen.

Supongan que accidentalmente medí un constructo completamente diferente con, por ejemplo, mi cuestionario sobre la depresión. ¿Qué pasa si el cuestionario mide sentimientos de exclusión social en lugar de la depresión? O supongan que cuidar un gato no afectó la soledad en absoluto, sin que en lugar de eso incrementó los sentimientos de responsabilidad y auto-estima. ¿Qué tal que la soledad se haya mantenido igual?

Bueno, entonces los resultados sólo *parecen* apoyar la hipótesis de que la *soledad* causa la depresión, cuando en realidad hemos manipulado una causa diferente y medido un efecto distinto. Desarrollar medidas y métodos de manipulación precisos es uno de los grandes desafíos en ciencias sociales y ciencias comportamentales. Discutiré esto más en detalle cuando demos una mirada a la operacionalización.

Por ahora avancemos hacia la **validez interna**. La validez interna es relevante cuando nuestra hipótesis describa una *relación causal*. Un estudio es internamente válido si el efecto observado *realmente obedece* a la causa formulada en la hipótesis.

Asumamos por un segundo que nuestros métodos de medición y manipulación son válidos. ¿Podemos concluir que la depresión bajó porque los adultos mayores se sintieron menos solos? Bueno... tal vez otra cosa causó la disminución de la depresión. Por ejemplo, si el estudio comenzó en invierno y terminó en primavera, cuando tal vez el cambio de estación hizo bajar la depresión. ¿O tal vez no fue la compañía del gato sino el aumento de ejercicio físico por limpiar la caja de arena y el bowl de la comida?

Explicaciones alternativas como estas amenazan la validez interna. Si hay una explicación alternativa plausible, la validez interna es baja. Ahora, hay muchos tipos diferentes de amenazas a la validez interna, que discutiré más adelante.

Ok, ahora veamos la **validez externa**. Un estudio es externamente válido si la relación planteada en la hipótesis, apoyada por nuestros hallazgos, también se sostiene en otros contextos y en otros grupos. Dicho con otras palabras, si los resultados *se generalizan* a diferentes personas, grupos, ambientes y momentos.

Volvamos a nuestro ejemplo. ¿Cuidar de un gato también disminuye la depresión en adolescentes y personas de mediana edad? ¿El efecto será el mismo en hombres y en mujeres? ¿Qué hay de las personas de culturas diferentes? ¿Un perro será tan efectivo como un gato?

Por supuesto es difícil decirlo con base apenas en los resultados de adultos mayores y gatos. Si hubiésemos incluido gente joven, gente de diferentes contextos culturales y utilizado otros animales, podríamos tener más confianza en la validez externa del estudio. Volveré sobre la validez externa después, y sobre cómo puede verse amenazada, cuando hablemos sobre el muestreo.

Entonces para resumir: la **validez del constructo** se relaciona con si nuestros métodos realmente reflejan las propiedades que queremos *manipular y medir*. La **validez interna** se relaciona con si la causa formulada en la hipótesis es la causa real del efecto observado. La validez interna es amenazada por explicaciones alternativas. La **validez externa** o la posibilidad de generalizar los resultados tiene que ver con el hecho de que la relación causal formulada en la hipótesis se sostenga en otros contextos.

## 2.04 Método científico: causalidad

Las hipótesis más interesantes son aquellas que describen una **relación causal**. Si sabemos qué causa un efecto podemos *predecirlo, influenciarlo, comprenderlo* mejor.

¿Entonces cómo identificamos una relación causal? Bueno, David Hume, con un poco de ayuda de John Stuart Mill, por primera vez hicieron una lista de los criterios que todavía utilizamos hoy. Estos son los cuatro criterios esenciales:

- Número uno: la causa y el efecto están **conectados**. El primer criterio significa que tiene que haber una manera de rastrear el efecto de vuelta a la causa. Si un paciente no ha sido expuesto a un virus, entonces no podemos argumentar que el virus *causó* la muerte del paciente.
- Número dos: la causa **precede** al efecto. Espero que esto sea obvio.
- Número tres: la causa y el efecto **aparecen juntas consistentemente**. Esto significa que causa y efecto deberían ir juntas. Cuando la causa está presente, deberíamos ver el efecto, y si la causa no está presente, entonces el efecto debería estar ausente. Si la causa influencia el efecto hasta un cierto punto, entonces deberíamos ver un efecto consistentemente más fuerte o más débil, en correspondencia.
- Número cuatro: las **explicaciones alternativas** pueden ser descartadas.

Entonces déjenme ilustrar estos criterios con un ejemplo. Supongo que yo formulo una hipótesis según la cual la soledad causa sentimientos de depresión. Les doy a algunas personas solitarias y deprimidas un gato para que cuiden de él. Ahora ya no se sienten solitarias. Si mi hipótesis es correcta, entonces podría esperar que esto disminuya su depresión.

La causa y el efecto, soledad y depresión, son muy cercanas, ocurren en las mismas personas y están razonablemente próximas en el tiempo, entonces podemos mostrar que están conectadas. La causa, la disminución de la soledad, necesita suceder *antes* que el efecto, o sea, la disminución de la depresión. Podemos mostrar esto porque podemos controlar la presencia de la causa, que es la soledad.

La causa y el efecto deberían ocurrir juntas consistentemente. Esto significa que menos soledad debería ir a la par con menos depresión. Puedo encontrar un grupo de comparación con personas solitarias y deprimidas que no tengan un gato. Como la causa está ausente: su soledad no cambia, no debería haber efecto.

Hasta aquí fue fácil. La verdadera dificultad reside en el último criterio, excluir todas las explicaciones alternativas, esto es, otras causas posibles. Vamos a buscar una explicación alternativa en nuestro ejemplo. Tal vez el incremento de la actividad física, requerida para cuidar del gato, en realidad causó la disminución de la depresión, en lugar de haberlo hecho la reducción de la soledad. **Explicaciones alternativas** conforman *amenazas a la validez interna* de una investigación. una parte importante de la metodología consiste en desarrollar y escoger diseños de investigación que minimicen estas amenazas.

Hay una cosa más que quiero decir acerca de la causalidad. La *causalidad requiere correlación*; la causa y el efecto deben ocurrir consistentemente. ¡Pero **la correlación no implica causalidad**!

Les daré un ejemplo: si observamos consistentemente un comportamiento agresivo después de que los niños juegan un videojuego violento, esto no significa que el juego *causó* el comportamiento agresivo. Podría ser que los niños agresivos buscan más estímulos agresivos, y esto reversaría la relación causal. O tal vez los niños cuyos padres les permiten jugar videojuegos violentos no son supervisados de cerca. Tal vez ellos son tan agresivos como otros niños, sólo que se sienten menos inhibidos para mostrar este comportamiento agresivo. Así que recuerden: ¡la correlación no implica causalidad!

## 2.05 Método científico: amenazas a la validez interna – participantes

La validez interna es amenazada si hay una explicación alternativa plausible para los resultados de un estudio. Para juzgar la validez interna de un estudio en particular debemos estar familiarizados con los tipos de amenazas que pueden ocurrir. Comenzaré con tres tipos de amenazas que están de alguna manera asociadas con los participantes o los sujetos del estudio. Estas amenazas se llaman **maduración**, **selección** y **selección por maduración**.

Comencemos con la maduración. La maduración se refiere a una explicación alternativa formada por *cambio natural*. Supongamos que yo formulo una hipótesis según la cual la soledad causa depresión. Disminuyo la soledad en adultos mayores, que son proclives a la depresión, dándoles



un gato para cuidar de él. Espero que la depresión de ellos disminuya porque ahora están menos solos.

Encuentro un hogar de retiro que tiene el deseo de participar. Mido la depresión en un grupo de residentes que parecen infelices. Les doy un gato para lo cuiden durante cuatro meses y después mido la depresión otra vez. Vamos a suponer que la depresión está más baja después de haber cuidado del gato. ¿Significa esto que la compañía del gato *causó* la disminución en la depresión? Bueno, no necesariamente. La disminución en la depresión puede haber ocurrido naturalmente. La gente se desarrolla, cambia. Muchos problemas mentales y físicos simplemente desaparecen después de un tiempo. Incluso si no recibimos tratamiento, hay depresiones que desaparecen por sí mismas.

Afortunadamente hay una manera de eliminar esta explicación alternativa para el cambio natural al cual nos referimos como maduración. Podemos introducir un **grupo de control** que es medido al mismo tiempo, que *no es expuesto* a la causa formulada en la hipótesis. Ambos grupos deben “madurar” o cambiar en igual medida. Cualquier diferencia entre los grupos puede ser atribuida ahora a la causa formulada en la hipótesis y no al cambio natural.

La amenaza de la maduración es eliminada, pero infortunadamente un estudio que incluya un grupo de control sigue siendo vulnerable a otras amenazas a la validez interna. Esto nos lleva a la amenaza por **selección**. La selección se refiere a cualquier diferencia sistemática en las características de los grupos, distintas a la causa manipulada.

Supongamos que en mi estudio yo incluí un grupo de control que no cuidó un gato. ¿Qué pasa si la asignación de los participantes a los grupos se basó en la movilidad? Supongamos que los que no eran capaces de agacharse para limpiar la caja de arena fueron puestos en el grupo de control. Ahora supongamos que el grupo experimental estaba de hecho menos deprimido que el grupo de control. Esto podría ser causado, no por la compañía del gato, sino porque la gente del grupo experimental estaba más apta físicamente.

Una solución para esta amenaza es utilizar un método de asignación de grupos que asegure que una diferencia sistemática en las características de los miembros de cada grupo sea poco probable. El método se llama **aleatorización**. Discutiremos esto mucho más en detalle más adelante.

La última amenaza a la validez interna relacionada con los participantes es la amenaza *combinada* de maduración y selección. La llamamos **selección por maduración**. Esto pasa cuando los grupos difieren sistemáticamente en su tasa de maduración. Por ejemplo, supongamos que la efectividad del tratamiento con el gato fue examinada en un grupo experimental que consistía en voluntarios que eran abiertos a cosas nuevas. En contraste, el grupo de control consistía en personas más conservadoras, a las que no les gusta el cambio.

Los participantes fueron seleccionados de forma tal que ambos grupos tuvieran niveles similares de depresión al comienzo del estudio. Pero ¿qué pasa si nos encontramos con que el grupo experimental muestra una depresión más baja? Bueno, tal vez la tasa más baja de depresión en el grupo experimental, el de la terapia gatuna, es simplemente debida al hecho de que la gente de

mente abierta tiende naturalmente a superar sus sentimientos de depresión más rápidamente que la gente conservadora. Así como la selección en sí misma, la amenaza de la selección por maduración puede ser eliminada mediante la *asignación aleatoria* de los grupos.

Pueden ver entonces que el diseño de investigación que escogemos –por ejemplo, incluir un grupo de control– y los métodos que usamos –por ejemplo, asignación aleatoria– pueden ayudar a minimizar las amenazas a la validez interna.

## 2.06 Método científico: amenazas a la validez interna – instrumentos

Otra categoría que amenaza la validez interna está asociada con los instrumentos que son usados para medir y manipular los constructos de nuestra hipótesis. Las amenazas **baja validez del constructo, instrumentación y prueba** [*testing*] caen dentro de esta categoría.

Comenzaré con la **baja validez del constructo**. La validez del constructo es baja si nuestros instrumentos contienen un sesgo sistemático o miden un constructo o propiedad totalmente diferente. En este caso no tiene mucho sentido seguir considerando la validez interna de un estudio. Como lo dijimos antes, la validez del constructo es un *prerrequisito* de la validez interna. Si nuestros instrumentos de medida o métodos de manipulación son de pobre calidad, entonces no podemos inferir nada sobre la relación entre los constructos contenidos en la hipótesis.

Supongamos que yo formulo la hipótesis según la cual la soledad causa depresión. Intento bajar los niveles de soledad de adultos mayores que están en un hogar de retiro, dándoles un gato para que lo cuiden, esperando que su depresión disminuya. Supongamos que cuidar un gato no afecta la soledad en absoluto, sino que al contrario les dio a los residentes un estatus social más alto: son especiales porque les es permitido tener un gato. La manipulación, inicialmente dirigida a disminuir la soledad, en realidad vino a modificar una propiedad diferente, el estatus social.

Ahora consideremos mi medición de la depresión. ¿Qué pasa si el cuestionario que yo utilicé en realidad midió si los residentes se sentían socialmente aceptados, en lugar de medir si se sentían deprimidos? Si accidentalmente manipulamos el status social o medimos el “sentirse aceptado”, entonces no podemos concluir nada acerca de la relación entre la soledad y la depresión.

La segunda amenaza que se relaciona con los métodos de medición es la **instrumentación**. La amenaza de la instrumentación ocurre cuando se cambia un instrumento durante el curso del estudio. Supongamos que yo utilizo un cuestionario de autoevaluación para medir la depresión al comienzo del estudio, pero lo cambio por un cuestionario diferente o por una entrevista abierta al final del estudio. Bueno, cualquier diferencia en los puntajes de depresión puede ser explicada por el uso de instrumentos diferentes. Por ejemplo, los puntajes en el post-test, al final del estudio, pueden ser más bajos porque el nuevo instrumento mide aspectos ligeramente diferentes de la depresión.

Por supuesto parece bastante estúpido cambiar los métodos o instrumentos a mitad de camino, pero algunas veces un investigador tiene que depender de otros para los datos, por ejemplo, cuando utiliza tests que son contruidos por agencias nacionales de evaluación o por agencias encuestadoras. Un buen ejemplo es el uso de la herramienta de diagnóstico estandarizado



llamada MDE (Manual de Diagnóstico y Estadística). Este MDE es utilizado para clasificar cosas como enfermedad mental, depresión, autismo, y es actualizado cada diez o quince años.

Ahora podemos imaginarnos los problemas que esto puede causar, por ejemplo, para un investigador que esté haciendo un estudio de largo plazo sobre la esquizofrenia. En una actualización reciente del MDE, hay varios subtipos de esquizofrenia que ya no son reconocidos. Entonces, si vemos una disminución en la esquizofrenia en los años venideros, ¿es esto un efecto “real” o es debido al cambio en el sistema de medición?

La última amenaza que quiero discutir es el **testing**, también conocido como sensibilización. Administrar un test o un procedimiento de medición puede afectar el comportamiento de las personas. Una amenaza por sensibilización ocurre si este efecto sensibilizador del proceso de medición constituye una explicación alternativa para nuestros resultados. Por ejemplo, tomar el pre-test de depresión al comienzo del estudio puede alertar a la gente sobre sus sentimientos de depresión. Esto puede causar que ellos estén más proactivos respecto a mejorar su estado emocional, por ejemplo, siendo más sociables. Por supuesto esta amenaza puede ser eliminada introduciendo un grupo de control. Ambos grupos serán afectados de modo similar por el efecto *testing*. Sus puntajes de depresión decrecerán en ambos casos, pero ojalá más en el grupo acompañado por gatos.

Sin embargo, añadir un grupo de control no es siempre suficiente. En algunos casos existe el riesgo de que el pre-test sensibilice a la gente del grupo de control de una forma diferente que a la gente del grupo experimental. Por ejemplo, en nuestro estudio con gatos, el pre-test en combinación con obtener un gato puede alertar a la gente del grupo gato-experimental del propósito del estudio. Esto podría reportar niveles de depresión más bajos en el post-test, no porque estén menos deprimidos, sino para asegurarse de que el estudio parezca exitoso, ¡para quedarse con el gato!

Supongamos que la gente del grupo de control no se da cuenta del propósito del estudio, porque no se les da un gato. No están sensibilizados y no se sienten motivados a cambiar sus puntajes en el post-test. Esta diferencia en la sensibilización provee una explicación alternativa. Una solución es *añadir* un grupo experimental y un grupo de control a los que no se les aplicó el pre-test. Discutiré esta solución en más detalle cuando hablemos de los diferentes diseños experimentales.

Para resumir, la *validez interna* puede ser amenazada por la **baja validez del constructo**, la **instrumentación** y el **testing**. Las dos primeras pueden ser eliminadas utilizando *instrumentos válidos* y *métodos de manipulación válidos* y, por supuesto, usándolos *consistentemente*. El *testing* puede ser eliminado utilizando un *diseño especial* que incluya grupos que sean expuestos al pre-test y grupos que no lo sean.

## 2.06 El método científico: amenazas a la validez interna – artefactos

Otra categoría que amenaza la validez interna está asociada con la reacción “artificial” o no natural causada por la participación en la investigación. esta puede ser una reacción del participante, ¡pero también del investigador!

En cualquier estudio científico el investigador tiene que observar el comportamiento de los individuos o grupos. Y en muchos casos, la gente bajo observación está al tanto de estar siendo observada. Tanto el investigador como los participantes tienen expectativas sobre el propósito del estudio y el comportamiento deseado. Estas expectativas pueden conducir a un cambio en el comportamiento que puede proveer una explicación alternativa. Discutiré dos de estas amenazas a la validez interna aquí: la expectativa del experimentador y las características de la demanda.

Si las expectativas del investigador tienen un efecto predisponente, llamamos a esta amenaza a la validez interna un efecto de **expectativa del experimentador**. La expectativa del experimentador se refiere a un cambio inconsciente en el comportamiento del investigador, causado por expectativas sobre el resultado, que influencia las respuestas de un participante. Por supuesto esto se vuelve un problema mayor si el investigador inconscientemente trata a la gente del grupo de control de manera diferente que a la gente en el grupo experimental.

Una de las demostraciones más sutiles y sorprendentes de este efecto de expectativa del experimentador fue indudablemente provista por Rosenthal en los años 1960s. Unos estudiantes de psicología fueron conducidos a creer que estaban tomando parte en un curso de competencias prácticas en investigación, manipulando animales para “duplicar resultados experimentales”:

Los estudiantes manipularon uno o dos lotes de ratas que fueron comparadas su desempeño al recorrer un laberinto. El primer lote, les dijeron a los estudiantes, habían sido cruzado para nacer con una excelente habilidad espacial. El otro lote había sido cruzado con el propósito opuesto. Y he aquí que hubo una diferencia en el desempeño de las ratas; las ratas que eran brillantes para los laberintos superaron a las mazas que eran tontas para los laberintos. Por supuesto en realidad no había diferencia en cuanto a la brillantez para los laberintos entre los dos grupos de ratas, ambos fueron el resultado de los mismos cruces y fueron asignados aleatoriamente a los grupos “brillantes” y “tontas”.

Aparentemente, sólo con saber lo que se esperaba de su lote de ratas condujo a los estudiantes a tratarlas de forma diferente, resultando de hecho en una diferencia en el desempeño de las ratas. ¡Imaginen cuál puede ser el efecto cuando un investigador interactúa con participantes humanos!

Afortunadamente hay una solución. Si el experimentador que interactúa con los participantes no sabe qué desempeño se espera de ellos, no podrá inconscientemente influenciar a los participantes. A esto le llamamos diseño de **experimentador ciego**.

Por supuesto los participantes también pueden tener expectativas sobre el propósito del estudio. Los participantes que están al tanto de que son sujetos de un estudio pueden buscar pistas o señales para averiguar de qué se trata el estudio. Las **características de la demanda** se refieren a un cambio en el comportamiento del participante debido a sus expectativas sobre el estudio. Las características de la demanda son especialmente problemáticas si la gente responde de manera diferente en el grupo de control y en el grupo experimental.

Una forma bien conocida de características de la demanda ocurre cuando las personas están al tanto de que están en un grupo experimental y están en el curso de un tratamiento, especialmente si el tratamiento se propone ayudarles. Los participantes pueden estar agradecidos

de estar en este grupo, o esperanzados en que el tratamiento va a ser efectivo. Y esto puede conducirlos a ser más positivos respecto del tratamiento que si no están al tanto de él.

Pero las señales que los participantes buscan en el estudio pueden no ser adecuadamente interpretadas por ellos. En la medida en que los participantes del mismo grupo interpreten esas señales de la misma manera y cambien su comportamiento de la misma manera, las características de la demanda constituyen un problema real.

Esta es la razón por la cual siempre es una buena idea que los estudiantes no estén al tanto del propósito real del estudio o que por lo menos no sepan en qué grupo están, si el experimental o el de control. Si tanto el sujeto como el investigador ignoran estas cosas, llamamos a este diseño de investigación **doblemente ciego**.

Dado que cualquier señal puede llevar a un sesgo en el comportamiento, los investigadores por lo general inventan una **historia fachada**. Una historia fachada es una explicación plausible del propósito del estudio. Esto puede darles a los participantes señales que difícilmente sesgarán su comportamiento. Por supuesto este engaño temporal debe ser *necesario*, y el riesgo de sesgo debido a las características de la demanda debe ser un riesgo real. Y por supuesto debemos informar a los participantes después sobre cuál fue el propósito real del estudio.

## 2.08 El método científico: amenazas a la validez interna – diseño / procedimiento

La última categoría de amenazas a la validez interna es una categoría tipo recordatorio. De forma muy general, los tres tipos de amenazas en esta categoría están relacionadas con el procedimiento de investigación. estas amenazas son: **precedencia temporal ambigua, historia y mortalidad**.

Una **precedencia temporal ambigua** en la relación causal formulada en la hipótesis es simplemente una forma elegante de decir que no es claro si la causa contenida en la hipótesis en realidad *precede* el efecto observado. Supongamos que estoy interesada en la relación entre jugar un videojuego violento y el comportamiento agresivo. Les pregunto a estudiantes de secundaria cuántas horas a la semana pasan jugando videojuegos violentos y le pido a su profesor que califique su agresividad en la clase.

¿Qué pasa si encuentro una fuerte relación: los niños que juegan videojuegos violentos durante muchas horas a la semana también muestran un comportamiento más agresivo? Bueno, esto *no significa* que jugar videojuegos violentos cause un comportamiento agresivo. Puede ser que los niños que juegan más videojuegos violentos eran más agresivos desde el comienzo y más proclives a buscar estímulos violentos.

La amenaza de la precedencia temporal ambigua puede ser eliminada manipulando o introduciendo la causa formulada en la hipótesis. Por supuesto no todos los constructos pueden ser manipulados. Pero si yo manipulo la causa puedo asegurarme de que esta ocurra *antes* que el efecto. Por ejemplo, puedo hacer que *todos* los niños jueguen videojuegos violentos. Si los niños que no fueron agresivos al comienzo se vuelven más agresivos *después* de jugar el videojuego violento, entonces mi inferencia causal es mucho más fuerte.

Avancemos hacia la amenaza conocida como **historia**. El efecto de la historia es un evento imprevisto que ocurre durante el estudio y provee una explicación alternativa. Este puede ser un evento a gran escala o algo pequeño que sale mal durante la recolección de datos.

Consideremos un estudio para mitigar estereotipos negativos sobre un grupo minoritario. La manipulación consiste en un grupo de discusión, orientado por un investigador. El investigador se enfoca en el punto de vista del grupo minoritario, pidiéndoles a los estudiantes que se pongan en los zapatos de los miembros de ese grupo. En el grupo de control el investigador se centra en las diferencias entre la mayoría y la minoría y enfatiza en el punto de vista de la mayoría. En ambos grupos hay tres discusiones semanales.

Para dar un ejemplo del efecto de la historia a pequeña escala, imaginen que durante la última sesión del grupo de control el experimentador se desmaya. Por supuesto los participantes están sorprendidos y angustiados por esto. Y esto puede traducirse en una actitud general más negativa en el grupo de control, lo cual también hace que la actitud del grupo de control hacia la minoría sea más negativa. El tratamiento puede parecer efectivo, porque el grupo experimental es más positivo pero la diferencia se debe, no a la técnica de la discusión, sino al incidente del desmayo.

Ahora consideremos el efecto de la historia a gran escala. Supongamos que durante el estudio un horrible asesino es comedido, aparentemente, por un miembro de la minoría. El crimen atrae una enorme atención por parte de los medios, reforzando el estereotipo negativo sobre el grupo minoritario. Este evento puede deshacer cualquier efecto positivo de la intervención.

La amenaza de la historia es difícil de eliminar. No es posible evitar eventos de gran escala. Sí podemos evitar los eventos de pequeña escala que ocurren durante, al menos en cierta medida, examinando separadamente a los participantes, siempre que sea posible. De esta manera, si algo sale mal, los resultados de uno o unos pocos sujetos tendrán que ser descartados.

La última amenaza que vamos a discutir aquí es la **mortalidad**. La mortalidad se refiere a un participante que *deserta* del estudio. Si se comparan los grupos y la desertión es diferente entre ellos, entonces esto puede proveer una explicación alternativa. Por ejemplo, supongamos que estamos investigando la efectividad de una droga para la depresión. Supongamos que la droga en realidad no es muy efectiva, y tiene un muy fuerte efecto secundario. ¡Causa flatulencias extremas!

Por supuesto esto puede ser tan incómodo y vergonzoso que no sería extraño que la gente desertara del estudio por causa de este efecto secundario. Supongamos que el 80% de los miembros del grupo experimental han desertado. En el grupo de control a los participantes se les administra un placebo, que no tiene el ingrediente activo y, por lo tanto, tampoco los efectos secundarios. La desertión en este grupo de control es de sólo el 10%. Es obvio que los grupos ya no son comparables.

Supongamos que para el restante 20% de los participantes del grupo experimental la droga es lo suficientemente efectiva como para soslayar el efecto secundario negativo. Este no fue el caso para el 80% que desertó. Basándonos en los sujetos restantes podríamos concluir que la droga es

muy efectiva. Pero si todos los sujetos hubiesen permanecido en el estudio la conclusión hubiese sido muy diferente.

La amenaza de la mortalidad es muy difícil de eliminar, en muchos casos lo mejor que se puede hacer es documentar las razones para la deserción de manera que esas razones puedan ser investigadas y posiblemente mitigadas en estudios ulteriores.

## 2.09 El método científico: variables de interés

Quiero adentrarme ahora en los diseños de investigación y darles una idea más concreta de cómo un diseño de investigación puede minimizar las amenazas a la validez interna. Pero antes de poder hacer eso tenemos que estar familiarizados con los términos **constructo**, **variable**, **constante** y **variable independiente** y **dependiente**.

Una hipótesis describe o explica la relación entre *constructos*. El término **constructo** es utilizado para indicar que estamos hablando de una propiedad en términos abstractos, generales. Por ejemplo, yo puedo plantear la hipótesis según la cual la soledad y la depresión están asociadas. Los términos soledad y depresión son los *constructos* en este caso.

Por supuesto la soledad y la depresión pueden ser expresadas de maneras muy diferentes. El término **variable** se refiere a una **versión operacionalizada de un constructo**. Una variable es una expresión específica, concreta, del constructo y es **medible** o **manipulable**. Por ejemplo, yo puedo operacionalizar la soledad en un grupo de adultos mayores en un hogar geriátrico utilizando un cuestionario de autoevaluación. Puedo administrar la escala de soledad diseñada por la UCLA. Esta escala es un cuestionario de 20 puntos consistente en ítems como “No tengo a nadie con quien hablar”. La variable soledad ahora se refiere a la soledad como se expresa en los puntajes propuestos por la escala UCLA.

Si yo planteo la hipótesis de que la soledad *causa* depresión, me iría mejor manipulando en lugar de midiendo la soledad. Puedo darle a un grupo de adultos mayores un gato para cuidarlo, comparándolos con el grupo de control que no tiene gato. Ahora he operacionalizado la soledad creando dos niveles de soledad. La variable soledad ahora se refiere a “alta” o “baja” expresada a través de la presencia o ausencia de la compañía felina.

Finalmente, puedo operacionalizar la depresión utilizando la Escala de Depresión Geriátrica, EDG, que consiste en 15 preguntas del tipo “¿Se siente feliz la mayor parte del tiempo?”. La variable depresión ahora se refiere a la depresión como se expresa en los puntajes de la EDG.

Entonces las variables son propiedades que son medidas o manipuladas y que adquieren diferentes valores. Esta última parte es muy importante, esto es, que los valores de las variables tienen que variar; de otro modo la propiedad no sería muy interesante. Supongamos que el hogar geriátrico es tan horrible que todos los residentes tienen el puntaje máximo en depresión. Bueno, entonces no podemos demostrar una relación entre soledad y depresión, o al menos no en este grupo de sujetos. Tanto la gente solitaria como la menos solitaria estará igualmente deprimida. La depresión sería una **constante**.

Así que las variables centrales para nuestra hipótesis deberían ser *variables*, deberían mostrar variación. Por supuesto es una buena idea mantener *otras* variables irrelevantes constantes, de forma que no puedan proveer explicaciones alternativas, pero veremos eso más adelante.

Ahora que hemos definido lo que es una variable, veamos los diferentes tipos de variables de acuerdo con el papel que juegan en la descripción o la explicación del fenómeno. Me referiré a las variables centrales para nuestra hipótesis como **variables de interés**.

Cuando una causa y un efecto no pueden ser identificados o cuando la dirección causal no es obvia o incluso no es de interés, nuestras variables están “en pie de igualdad”. En ese caso sólo nos referimos a ellas como variables. Pero cuando nuestra hipótesis es causal, podemos identificar la causa y el efecto. Y entonces nos referimos a la variable causa como la **variable independiente** y a la variable efecto como la **variable dependiente**.

También nos referimos a la variable *independiente* como variable **causa**, **explicativa**, variable de **entrada** o **predictora**. Esto se refiere a una variable que asume el papel de causa o predictora de otra variable en la hipótesis. En nuestro ejemplo se formula la hipótesis según la cual la soledad causa depresión. La variable independiente aquí por supuesto es la soledad, y es operacionalizada a través de la presencia o ausencia de un gato.

La variable *dependiente* en la hipótesis es influenciada por la variable causa o se le considera resultado de otra variable. Estos valores *dependen* de otra variable. En nuestro ejemplo la variable *dependiente* fue la depresión, medida a través de los puntajes en el cuestionario EDG. La variable dependiente también es llamada variable **efecto**, variable **respuesta**, variable de **salida** o variable de **resultado**.

Ahora si están teniendo problemas diferenciando los términos independiente y dependiente, traten de recordar que la variable independiente es la que el investigador quiere controlar, es la causa que viene primero.

## 2.10 El método científico: variables de desinterés

He discutido variables que están en el centro de nuestra hipótesis: las *variables de interés*. Por supuesto en cualquier estudio habrá *otras* propiedades superficiales asociadas con los participantes y el diseño de la investigación que cambian de un participante a otro.

Estas propiedades no son el principal foco de nuestro estudio, pero pueden ser asociadas con nuestras variables de interés, proveyendo posibles explicaciones alternativas. Esas variables de *desinterés* vienen en tres sabores: **confusores**, **variables de control** y **variables de antecedentes**.

Un confusor o variable latente es una variable que está relacionada tanto con la variable independiente como con la variable dependiente y que en parte o por completo da cuenta de la relación entre ellas dos. Supongamos que investigo el efecto de reducir la soledad y la depresión en un grupo de adultos mayores. Yo disminuyo la soledad entregándole un gato a cada adulto mayor participante en el grupo experimental y los adultos mayores del grupo de control reciben un muñeco de peluche.



Ahora, además de la soledad, los dos grupos pueden también diferir en términos del ejercicio físico que hacen, su edad o la susceptibilidad a la depresión, las cuales son todas variables de desinterés. Tomemos el ejercicio físico. El grupo experimental será más proclive a la actividad física porque tienen que alimentar al gato y limpiar. Entonces el ejercicio físico está relacionado con la soledad, el grupo del gato –el grupo menos solitario– es más activo que el grupo de control.

Supongamos que la actividad física está también causalmente relacionada con la depresión –ser más activo disminuye la depresión. Bueno, entonces la actividad física, y no la soledad, puede dar cuenta un decrecimiento en los puntajes de depresión en el grupo experimental del gato. La relación entre soledad y depresión, se dice en este caso, es **espuria**. La relación puede ser explicada por la *variable de confusión*, la actividad física.

Algo importante que debe ser notado acerca de los confusores es que no están incluidos en la hipótesis y generalmente no son medido. Esto hace imposible determinar cuál es el efecto real de un confusor. La única cosa que podemos hacer es repetir el estudio y controlar el confusor asegurándonos de que tenga el mismo valor para todos los participantes. Por ejemplo, si todos los adultos mayores en ambos grupos tienen que ser igualmente activos, entonces la actividad física no puede explicar diferencias en la depresión.

Otra posibilidad es convertir un confusor en una variable de control. Una **variable de control** es una propiedad que es más proclive a ser relacionada con la variable independiente y dependiente, tal como un confusor. Pero a diferencia de este último, la variable de control es medida. Sus efectos pueden por lo tanto ser evaluados y controlados.

Por ejemplo, podríamos ver si la actividad física provee una explicación alternativa que deba ser medida y tomada en cuenta. Supongamos que podemos distinguir a la gente activa de la gente inactiva. En la gatoterapia hay más gente activa, pero algunos son inactivos. En el grupo de control la mayoría de la gente es inactiva, pero algunos son activos.

Ahora consideremos la diferencia en la depresión entre los grupos de gatoterapia y de control, primero para gente activa y después para la gente inactiva. Controlamos la actividad, manteniéndola constante y considerando cada nivel de actividad separadamente.

Si la relación entre soledad y depresión desaparece cuando miramos a cada nivel de actividad separadamente, entonces la actividad “explica el por qué” de la relación espuria entre soledad y depresión. Pero si la relación se sigue mostrando en cada nivel de actividad, entonces hemos eliminado la actividad física como una explicación alternativa para la caída en los puntajes de depresión.

El último tipo de variable de desinterés es una variable de **antecedentes**. Este tipo de variable no es inmediatamente relevante para la relación entre las variables de interés, pero es relevante para determinar qué tan representativos son los participantes en nuestro estudio, con respecto a un grupo más grande, tan vez adultos mayores de todo el mundo, o incluso toda la gente de todas las edades. Por esta razón es interesantes saber cuántos hombres y cuántas mujeres participaron, cuáles eran sus edades promedio, sus antecedentes étnicos o culturales, su nivel socioeconómico, su nivel educativo y cualquier otra cosa relevante para el estudio.

Para resumir, un **confusor** es una variable que explica parcial o totalmente un efecto sobre la variable dependiente en lugar de o adicional a la variable independiente. Un confusor no es tenido en cuenta en la hipótesis y no es medido o controlado por el estudio. Un posible confusor puede ser controlado manteniendo constante el valor de la propiedad o convirtiéndolo en una variable de control.

Una **variable de control** da cuenta de un posible confusor midiendo la propiedad relevante y revisando la relación entre las variables de interés, en cada valor o nivel de la variable de control.

Las **variables de antecedentes**, finalmente, son medidas, no porque se espere un posible efecto sobre las variables de interés, sino porque las propiedades contextuales de los participantes son útiles para evaluar la posibilidad de generalizar el estudio con base en las características de la muestra.

DOCUMENTO DE TRABAJO